Building on the Host: Audit-Oriented Data Reuse and the Symbiotic Knowledge Infrastructures of Climate Crisis Skeptics

by

Morgan F. Wofford

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Information)
in The University of Michigan
2026

Doctoral Committee:

Associate Professor Libby Hemphill, Co-Chair
Associate Professor Andrea Thomer, University of Arizona, Co-Chair
Professor Anna Kirkland
Emeritus Professor Elizabeth Yakel

Morgan F. Wofford

mwofford@umich.edu

ORCID iD: 0000-0002-4688-0133

© Morgan F. Wofford 2026

Dedication

To Mikey, I love you more.

Acknowledgements

First and foremost, I want to thank to my advisors and dissertation co-chairs, Andrea Thomer and Libby Hemphill. Andrea, you are the reason I came to Michigan, and even after you left for Arizona you supported me just as fully from a distance. You always pushed me toward precision in my definitions and methods. I promise to clearly define how I operationalize infrastructural inversion in all future works. You taught me to think infrastructurally, and, most meaningfully, you helped me see how to bring my love of the natural world together with information science, connecting me to a wider community at ESIP and beyond. Libby, you brought an interdisciplinary perspective that reshaped how I approach this work, and you consistently tied my research back to the real world outside it. I am grateful for that grounding, and for what you taught me, through the way you advise, about what it means to be a mentor. You both gave me the freedom to define my own research problem and the support to see it through, and you showed me that a life outside the work is not a concession but part of doing the work well.

I am grateful to my committee members, Anna Kirkland and Elizabeth Yakel, whose attention made this a better piece of work than I could have written alone. Anna, our recurring meetings and your outside perspective were invaluable, especially your guidance on working with controversial communities and on conceptualizing data misuse. Beth, thank you for your close readings, and for sharing your expertise on trust, data curation, data reuse, and so much more. Your standards were high and a genuine pleasure to meet, and it is an honor to be the last dissertation committee you sat on.

I am indebted to the mentors who set me on this path long before the dissertation began. Christine Borgman, without you I would not be in a PhD program or drawn to this topic at all. You began my journey as an information science scholar. You never stopped offering support, you celebrated every success along the way, you never treated academia as a hierarchy, and you opened your community to me. I cannot thank you enough. Irene Pasquetto, traveling from UCLA to Michigan with you made the transition so much easier, and your intellectual

fingerprints are all over this work. You included me continuously and generously. Your expertise in data misuse, scientific misinformation, and open data runs throughout this dissertation. Just search your name. I am grateful, too, to the members of the UCLA Center for Knowledge Infrastructures I overlapped with, including Bernadette Boscoe, Milena Golshan, Michael Scroggins, and Peter Darch, each of whom helped me grow as a scholar.

Many audiences read and listened to earlier versions of this work. I thank the mentors of the doctoral colloquia I took part in: at the ASIS&T colloquium, Devan R. Donaldson, whose help with my semi-structured interview protocol was formative, and the participants of the iSchool colloquium. I thank the members of LARC for their feedback as this project took shape including Ricky Punzalan, David Wallace, Jesse Johnson, Rebecca Frank, Patricia Garcia, Dallas Card, Nazelie Doghramadjian, Sony Prosper, Tam Rayan, Alli Rayburn, Yvette Ramirez, Lavinia Dunagan, Pelle Tracy, and Suanmuanlian Tonsing. And to my lab group — Angela Schopke-Gonzalez, Megan Brown, Lingyao Li, Anmol Panda, and Ji Eun Kim — thank you for the community and the interdisciplinarity. Angela, thank you for asking the deep questions about research and about life, and for calling me back to the ethical dimensions of this work whenever it began to feel routine.

The friends I made in the program carried me through it as much as any part of the work did. Alli Rayburn and Tanisha Afnan, thank you for never asking about my dissertation unless I brought it up first. Even when our research overlapped, you gave me the rare gift of friendship that had nothing to prove. To my friends beyond academia, especially Sarah and Melanie, thank you for staying beside me through so many versions of myself, and for forgiving me every time I went MIA.

This dissertation was written at the Alameda Free Library, in more coffee shops than I can count, and mostly in my home office, where I spent far too many hours. I am grateful, too, for the practices that kept me steady: crochet, dancing, and time outdoors, each of which kept my body busy and my mind quiet when the writing would not come.

Shubham, you could be thanked in every paragraph of this acknowledgement. But mostly, thank you for tethering me to life outside the dissertation even at its busiest, for taking me on adventures, and for your advice. When the work felt endless, you kept showing me how much the world beyond the page mattered, and you made sure I actually went out into it.

To my parents, thank you for supporting me in everything, always, even from across the country. You and my grandparents taught me the importance of education and of hard work, and you loved me unconditionally through every step of a long path. Everything I am and everything in these pages began with you.

And to Mikey. You did not make this easier, but you put everything in perspective and reminded me what truly matters. Thank you for teaching me to enjoy the little things, to love, and to meet people very different from myself with empathy and understanding, a lesson that sits at the very heart of this research. Thank you for showing me how to be spontaneous, to find joy, to never take a single day for granted, and to ultimately live for myself and no one else. I know that if you were here, you would be so proud, and we would be celebrating just as we did for UCLA.

Finally, to Vini the cat, who kept me in my chair by falling asleep on my lap exactly when the writing got hardest, and to Charlie the dog, who was there at the beginning, a steady, cuddly constant through an unstable time.

Preface

Portions of this dissertation draw on previously published collaborative work. Section 2.1.1 adapts material from (Pasquetto et al., 2026), where I was the primary author of the ACC skeptic case study. Sections of Chapter 5, 6, and 7 draw on Pasquetto et al. (2026), Wofford (2022), and Wofford & Thomer (2023). I am the primary author of that material as it appears here, and I have revised it for this dissertation. Section 2.4 and Section 7.5.3 additionally draw on (2024), a collaborative article on which I was not the lead author. Passages adapted from that work are identified in footnotes at relevant sections, and specific claims are cited in text.

My co-authors' contributions to the original publications are gratefully acknowledged.

Table of Contents

Dedication ii

Acknowledgements iii

Preface vi

List of Tables xiii

List of Figures xiv

List of Illustrations xv

List of Appendices xvii

Abstract xviii

Chapter 1 Introduction 1

1.1 The Promise and Limits of Open Data 1

1.2 Study Contributions 4

Chapter 2 Theoretical Frameworks: Tracing Data Reuse Practices in Communities and the Knowledge Infrastructures They Rely Upon Through Data Journeys 8

2.1 Data Communities 9

2.1.1 The ACC Skeptic Data Community 10

2.2 Examining data reuse practices: A practice-theoretical perspective 12

2.2.1 Data practices are heterogeneous and situated 13

2.2.2 Applying a practice-theoretical perspective to ACC skeptics' data reuse 15

2.3 Situated activities in data reuse practices 16

2.3.1 Data reuse prompt and design 17

2.3.2 Seeking and selecting data 17

2.3.3 Processing and transforming data 19
2.3.4 Analyzing, interpreting, and abstracting data 19
2.3.5 Managing, documenting, and preserving data 20
2.3.6 Publishing and sharing research 20
2.3.7 Attributing and citing data 21
2.3.8 Publishing and sharing data 21
2.4 Data misuse 22
2.4.1 Defining data misuse as a sociotechnical phenomenon 22
2.4.2 Implications for this study 26
2.5 Understanding knowledge infrastructures: Foundations of data practices 27
2.5.1 Climate science knowledge infrastructure and the persistent presence of skeptics . 29
2.5.2 The open data infrastructure 33
2.6 Studying data practices and knowledge infrastructures through data journeys 37
2.7 Conclusion 39
Chapter 3 Methods 41
3.1 Digital Ethnography 43
3.1.1 Boundaries of Digital Ethnography 45
3.2 Content analysis 48
3.3 Participant observation 51
3.4 Semi-structured interviews 53
3.4.1 Recruitment and sampling 55
3.4.2 Member checking 57
3.5 Data analysis 59
3.6 Addressing Positionality and Epistemology 62
3.7 Ethical considerations 64

3.8 Limitations and Risks 67

Chapter 4 The Data Community and Four Data Journeys 69

4.1 The Data Community 69

4.1.1 Pathways into ACC skepticism and data reuse 70

4.2 Infrastructuring: Building Parallel Data Systems (Dave Burton — sealevel.info) 73

4.2.1 Data reuse prompt and project design 74

4.2.2 Seeking and selecting data 76

4.2.3 Processing and transforming data 78

4.2.4 Managing, documenting, and preserving data 79

4.2.5 Analyzing, interpreting, and abstracting data 83

4.2.6 Attributing and Citing Data Sources 85

4.2.7 Publishing and sharing data 86

4.2.8 Analysis: 87

4.3 Rapid Reuse: Contesting Media Narratives (Burnett: *Climate Realism*) 89

4.3.1 Data Reuse Prompt and Design 89

4.3.2 Seeking and Selecting Data 91

4.3.3 Processing and transforming data 94

4.3.4 Analyzing and Interpreting Data 95

4.3.5 Publishing, and sharing data 96

4.3.6 Attributing and citing data 96

4.3.7 Publishing and sharing research 96

4.3.8 Managing, documenting, and preserving data: 97

4.3.9 Analysis: 97

4.4 The Human Contribution to Atmospheric Carbon Dioxide: Skeptical data practices and a mainstream-aligned conclusion (Engelbeen - CO_2 Coalition) 99

4.4.1 Data reuse prompt and design 99

4.4.2 Seeking and selecting data 100
4.4.3 Processing and transforming data 102
4.4.4 Analyzing, Interpreting, and Abstracting Data 102
4.4.5 Managing, documenting, and preserving data 107
4.4.6 Publishing and sharing research data 108
4.4.7 Attributing and citing data 108
4.4.8 Publishing and sharing research 108
4.4.9 Analysis 109
4.5 Replication and Quality Assessment Reuse: Academic Challenges to Policy (McKitrick - Nature Scientific Reports) 111
4.5.1 Background and Research Design: 111
4.5.2 Seeking and selecting data 113
4.5.3 Processing and transforming data 114
4.5.4 Analyzing and Interpreting Data 115
4.5.5 Managing, documenting, preserving, and sharing data 117
4.5.6 Attributing and citing data 118
4.5.7 Publishing and sharing research 118
4.5.8 Analysis 119
4.6 Conclusion 121
Chapter 5 Data Practices of ACC Skeptics 124
5.1 Data reuse prompt and project design 125
5.1.1 Research topics 125
5.1.2 What prompts data reuse 127
5.1.3 Research design and planning 131
5.2 Finding and selecting data 135
5.2.1 Assessing data quality and trustworthiness 138

5.3 Processing and transforming data 146
5.4 Analyzing, interpreting, and abstracting data 149
5.4.1 Comparing datasets, models, and claims 150
5.4.2 Identifying trends and relationships 152
5.4.3 Reproducing and reanalyzing prior studies 154
5.5 Managing, documenting, and preserving data 154
5.6 Publishing and sharing research 156
5.7 Attributing and citing data 158
5.8 Publishing and sharing data 160
Chapter 6 Knowledge Infrastructures They Rely Upon 166
6.1 Technical infrastructure 167
6.1.1 Data repositories and archives 167
6.1.2 Computational tools and platforms 170
6.1.3 Publication Platforms 173
6.2 Social infrastructure 175
6.2.1 Social networks and communities 176
6.2.2 Epistemic Values as Infrastructure 179
6.3 Infrastructural Inversion and Friction 183
6.3.1 The politics of access, archiving, and deletion 183
6.3.2 Data Misuse accusations as infrastructural friction 185
6.4 Pointing Forward 190
Chapter 7 Discussion 191
7.1 The audit orientation embedded into the KI 191
7.2 Radical scientific values 196
7.2.1 Transparency 197

7.2.2 Mechanical objectivity 198
7.2.3 Aperspectival objectivity 199
7.2.4 Interoperability 200
7.3 Symbiotic knowledge infrastructures 201
7.4 Data misuse as infrastructural friction 204
7.5 Practical implications for data intermediaries and repositories 206
7.5.1 Reframing open data governance 207
7.5.2 Designing for trust across differences 211
7.5.3 Designing for and responding to potential data misuse 213
7.5.4 The role of information professionals with skeptic communities 217
Chapter 8 Conclusion 220
8.1 Final Thoughts 220
8.2 Limitations 222
8.3 Future Work 223
Appendices 224
Bibliography 257

List of Tables

Table 2.1: Ways of misusing research data identified in the literature on data misuse, and examples of such misuses, drawn directly from Pasquetto et al. (2024). 23

Table 3.1: Overview of Research Activities 44

Table 3.2: Documents archived and coded during content analysis, September 2021 through December 2025. 50

Table 3.3: Overview of semi-structured interview participants, including interview format and role in the ACC skeptic community. 56

Table 4.1: Datasets obtained by McKitrick for his article. 113

Table 4.2: Sample size progression across McKitrick's dataset recovery process. 115

Table 4.3: Comparative summary of four ACC skeptic data reuse journeys. 121

List of Figures

Figure 2.1: Theoretical framework. Data journeys and data misuse accusations are used to study the data practices of data communities, and the knowledge infrastructures on which they rely. 9

Figure 3.1: Workflow during a monitoring session for ACC skeptics following the data and questioning fixed locations. 48

Figure 4.1: The data reuse journey map of Dave Burton's SeaLevel.info. 73

List of Illustrations

Image 4.1: Screenshot of NOAA's linear mean sea level (MSL) trends page, as archived by the Wayback Machine. Burton went to this site during his semi-structured interview to illustrate the page he first encountered in 2009. 74

Image 4.2: Screenshot of sealevel.info's mean sea level visualization and analysis tool for the Świnoujście, Poland tide gauge station, showing the site's regression analysis and display options (Burton, 2026b). 83

Image 4.3: Screenshot of sealevel.info's citation of the Permanent Service for Mean Sea Level (PSMSL) dataset, illustrating Burton's attribution practices. 85

Image 4.4: Screenshot of attribution metadata displayed at the top of sealevel.info's archived text file for historical CH_4 mixing ratios from Law Dome (Antarctica) and Summit (Greenland) ice cores. 86

Image 4.5: Screenshot of sealevel.info's data citations for the Świnoujście, Poland tide gauge station, illustrating Burton's attribution of IMGW as the data source. 86

Image 4.6: The U.S. Heat Wave Index graph as embedded in Burnett's *Climate Realism* blog post, sourced from the EPA's Climate Change Indicators: Heat Waves page. 91

Image 4.7: Screenshot of the FAOSTAT Compare Data interface showing Burnett's variable selections for his apple production analysis: United States of America, 1979–2022, with area harvested, production quantity, and yield selected as elements, and apples as the item. 94

Image 4.8: Screenshot of Burnett's FAOSTAT graphs posted on blog post. 95

Image 4.9: Figure 2.3.2 from Engelbeen et al. (2024, p. 51) showing "Ice core CO_2 in four ice cores, Law Dome DE08 and DE08-2 have a resolution of eight years and Law Dome DSS and Siple Dome a resolution of around 20 years. The Law Dome ice cores have an overlap with the South Pole CO2 data (1958–1978) within a few ppmv of each other (NOAA ongoing)." 103

Image 4.10: Figure 2.3.3 from Engelbeen et al. (Engelbeen et al., 2024, p. 52) showing "CO_2 levels as measured at Linden/Giessen under inversion (HLNUG, ongoing) compared to hourly averaged raw CO_2 data from Barrow, Mauna Loa and the South Pole for the same days, including all local outliers. Data from (NOAA ongoing)." 104

Image 4.11: Figure 2.3.4 from Engelbeen et al. (Engelbeen et al., 2024, p. 53) "Diurnal CO2 change at Linden/Giessen under inversion (HLNUG ongoing)." 105

Image 4.12: Figure 2.3.5 from Engelbeen et al. (2024, p. 53) showing “Monthly averages at Linden/Giessen compared to Mauna Loa data.” 106

Image 4.13: Figure 1 from McKitrick (2025) showing yield change simulations (%) for four crop types, maize/millet/sorghum, rice, wheat, and soybean, at warming levels from 1°C to 6°C. Blue lines show results based on the incomplete C14 dataset (N=862); green lines show results based on the expanded dataset (N=1,222). The shift from predominantly negative to neutral or positive yield projections across crop types illustrates the effect of the dataset expansion. Source: McKitrick (2025, Figure 1). 117

Image 5.1: Dr. John Christy’s comparison of modeled and observed mid-tropospheric temperature. Lines show five-year-average variations from 32 climate models (102 simulations); circles show weather-balloon observations and squares show satellite observations (*Testimony of John R. Christy*, 2017, p. 5). 151

Image 5.2. COMM001 posted UAH satellite temperature anomaly data directly into a blog comment thread in response to claims about warming trends (COMM001, Blog, 2025). 158

Image 5.3. “Figure 1.9/5.1: More Fossil Fuels, Fewer Climate-Related Deaths” (ACC038, Website, 2021). 164

List of Appendices

Appendix A Semi-structured Interview Protocol 225

Appendix B Codebook.....231

Appendix C Public Datasets Found 254

Abstract

The open data movement is often framed as a means of broadening access to scientific knowledge and enabling new forms of participation in research. This dissertation examines what happens when open scientific data are reused by communities whose goals often diverge from those of the scientists who generated them. Drawing on a digital ethnography of anthropogenic climate crisis (ACC) skeptics that combines content analysis, participant observation, and semi-structured interviews, this dissertation investigates how a technically capable data reuser community at the margins of mainstream science engages with open climate data and builds a knowledge infrastructure of its own. The dissertation finds that ACC skeptic data reuse was organized around a consistent audit orientation: community members approached open data as a resource for scrutinizing scientific claims and the interpretive decisions embedded in data production, processing, and presentation. This orientation shaped every stage of reuse, from what motivated engagement with data to how findings were interpreted and circulated. Community members generally preferred "raw" and observational data over processed or modeled outputs directed their skepticism toward institutional interpretive decisions rather than data collection itself and disseminated findings through a hybrid infrastructure spanning peer-reviewed journals, grey literature, blogs, and community-built analytical tools. Theoretically, the dissertation extends knowledge infrastructure theory by showing that the relationship between the ACC skeptic and mainstream climate science knowledge infrastructures is best understood as a symbiosis in the ecological sense: by turns parasitic, competitive, and occasionally mutualistic with mimicry running throughout. It also argues that the community does not reject scientific values, but instead intensifies them, pushing transparency, objectivity, and interoperability to the point where they begin to work against the mainstream institutions that uphold them. Finally, the dissertation reconceptualizes contested data reuse, often termed data misuse, as a form of infrastructural friction that reveals divergences between the values embedded in open data systems and the practices of their users. ACC skeptics, in this account, are not anomalous misusers of open data but a predictable product of openness itself. The findings have

implications for repositories, data stewards, and information professionals designing open data systems for diverse and sometimes contested user communities.

Chapter 1 Introduction

1.1 The Promise and Limits of Open Data

One of the most compelling promises of open data, or data that are made publicly available, is their potential to democratize and diversify who can use data and generate knowledge from them (Baack, 2015; Cavalier & Kennedy, 2016; Espinosa et al., 2014; Nielsen, 2011; Ricker et al., 2020; Zuiderwijk & Janssen, 2014). Rather than being available only to academics and professional researchers, open data are available to anyone, including citizen scientists, students, conspiracy theorists, community organizers, activists, teachers, and more. Open data can enable individuals and communities to generate their own knowledge, representing a "democratization of information" (Baack, 2015, p. 3). For example, citizen groups such as the Anti-Eviction Mapping Project and the Mapping Police Violence Database use open government data to support the needs of historically marginalized communities (*Anti-Eviction Mapping Project*, 2021; *Mapping Police Violence*, 2021). These data reusers often view their community activities as philanthropic endeavors (Kassen, 2021).

These two mapping projects demonstrate the promise of open data to benefit historically marginalized populations. However, data openness can also reproduce inequalities and facilitate forms of reuse that data creators, repositories, or other stakeholders may regard as data misuse. For instance, those with the time, resources, and expertise to use open data are often best positioned to harness them, which can reinforce existing power structures rather than disrupt them (Mirowski, 2018). In Mirowski's words, "for the 'widest possible dissemination,' read 'the most readily monetizable free resource'" (Mirowski, 2018, p. 178). A commonly discussed example is the digitization of land records in Karnataka, India (Donovan, 2012; Gurstein, 2011). Although framed as a "pro-poor" initiative, the project ultimately disenfranchised the poor by allowing those with financial resources and technical skills to access previously inaccessible data and re-appropriate land. The philosopher Leonelli proposes that the very appeal of open data "lies precisely in the impossibility to predict and quantify their potential as evidence in advance" (Leonelli, 2013b, p. 9). We cannot fully predict how data will be reused, and they may be reused

in ways that data creators, data repositories, or data intermediaries neither anticipated nor endorse.

Open data advocates often celebrate the broad reusability of data and the diversity of potential reusers, but they face a challenge when those reusers are actors whose intentions or conclusions they find politically, ethically, or scientifically troubling. This conundrum surfaces a thorny problem that the open data movement has not resolved: when data are released publicly, how much accountability do data providers bear for downstream reuses they did not anticipate and may not endorse? This dissertation does not resolve that problem, if such resolution is even possible. Instead, it provides an account of how anthropogenic climate crisis (ACC) skeptics reuse open data and what their reuse produces. Understanding what is done with open data is a prerequisite for designing open data infrastructures capable of responding more wisely to contested reuse — reuse that some stakeholders regard as legitimate and others as illegitimate.

Contested reuse is often labeled as data misuse by the communities whose data are being reused. What one community designates as misuse, another may consider legitimate critique, standard practice, or valid reanalysis. Data misuse is not an objective category; it is a designation applied by specific actors, from specific positions, with specific stakes. Because data misuse is situated in this way, the analysis developed here does not adjudicate which reuses are misuse. It treats accusations of data misuse as sites of infrastructural friction, signals that the values built into open data systems diverge from the practices of the communities that use them.

This dissertation, therefore, documents the specific infrastructural conditions that produce misuse accusations, and what those accusations reveal about the systems in which they occur. None of this means repositories can remain neutral. Curatorial decisions about what data to provide, how to document them, and how to respond when reuse is contested are themselves policy choices with consequences for public trust in data and public perceptions of scholarship. This dissertation takes the position that repositories must make those decisions, and that making them wisely requires a more precise account of the dynamics involved than currently exists.

Understanding the dynamics of contested data reuse is increasingly urgent. Open data mandates have expanded across research and government contexts (Holdren, 2013; Obama, 2013; Office of The Director, National Institutes of Health, 2023; Ryan, 2019; White House Office of Science and Technology Policy (OSTP), 2022), placing ever more data in the hands of more varied reusers. Data repositories that house these data, and data intermediaries that assist

data reusers, need to understand reusers' needs in order to make informed design and data curation decisions (Borgman, 2015; Faniel et al., 2013, 2020; Faniel & Yakel, 2017; Kansa, 2012; Weber et al., 2013).

Designing for and supporting one group of users is difficult enough. Repositories and intermediaries aiming to support broader communities have even more users to consider. Understanding those users' data practices, defined as the "ways in which people collect, make sense of and engage" with data, is an essential first step to designing effective data services (Lupton, 2015, p. 336). This dissertation investigates and analyzes the data practices of anthropogenic climate crisis (ACC) skeptics, drawing on a digital ethnography comprising participant observation, content analysis, and semi-structured interviews, built from a broad corpus of community activity — blogs, social media, comment sections, and community tools — as well as in-depth analysis of four data reuse journeys conducted by community members. This data community comprises individuals who are skeptical of mainstream claims about the existence, severity, interpretation, or political implications of anthropogenic climate change, and who invest significant time and effort in reusing open climate data to investigate those views.

ACC skeptics are not only consumers of open data from mainstream climate science knowledge infrastructure (KI). They build their own. They develop tools, maintain repositories, train new members, and produce and circulate knowledge claims through a network of blogs, conferences, journals, and social media platforms. This KI is organized around an audit orientation: a stance that treats the official climate science infrastructure not as a neutral resource but as an object of systematic scrutiny and contestation. Knowledge infrastructures are "robust networks of people, artifacts, and institutions that generate, share, and maintain specific knowledge about the human and natural worlds" (Edwards, 2010, p. 17). The KI ACC skeptics build is not separate from the one they contest; it is assembled from many of the same open data, code, tools, and norms, and produces a symbiotic relationship — by turns parasitic, competitive, and mutualistic — with the "vast machine" of climate science that Edwards describes. The alternative KI depends on the mainstream KI even as it disputes it. ACC skeptics are active, technically capable users whose data practices raise fundamental questions about how open data systems are designed, governed, and sustained. Understanding their data practices and the knowledge infrastructures they depend on and build is essential for repositories and

intermediaries seeking to design for the full range of communities that reuse their data. I am guided by the following research questions:

- What are the data practices of ACC skeptics?
- How does this community develop trust or distrust in the data they reuse?
- What are the knowledge infrastructures that support data reuse within this community, and how do they operate?

Together, these questions advance information science research by examining the data practices and knowledge infrastructures of a community that is actively reusing open data, but whose research goals often place it in conflict with the scientific community whose data it reuses.

1.2 Study Contributions

This dissertation advances understanding of how diverse communities engage with open data, through an in-depth empirical and theoretical account of ACC skeptics' data practices and the knowledge infrastructure they build and depend on.

Theoretically, this dissertation contributes to science and technology studies (STS) and information science by documenting how this community interacts with knowledge infrastructures and constructs its own. Edwards observed that climate skeptics reframe their engagement with climate science as auditing it, a version of the infrastructural inversion he documented within climate science itself (Edwards, 2010, 2019). Infrastructural inversion is the work of foregrounding the normally background infrastructure (Bowker & Star, 1999). This dissertation's central theoretical claim is that this audit orientation organizes the entire knowledge infrastructure the community builds. By audit orientation, I mean the community's characteristic practice of treating official climate science knowledge infrastructure not as a neutral resource, but as an object of systematic scrutiny and contestation. This stance recurs across the chapters that follow.

This audit orientation rests on a distinctive relationship to normative scientific values. The dissertation argues that ACC skeptics do not simply reject the epistemic values of mainstream science but often embrace intensified versions of them, pushing transparency, objectivity, and interoperability — values mainstream science itself holds — to points where they begin to turn against mainstream science. Following Biagioli and Pottage (2022), this dissertation shows how the demand for total transparency can produce dark transparency, where

the maximal openness yields paralysis rather than accountability. The dissertation frames this as a consequence rather than an attributed strategy: the same effects can follow whether a data reuser's commitment to these values is strategic or sincere. The contribution is to show that contested data reuse can arise not from a deficit of scientific values but from their intensification.

The dissertation also theorizes the relationship between the ACC skeptic knowledge infrastructure and the mainstream climate science knowledge infrastructure it depends on and contests. Rather than characterizing this relationship as simply parasitic, the dissertation frames it as a symbiosis in the ecological sense: at times parasitic, at times competitive, and occasionally mutualistic, with mimicry running throughout. No single characterization captures the full range of dynamics documented here.

A further theoretical contribution concerns data misuse. Building on prior collaborative work, this dissertation moves beyond our harm-centered definition of misuse toward a framework that treats contested reuse as infrastructural friction. It treats misuse accusations, or contested reuse, as objects of analysis rather than verdicts, attending to accusations that run in both directions: skeptics charge mainstream scientists and repositories with manipulating data, while scientists characterize skeptic reuse as cherry-picking and misrepresentation. This reframing has concrete implications for how repositories and intermediaries design for and respond to contested reuse. This move does not discard harm but situates it: a harm-centered account of data misuse holds only from a situated and explicit position.

These theoretical contributions have direct practical implications. Open data guidelines and repository design still largely cater to expert communities within mainstream science, with citizen science, data literacy initiatives, and educational programs as the main exceptions. This dissertation provides repositories, intermediaries, and information professionals with an empirically grounded account of one reuser community that falls outside those categories: how ACC skeptics reuse data; where they develop distrust in repositories and why; and what their usability concerns reveal about the limits of existing infrastructure design. These findings inform more responsive strategies for documentation, governance, and engagement — strategies that benefit not only this community but also the broader range of reusers that repositories increasingly serve.

More broadly, this dissertation uses this case to make the argument, by negative example, for a more reflexive open science. Drawing on Leonelli's philosophy of open science (2023), the

dissertation treats openness not as the mere availability of data but as a relational practice that depends on trust, documentation, and the capacity of reusers to take data up meaningfully, and it argues that the benefits of openness are never automatic or evenly distributed. The skeptic community shows what happens when data travel without the trust and relationships that give them meaning. This case makes concrete the importance of working towards an open science that acknowledges the full range of reusers and is explicit about whom its infrastructures privilege, rather than presuming a scientific norm and treating openness as good for everyone by default.

Ultimately, the relationship between ACC skeptic data practices and the mainstream climate science KI they engage with is neither purely harmful nor purely benign. Designing open data systems capable of responding to that complexity wisely begins with understanding what that complexity is, which is what this dissertation provides.

The dissertation proceeds as follows. Chapter 2 develops the theoretical framework, bringing together five concepts, three objects of study and two analytic lenses, to explain ACC skeptic data reuse: data communities, data practices, and knowledge infrastructures, examined through data journeys and accusations of data misuse. Chapter 3 describes and defends the digital ethnographic method, which combines content analysis, participant observation, and semi-structured interviews. Chapter 4 introduces readers to the ACC skeptic data community through four data reusers, following each through a specific data reuse journey. Together, the journeys establish the empirical ground for the chapters that follow. Chapter 5 broadens the description of data practices from the individual to the community level. Chapter 6 turns from data practices to the knowledge infrastructures that support them, arguing that ACC skeptics have built a functioning alternative KI designed to facilitate infrastructural inversion. Chapter 7 develops both theoretical and practical implications, asserting that skeptics are building an audit-oriented KI that is entangled with the mainstream climate science KI, relies on intensified versions of mainstream scientific values, reframes accusations of data misuse as sites of infrastructural friction, and that these dynamics carry concrete implications for how open data practitioners design, govern, and sustain open data systems. Chapter 8 concludes the dissertation.

Taken together, the chapters build a single argument: ACC skeptics are not anomalous misusers of open data, but rather skilled and rhetorically effective data reusers and infrastructure builders whose audit-oriented practices and alternative KI are a predictable product of openness

itself, and that taking their reuse seriously is required for designing effective and responsive open data systems.

Chapter 2 Theoretical Frameworks: Tracing Data Reuse Practices in Communities and the Knowledge Infrastructures They Rely Upon Through Data Journeys

This chapter develops the theoretical framework guiding my dissertation. It has three objects of study and two analytic lenses. The objects of study are data communities, the data practices that occur within them, and the knowledge infrastructures those practices depend on. Data communities are collectives that form around shared data needs, methods, and uses rather than disciplinary boundaries. ACC skeptics are a data community because they are drawn from a variety of disciplines like engineering, and economics and are held together by similar data needs. Within these communities are data practices, the concrete things people do with data, such as seeking, selecting, processing, analyzing, managing, and sharing data. The emphasis falls on what ACC skeptics do with data. These data practices are supported by knowledge infrastructures, or the robust networks of people, artifacts, and institutions that generate, share, and maintain knowledge about the human and natural worlds" (Edwards, 2010, p. 17).

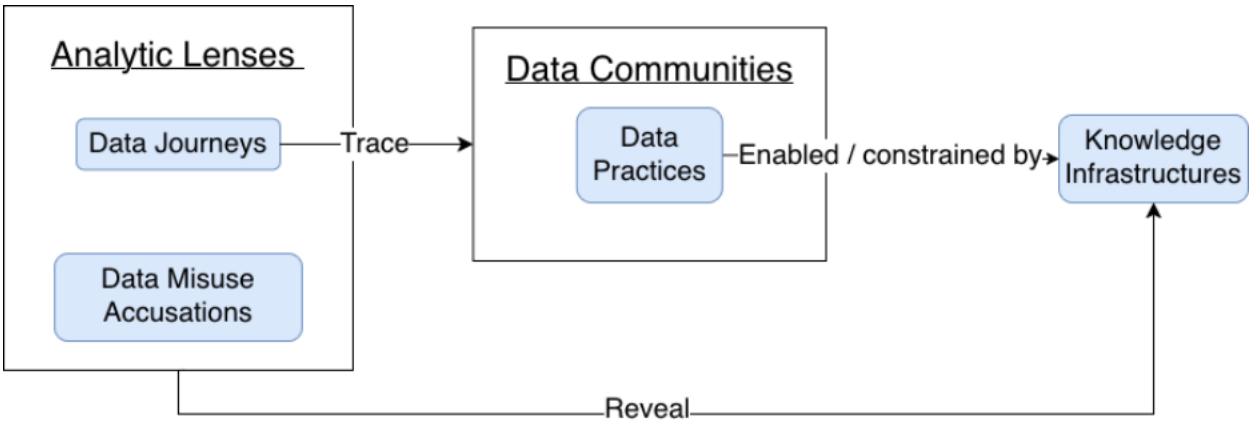
To study these communities, their data practices, and the knowledge infrastructures, I employ two analytic lenses. The first is data journeys, which highlight the trajectories, reuses, and reinterpretations of data across infrastructures and communities. Data journeys are closely tied to infrastructural inversion, an analytic strategy that brings to light the often-invisible sociotechnical systems, standards, and routines that underpin data work. The second are misuse accusations, which I treat as moments of infrastructural friction, and natural inversion where power, values, and epistemic authority collide and are revealed as data travel across communities, Figure 2.1 shows how these objects of study and analytic lenses relate.

Taken together, these frameworks situate practices not only in terms of what actors do but also in the social groupings and infrastructural arrangements that make those practices possible. A large body of work shows that data practices dramatically shape the knowledge communities produce and disseminate (Edwards et al., 2011; Karasti & Baker, 2008; Stvilia et al., 2013; Thomer & Wickett, 2020; Vertesi & Dourish, 2011; Zimmerman, 2008), in essence creating situated knowledges limited to specific contexts (Haraway, 1988). Knowledge

infrastructures afford these data practices. KIs include the technical structures as well as social and disciplinary contexts in which the practices are developed, adapted, shared, and policed.

In this chapter, I begin with data communities, defining the concept and situating ACC skeptics within it. I then examine data practices from a practice-theoretical perspective, reviewing the high-level activities that constitute data reuse. A central focus is how data reusers develop trust in data they did not collect themselves. I then address data misuse, examining its forms, its role as a driver of infrastructural adaptation, and the analytical stance this dissertation takes toward it. I next turn to knowledge infrastructures, with attention to both the climate science knowledge infrastructure and the open data infrastructure on which ACC skeptics and climate scientists alike depend. Finally, I introduce data journeys as the method through which I trace how data and their meanings travel across these communities and infrastructures. Using these frameworks, I address my research questions regarding the data practices of ACC skeptics, how they develop trust, and the knowledge infrastructures on which they depend.

Figure 2.1: Theoretical framework. Data journeys and data misuse accusations are used to study the data practices of data communities, and the knowledge infrastructures on which they rely.



2.1 Data Communities

In discussions of data reuse, the term "communities" is often used to refer to broad disciplinary domains (Borgman, 2015). I follow the conceptualization of data communities as explained by Gregory et al. (2023). Rather than centering on disciplines, they assert that communities "form around shared data, common data needs, shared methodologies, or common data uses" (K. M. Gregory et al., 2023, p. 195). An individual may belong to several data communities. This is often the case with ACC skeptics, who frequently have very different disciplinary backgrounds but share the same data, uses, and needs. Gregory et al.'s

conceptualization draws on foundational STS literature by Karin Knorr-Cetina, whose work on epistemic cultures characterizes communities as fluid collectives formed through shared practices, knowledge standards, and objects of inquiry (Knorr-Cetina, 1999). Rather than being defined by disciplinary boundaries like much reuse literature, community membership is dynamic, with individuals often participating in multiple communities simultaneously. These communities also actively shape individual behaviors. To truly understand these communities, it is essential to examine their real-life practices.

Knowledge is held collectively, not individually (House, 2004; Knorr-Cetina, 1999). For a data community, this means an individual's data behavior is legible only in light of the community they belong to. One person's choice reflects shared standards, methods, and ways of knowing. Observing individuals' data behaviors online can reveal broader community data practices. The notion of data communities helps us understand the collective nature of data engagement, practice theory allows us to analyze the specific activities and routines that produce and circulate knowledge within these communities.

2.1.1 The ACC Skeptic Data Community

This dissertation focuses on the ACC skeptic data community, which arose in reaction to growing scientific concern around climate change. Through the 1980s, scientific concern about human-caused warming grew. In 1988, NASA scientist James Hansen's congressional testimony brought anthropogenic climate change to national attention, and the Intergovernmental Panel on Climate Change (IPCC) was established that same year (Dunlap & Jacques, 2013). Shortly thereafter, a counter-campaign developed. Industries, especially the fossil fuel industry, and conservative political interests converged into a coalition opposed to regulating CO_2 emissions (McCright & Dunlap, 2011b). By the late 1990s, fossil fuel companies began to play a less central role, with many publicly acknowledging the scientific consensus, though organized skepticism continued through think tanks, advocacy groups, and individual researchers operating at the intersection of science and politics (McCright & Dunlap, 2011b). This skepticism was modeled in part on strategies developed by the tobacco industry, which had spent decades

manufacturing uncertainty about the harms of smoking to delay regulation (Oreskes & Conway, 2011).[1]

Older (over 65), white, politically conservative men are more likely than other groups to express ACC skepticism (Bolsen & Druckman, 2018; McCright & Dunlap, 2011a; Whitmarsh, 2011). Organized skepticism has been particularly concentrated and visible in the anglosphere including America, England, Canada, and Australia (Dunlap & Jacques, 2013; Hornsey & Lewandowsky, 2022; Painter & Ashe, 2012). For these individuals, higher levels of education or self-reported understanding can actually increase polarization, as they use their knowledge to better counterargue against scientific findings that conflict with their identity (A. Leiserowitz & Smith, 2010; McCright & Dunlap, 2011b). ACC skeptics hold a range of beliefs, from denying that climate change is occurring at all to accepting that anthropogenic climate change is real but not a crisis. At the community level, these positions have not remained static. Scholars have documented a shift from "old denial", disputing that climate change is occurring at all toward "climate delayism," which accepts climate change as real but downplays its severity and proposed responses (Solomun et al., 2025).

One event marked a turning point in how climate science came to be publicly contested. In 2009, private emails from the Climatic Research Unit at the University of East Anglia were hacked and released online. Skeptics and media outlets highlighted selective phrases — most notably "trick" and "hide the decline" — to allege that scientists had manipulated data and suppressed dissenting research. The incident was dubbed Climategate in blogs and popular media. Nine independent investigations in the UK and US found no evidence of scientific misconduct (A. A. Leiserowitz et al., 2013). However, some of these reviewers did fault the scientists "for being insufficiently open and transparent in their dealings with contrarian requests for information, including Freedom of Information requests for raw data used to estimate changes in global mean surface temperature over land" (Parker, 2020, p. 191). The public effects were significant. Trust in climate scientists declined, particularly among politically conservative audiences (A. A. Leiserowitz et al., 2013), and attention shifted from scientific findings to the conduct and transparency of scientists themselves. The scientific community and ACC skeptics

[1] Portions of this paragraph are adapted from Pasquetto et al (2026), where I was the primary author of the ACC skeptic case study.

interpret the event differently to this day; researchers describe it as a PR disaster in which private emails were selectively quoted out of context, while ACC skeptics frame it as a genuine expose (Biagioli & Pottage, 2022; La et al., 2024; A. A. Leiserowitz et al., 2013; Whitmarsh, 2011).

Climategate legitimized and strengthened a particular mode of engagement with the mainstream climate science KI: the scrutiny of data, methods, and scientific communication. In the years that followed, organized skeptic communities increasingly framed themselves not as rejecting science, but as auditing it, subjecting it to independent verification (Edwards, 2019). They downloaded datasets. They reanalyzed temperature records. They photographed weather stations. They demanded raw code and metadata. In other words, they engaged directly with the same open data infrastructures that mainstream scientists use. The skeptic community found its primary home in the blogosphere, where sites function as alternative sites of expertise, translating and reinterpreting mainstream climate findings for skeptical audiences (Lewandowsky, Mann, et al., 2016; Matthews, 2015; Sharman, 2014). For instance, one blog is literally called *Climate Audit*. Its author treats auditing as a step above peer review, which he sees as inherently flawed, or in his own words "pal review" (Edwards, 2019). It is these data reuse practices, the downloading, reanalyzing, and reinterpreting of publicly available climate data, that this dissertation examines.

2.2 Examining data reuse practices: A practice-theoretical perspective

If a data community is where knowledge is produced, then understanding one means examining what its members actually do with data. Practice theory offers a way out of a persistent problem: whether to explain behavior by individual choice or social structure (Shove et al., 2012). Developed in the 1970s, it treats routine action as embedded in collective practices shaped by both the individual and society (Bourdieu, 2008; Giddens, 1979; Shove et al., 2012).

Interpretive qualitative researchers took it up widely (Palmer & Cragin, 2008; Reckwitz, 2002; Schatzki, 2005; Shove et al., 2012), and continues to be adopted in fields such as HCI, and information science (Chanen & Soden, 2025; Chao et al., 2015; Crooks & Currie, 2021; Karasti & Baker, 2008; Lee et al., 2021; Soden et al., 2025).

Practices, broadly defined by the sociologist Schatzki, are "arrays of human activity" that are "materially mediated" and "organized around shared practical understanding" (Schatzki, 2005, p. 2 as cited in Palmer & Cragin, 2008). A practice-theoretical framework treats these patterned activities as the primary unit of analysis. Rather than focusing solely on individual

behaviors or abstract norms, this approach examines how actions, tools, and shared meanings converge in specific contexts. In this sense, agency is distributed: outcomes emerge from the interaction of human actors and material artifacts within specific socio-technical arrangements (Lämmerhirt et al., 2024). In research, applying a practice framework involves studying both what people do and how their activities are shaped by artifacts and infrastructures. For instance, in HCI, Kuutti and Bannon promote a practice-led framework as opposed to a focus solely on the interaction between human and computer, where the spatial and temporal contexts of actors can also be taken into account (Kuutti & Bannon, 2014). This orientation makes it particularly well-suited to analyzing data work, where people, technologies, society, and knowledge claims are tightly intertwined.

Within this orientation, data practices refer to the routine activities "around data that generate meaning through action[s]" associated with different actors in specific contexts (Lämmerhirt et al., 2024, pp. e24-3). Data practices encompass a broad range of activities involved in the identification or creation, management, curation, preservation, and use of data. Data are different from other artifacts, as they are regarded as representations of the physical world and evidence to support research claims (Latour, 1987). As Borgman (2015, p. 28) notes, entities "become data only when someone uses them as evidence of a phenomenon, and the same entities can be evidence of multiple phenomena."

This practices framework naturally positions itself against information determinism, or the belief that "information by itself, and in the abstract, can bring about social change" as people become more informed about social issues (Srinivasan et al., 2017 as cited in Soden et al., 2025;). Information determinism attributes the power and agency of social actors to information; in this case, the data, by themselves, cut off from the rest of their socio-technical underpinnings (Soden et al., 2025). Kuuttii and Bannon describe practices as "situated in time and space, and…dependent on many features of the surrounding material and cultural environment" (Kuutti & Bannon, 2014, p. 3543). Soden and colleagues extend this to data themselves, noting that a practice approach draws attention to the ways data are similarly situated (Soden et al., 2025). Because data are both highly mobile and context-dependent, data practices provide a productive lens for examining how knowledge is produced and interpreted across communities.

2.2.1 Data practices are heterogeneous and situated

Recognizing the diversity in practices across data communities is critical. Generalized models of the data or research lifecycle (Ball, 2012), often assume a one-size-fits-all approach, but such models fail to capture the variation in how different groups actually engage with data (Chao et al., 2015). Data practices are always situated in particular communities and shaped by their socio-technical arrangements, including disciplinary norms, organizational structures, political contexts, and available technologies (Palmer & Cragin, 2008). Tailoring analyses and infrastructures to these differences is essential; solutions designed to serve all users uniformly often fall short because practices vary so widely.

Much of what is known about data practices comes from studies of traditional research communities, such as academic scientists, government researchers, industry data scientists, and domain specialists (Burkhardt et al., 2022; Chao et al., 2015; Fiske et al., 2023; Lämmerhirt et al., 2024; Leonelli, 2020; Palmer & Cragin, 2008; Wallis et al., 2008). These studies reveal disciplinary variation in how data are generated, documented, interpreted, and reused. For instance, detailed models have been developed for data-intensive work in the Earth and environmental sciences (Faundeen et al., 2014), social sciences (Green & Gutmann, 2007), physical sciences (I2S2 Project, 2011), and habitat ecology (Borgman et al., 2007; Wallis et al., 2008).

More recent research has extended this attention to communities outside conventional academic, industry, or governmental settings. Studies of activists, citizen scientists, community organizations, journalists, and other public-interest groups demonstrate that data practices emerge within distinct sociopolitical contexts and orient toward different goals, values, and epistemic commitments (Bowser et al., 2017, 2020; Bui, 2023; Choudry, 2013, 2014; Crooks, 2019; Gray, 2018; K. M. Gregory et al., 2024; Gutiérrez & Milan, 2019; Hinojosa et al., 2021; Lagoze, 2014; Muravyov, 2024; O'Grady & Mangina, 2024; Pei et al., 2022; Ruckenstein, 2019; Tran et al., 2022; Van Tuyl & Whitmire, 2018). This work underscores that data practices cannot be assumed to conform to academic norms, nor should they be evaluated solely by academic criteria.

Across these literatures, scholars have called for more granular descriptions of the activities that data enable (Burkhardt et al., 2022; Christin, 2020; Fiske et al., 2023; Leonelli, 2020; Wallis et al., 2008). Despite variations across groups, some common practices are observed, providing a cohesive lens through which to view and analyze data practices. Usefully,

Chao, Cragin, and Palmer suggest that high-level, more universal data practices are tied to a series of more granular actions in their taxonomy, the Data Practices and Curation Vocabulary (DPCVocab) (Chao et al., 2015). Rather than imposing a universal lifecycle, DPCVocab identifies high-level categories of data practice and ties them to specific, empirically observed sub-activities. This functional vocabulary was validated through studies of scientific data work in volcanology, oceanographic and climate modeling, and soil ecology, domains whose data infrastructures and reuse challenges closely resemble those relevant to climate science. While this dissertation does not adopt the DPCVocab as a normative model, it is methodologically useful for analyzing data reuse practices and for generating actionable recommendations for data repositories and intermediaries. Accordingly, Chapter 3 explains how I operationalize elements of this vocabulary in my analysis of ACC skeptics' data reuse.

2.2.2 Applying a practice-theoretical perspective to ACC skeptics' data reuse

For this dissertation, adopting a practice-theoretical perspective provides a lens for examining how ACC skeptics reuse open data. I assume data work is constituted through situated, materially and socially mediated activities rather than following a universal lifecycle. This framework suggests that I should expect to find data reuse practices among ACC skeptics that diverge from generalized models, shaped by their unique epistemic commitments, social networks, and political goals. Building on this practice-theoretical foundation, I examine the concrete activities that comprise data reuse, highlighting the stages and tasks most relevant to Earth and climate science data.

This theoretical orientation foregrounds data reuse as a situated activity whose forms and meanings are contingent on the infrastructures and communities in which it unfolds. Data practices emerge in response to practical constraints, disciplinary norms, political demands, and more. This is especially true for Earth and climate science, where the complexity of the data, including a variety of observational systems, historical records, multi-sensor integration, calibration chains, and data assimilation, makes reuse highly contingent, as discussed more in chapter 4 (Edwards, 2010, 2011). Recent work in human-computer interaction and climate justice further emphasizes that climate data practices are not merely scientific but also politically and ethically charged (Chanen & Soden, 2025; Soden et al., 2025). Next, I review the specific high-level activities that can constitute data reuse and the ways these practices vary across both

academic and non-academic settings, including Earth and climate science contexts most relevant to ACC skeptics.

2.3 Situated activities in data reuse practices

Data reuse is a specific overarching data practice that involves the "usage of a dataset by someone other than the data creator" (Pasquetto et al., 2017, p. 3). While definitions of data reuse vary across the literature, I adopt this definition throughout the dissertation for applicability for ACC skeptics. Understanding data reuse requires examining the practices and the more granular situated activities through which people interact with data. To describe the data practices I observed among ACC skeptics, I draw primarily on Chao et al.'s Data Practices and Curation Vocabulary (DPCVocab) (Chao et al., 2015), a taxonomy of high-level data practices and the granular activities that constitute them. Unlike stage-based lifecycle models, the DPCVocab does not presuppose a fixed sequence of activities. This flexibility is particularly useful for describing ACC skeptic data practices, which rarely follow a linear path. This lack of standardized linearity in data practices is not unique to ACC skeptics, data reuse is rarely seamless or linear (Burkhardt et al., 2022; Christin, 2020; Fiske et al., 2023; Leonelli, 2020; Pasquetto et al., 2019; Wallis et al., 2008; X. Wang et al., 2021). Where DPCVocab terms did not fully capture what I observed, I supplemented them with terminology from the USGS Data Lifecycle Model (Faundeen et al., 2014). Importantly, the categories described below first emerged inductively from the ACC skeptic data practices I observed. I subsequently named and organized them using existing scholarly vocabulary to facilitate comparison with other data communities. The order in which these practices are presented below reflects neither a prescribed sequence nor a typical workflow. In practice, ACC skeptics move among them in varying order. Additionally, although I use the term data reuser in the dissertation, I do so with care, as members of these data communities enact multiple roles and inhabit multiple contexts in which technologies are present.

Below, I review the high-level practices relevant to data reuse: research idea and design; seeking and selecting data, which includes assessing data and repository trustworthiness; processing data; analyzing, interpreting, and abstracting data; managing, documenting, and preserving data; attributing and citing data; publishing and sharing data; and publishing and sharing research.

2.3.1 Data reuse prompt and design

Data reuse prompt and project design refers to the factors that led to the data reuser beginning the project as well as the initial framing of the project. It includes activities like identifying a problem, generating research questions and hypotheses, planning sampling strategies, and identifying what data will be needed (Stvilia et al., 2013; Wallis et al., 2008). Topics chosen can be prompted a variety of factors, including real-world problems, community needs, disciplinary agendas, existing datasets, unsolved scientific questions, influential texts, policy disputes, gaps in available evidence, and funding pressures (Chao et al., 2015; Choudry, 2013, 2014; Karasti & Baker, 2008; Palmer et al., 2009; Zimmerman, 2003).

The initial design and framing of a project also include the kind of reuse it involves. Several studies have identified distinct types of data reuse that reflect differing goals (Coady et al., 2017; Federer, 2019; K. Gregory et al., 2020; Pasquetto et al., 2019; Thomer et al., 2023; Wofford & Thomer, 2023). Pasquetto et al. (2019) distinguish comparative data reuse, where data are reused for ground-truthing, calibration, comparison, and confirmation, from integrative data reuse, where researchers repurpose data to answer new questions, perform meta-analyses, or develop new statistical models. Rather than two discrete categories, comparative and integrative reuse can be understood as ends of a continuum defined by analytical depth and the expertise a reuse demands, with comparative reuse tending toward the more routine and integrative reuse toward deeper, often contributory, engagement with the data. Reusing data for reproducibility tests, where researchers use the same data and analysis methods to verify findings, is another prominent type of reuse (Borgman, 2015), sitting toward the integrative end of this continuum in the depth it requires. Other typologies describe reuse aimed at producing summaries, visualizations, or analysis tools (K. Gregory et al., 2020; Thomer et al., 2023).

2.3.2 Seeking and selecting data

The next category of data reuse practices involves seeking and selecting data for reuse, which encompasses the activities through which reusers locate, evaluate, and retrieve datasets for their purposes (Chao et al., 2015)). Data seeking, often called data discovery, is shaped by available search tools, social networks, disciplinary norms, and search literacy (K. Gregory et al., 2020). Researchers often encounter data through intermediary sources, such as literature, colleagues, or community forums, before searching for it directly (K. Gregory et al., 2020).

Selecting data entails evaluating its quality and determining its suitability for addressing the research question. Earthquake engineers assess completeness and documentation detail (Faniel & Jacobsen, 2010); ecologists weigh original collection methods and provider reputation (Zimmerman, 2008); and Earth science modelers rely heavily on communication with colleagues and repository reputation (Chao et al., 2015; Parsons, 2011). A central sub-activity here is assessing data and repository trustworthiness.

2.3.2.1 Assessing data and repository trustworthiness

Seeking and selecting data for reuse requires developing trust in the data and in the repositories that store them. Trust is a foundational factor for data reusers as they evaluate the credibility, authenticity, and reliability of data, data creators, and repositories. Unlike the institutional focus on certification mechanisms, such as the CoreTrustSeal (2024), data reusers often rely on social and contextual factors when determining whether to trust a dataset (Yakel et al., 2024). Trust is not simply bestowed upon a repository or dataset because of its formal accreditation; rather, it is actively negotiated by reusers through factors such as their lived experiences, disciplinary norms, perceptions of data quality, and other considerations (Faniel & Yakel, 2017).

One key factor in trust development is documentation and provenance information, which provides insight into how data were collected, processed, and curated (Zimmerman, 2008). Reusers are more likely to trust datasets that include clear metadata detailing who created the data, how it has been transformed, and any limitations in its scope (Faniel et al., 2012). Another crucial factor is reputation and credibility - both of the repository and the data producers (Van House et al., 1998; Zimmerman, 2008). Researchers tend to trust data created by known scholars, institutions, or projects with established track records. This aligns with findings that trust in data is often socially constructed through preexisting disciplinary networks rather than derived from repository certification alone (Yoon, 2017). Trust is also built through interaction with repository staff and data curators. Reusers who directly dialogue with repository personnel, whether through consultations, data user forums, or research collaborations, often express greater confidence in the data's reliability (Gabrielsen, 2020).

Distrust in data or repositories can emerge when inconsistencies, missing context, or past negative experiences are present. One common source of distrust is unclear provenance and a lack of transparency, in which reusers struggle to verify how the data were generated or whether

they have been altered (MacNeil, 2001). If a dataset lacks sufficient contextual information or has ambiguous methodological details, reusers may question its reliability (Zimmerman, 2008).

Data reusers must also evaluate whether the repository aligns with their epistemic frameworks. Activists, for instance, may distrust institutional repositories due to past experiences of data erasure, exclusion, or bias (Findlay, 2013; McKemmish et al., 2011). In some cases, trust in data is not necessary for reuse. Instead, there is epistemological ambiguity, where reusers simultaneously use a dataset while questioning its credibility (Muravyov, 2024). This is evident in activist data practices, where engagement with data is critical and strategic. Activists may correct errors, recontextualize information, and use data infrastructures in subversive ways while still relying on the data for advocacy (Bui, 2023). Chanen and Soden (2025) show that municipal climate governance actors often distrust the very data they are mandated to use, because those data fail to capture on-the-ground realities. Trust, therefore, is not simply a matter of accreditation mechanisms such as CoreTrustSeal; it arises from lived experience, community narratives, and epistemic commitments.

2.3.3 Processing and transforming data

Processing and transforming data includes activities through which reusers prepare data for analysis (Chao et al., 2015). This encompasses preprocessing, cleaning, normalizing, aggregating, and converting data into formats suitable for further work (Chao et al., 2015). These practices vary significantly across disciplines and are shaped by researchers' methodological training and the social norms of their fields (Leahey et al., 2003; Van House et al., 1998). For example, data editing requires an understanding of acceptable practices within specific disciplines or work sites (Van House et al., 1998). Data editing, including decisions about what to alter, retain, or discard, requires shared agreement about what constitutes acceptable practice within a given community or context (Palmer & Cragin, 2008; Star, 1983). Processing is therefore not merely a technical activity but an epistemologically loaded one, in which decisions about transformation reflect underlying assumptions about data quality and validity.

2.3.4 Analyzing, interpreting, and abstracting data

Analyzing, interpreting, and abstracting data are critical stages in data reuse practices, encompassing activities such as categorizing, classifying, comparing, annotating, searching for patterns, visualizing, data mining, and developing software and scripts (Chao et al., 2015). These

activities are foundational for transforming raw data into meaningful insights. Although the DPCVocab groups the three together, they name distinct kinds of work. Analyzing produces results from data, through activities such as comparing one dataset against another or against a model, searching for trends and correlations, and mining collections for structure not apparent on inspection. Interpreting assigns meaning to those results, reading them considering prior findings, theory, or the questions that prompted the reuse. Abstracting condenses data and results into more compact and communicable forms such as summaries, models, and visualizations, which then stand in for the underlying records in later reasoning and reporting. What a reuser searches for, and what they take a result to mean, is rarely given by the data alone. It is shaped by the purpose the reuse serves and the commitments the reuser brings to it, so that the same dataset can support a variety of findings (Leonelli, 2020).

2.3.5 Managing, documenting, and preserving data

Managing, documenting, and preserving data draws on both the DPCVocab's managing and representing categories (Chao et al., 2015) and the USGS Lifecycle Model's preserve stage (Faundeen et al., 2014). It encompasses organizing, labeling, annotating, generating metadata, and storing data in ways that support ongoing use and potential sharing. Documentation quality and granularity directly determine whether data can be meaningfully reused (Chao et al., 2015). When reusers archive data separately from its original repository, whether locally or publicly, decisions about what provenance information, processing history, and supplementary documentation to include or omit shape how the data can be interpreted and used by others. The USGS Data Lifecycle Model additionally emphasizes the creation of accessible storage formats and supplementary preservation products as part of this practice (Faundeen et al., 2014).

2.3.6 Publishing and sharing research

Publishing and sharing of research refers to the dissemination of findings derived from data reuse (Faundeen et al., 2014). In traditional academic settings, this typically takes the form of peer-reviewed journal articles or conference presentations, where findings are evaluated by disciplinary experts. Within and outside academia, findings are frequently shared through blogs, newsletters, public reports, or social media, and validation often occurs informally through community networks rather than formal peer review (Choudry, 2013; Eschle & Maiguashca, 2010). While publishing and sharing data and publishing and sharing research often occur in

close proximity, they are analytically distinct practices: one concerns making data itself available; the other concerns communicating what was done with it and what was found.

2.3.7 Attributing and citing data

Attributing or citing data refers to how reusers acknowledge the sources of the data they use (Chao et al., 2015). This can take the form of formal citation with persistent identifiers in reference lists, informal acknowledgment in methods sections or footnotes, or inclusion of data producers as co-authors (Chao et al., 2015). With regard to formally acknowledging data, citing data with persistent identifiers (PIDs) in reference lists is considered the ideal practice, as it gives proper credit to data creators and facilitates discovery (Lowenberg et al., 2019; E. Moss & Lyle, 2018).

2.3.8 Publishing and sharing data

Publishing and sharing data draws on Chao et a.'s (Chao et al., 2015) data publishing practice as well as the USGS DLM share stage (Faundeen et al., 2014). It describes how reusers make datasets available to others, including depositing data in a repository, distributing it on request, sharing it through collaborative channels, or building public-facing tools that incorporate and display datasets. Data sharing is shaped by disciplinary cultures, funder and publisher mandates, and community norms (Holdren, 2013; Key Perspectives Ltd, 2010; Ryan, 2019; Tenopir, Palmer, et al., 2011). Data reusers still publish data they are archiving from another source on websites, libraries, or other platforms (Dillon et al., 2019; Eschle & Maiguashca, 2010; Vera et al., 2018, 2019). For instance, the Environmental Data Governance Initiative (EDGI), through its project Data Together in collaboration with qri.io and Protocol Labs, facilitated grassroots archiving efforts and community stewardship of vulnerable federal data during the first Trump administration (Vera et al., 2018).

In conclusion, although data practices are contextually distinct, shared high-level activities appear across many different domains. Understanding these common practices is essential for designing systems and policies that support effective data reuse across diverse communities. The next section turns to data misuse.

2.4 Data misuse[2]

Having described the principal activities involved in data reuse practices, I now address the potential risks and conflicts that emerge as data move across communities: data misuse. This section explores the difficulty in defining misuse, how this dissertation approaches it analytically, and how moments of misuse reveal underlying tensions within open knowledge infrastructures. Data misuse is a recurring concern in discussions of open data, yet it remains analytically unstable and context dependent. While open science initiatives emphasize the benefits of making data widely available, concerns about misuse have long impeded data sharing (Enke et al., 2012; Hemphill et al., 2022; Sayogo & Pardo, 2013). Scientists' fears broadly cluster around three concerns: being scooped or losing credit for data collection efforts; reusers making errors in secondary analysis due to poor data quality or complexity; and reusers running replications that reveal flaws in the original work, damaging scientists' reputations (Pasquetto et al., 2024). These concerns are pronounced when data travel beyond the communities and infrastructures in which they were originally produced. While openness is often framed as a good, it also exposes data to reuse by actors operating under different assumptions, values, and epistemologies.

At the same time, treating misuse solely as a problem to be prevented risks obscuring what it reveals. Data misuse should be understood not only as a risk, or even failure, of open data infrastructures, but as a site where power, values, and epistemic authority collide and become visible in an act of infrastructural inversion. What makes reuse into misuse, after all? Misuse draws attention to questions that infrastructures often struggle to surface explicitly: Who gets to interpret data? Whose epistemology defines "proper use"? And who bears the harm when interpretations diverge? Approached in this way, misuse becomes an important lens for understanding how infrastructures are negotiated, contested, and adapted over time.

2.4.1 Defining data misuse as a sociotechnical phenomenon

A useful starting point for this discussion is the data misuse typology developed in recent work by Pasquetto, Cullen, Thomer, and myself (Pasquetto et al., 2024). That work grounds misuse in a harm-centered definition; this dissertation borrows the typology but expands on the

[2] Portions of this section are adapted from Pasquetto et al. (2024). Because this work was collaborative and I was not the lead author, I identify passages drawn from it in this section.

harm-based definition, for reasons developed in Chapter 7. Drawing on a critical review of interdisciplinary literature, we emphasize the difficulty of defining misuse in universal terms. Misuse is fundamentally relational: what counts as misuse depends on who is reusing data, for what purpose, in which community, and according to which norms. Misuse is also historically contingent. Epistemological norms shift, and what was once standard practice may later be recognized as misuse, or the reverse. In the 1950s, for instance, there were no consent standards for biopsy, and it was routine for doctors to own tissue samples and patient records they collected; the same practice would now be understood as a violation of consent (Acker, 2015; Landecker, 2007 as cited in Pasquetto et al., 2024).

Across the literature, certain forms of data reuse have been identified as misuse: analytical errors, misinterpretation, misrepresentation, reputational harm, privacy and geoprivacy violations, exploitation, and the uncritical reuse of biased or offensive data, each described in Table 2.1 (Pasquetto et al., 2024). Even within these categories, what counts as an instance varies significantly across communities. What one group considers an analytical error, another may treat as a legitimate methodological choice; what one community calls misrepresentation, another may frame as valid reinterpretation (Pasquetto et al., 2024). These categories are not exhaustive or mutually exclusive. I mention these categories not as universal definitions of misuse but as a way of organizing the accusations and instances of contested reuse documented in the corpus. Where I do assign a label, it is in cases where ACC skeptic data practices appear to violate not only the epistemic standards of the mainstream climate science community but the community's own stated commitments.

Table 2.1: Ways of misusing research data identified in the literature on data misuse, and examples of such misuses, drawn directly from Pasquetto et al. (2024).

Misuse Types	**Definition**	**Examples**
Analytical errors	Incorrect or flawed methods used during data analysis, leading to inaccurate or unreliable results	The GDELT Project is a digital archive that collects worldwide data about media content. Media reporters working for a famous data driven news site conducted an analysis of a GDELT Project, and based on such analysis, reported that certain events in Africa were happening at a much faster rate than they were actually happening. This happened because the reporters miscounted instances of such events in the data. D'Ignazio and Klein attribute this problem not only to the inattention of the journalists, but also to the lack of contextual information that the GDELT archive provides about the data it hosts (D'Ignazio & Klein, 2020)
Misinterpretation	Misunderstanding the	In 2019, NASA released research data about wildfires in

	meaning or implications of research data, potentially leading to incorrect conclusions	Australia. An artist created a map of wildfires using such data. The map became viral as it was shared by celebrities with the intention of raising awareness about the scale and gravity of the wildfires. In these viral posts, celebrities implied that the map was representing live events of wildfires. As it turned out, the map was visualizing data for a month of wildfires, not live data. Some other users used this "misunderstanding" as evidence that the wildfire emergency was overblown on social media and to argue that, instead, wildfires were happening at a normal rate in Australia (Rannard, 2020)
Misrepresentation	Deliberate or unintentional distortion, alteration, or absence of data that may convey a false or incomplete impression or support a particular agenda	The 2013 Arctic Report Card showed that the Arctic sea ice increased from its record low in 2012 but remained the sixth lowest since 1979. Different media outlets framed this differently; Climate. gov highlighted it as larger than the 2012 low but still the sixth smallest, while FoxNews.com emphasized a 60% increase from the previous year. Such framing highlights how data can be misrepresented in ways that lead to misconceptions and fuel polarization over climate change (Hardy & Jamieson, 2017)
Reputational harm	Damaging the reputation of the original data collectors, analyzers, or curators through actions such as lack of data citation, negated or unwarranted authorship, or generating public shame for perceived flaws in data	This can take many forms. In one of our previous research studies, we encountered a climate change skeptic who explicitly named curators from the Centre for Research on Epidemiology of Disasters (CRED) EM-DAT International Disaster Database in a blog post. The skeptic accused curators of endorsing a United Nations World Meteorological Organization (UN WMO) report. The skeptic argued that the report, which indicates increasing disaster rates using EMDAT data, fails to consider changes in reporting rates. The skeptic presented correspondence with the curators as evidence that they acknowledged this flaw but endorsed the report anyway. The blog post, in turn, damages the reputation of both the curators and the repository (Wofford & Thomer, 2023)
Privacy and geoprivacy violations	Breaching the confidentiality of individuals or revealing sensitive location-based information, potentially causing harm or discomfort	In a research project, still in Australia, sharks were tagged by researchers for scientific purposes to gain insights into marine ecology, shark movement, behavior, and conservation efforts. A government agency identified a tagged shark as an imminent threat, leading to its targeted killing. This misuse of tagged data undermined the conservation and knowledge generation efforts of the research (Meeuwig et al., 2015)
Exploitation	Unethical or harmful use of research data for personal gain or advantage, often at the expense of the subjects of the data	Historian of science Joanna Radin analyzed the problematic reuse of genetic data from indigenous communities. The case involves the collection of blood samples from an indigenous community in the American Southwest for a diabetes study in the 1960s. In the 1980s, this data was added to a repository for training machine learning algorithms without obtaining a new informed consent from the tribe. The data was then reused to train ML algorithms for all sorts of purposes, including many commercial purposes. The indigenous tribe never received compensation nor signed an informed consent (Radin, 2017)
Uncritical use of	Failure to critically	In *Automating Inequality*, Eubanks discusses a case of uncritical

biased and offensive data	evaluate and challenge data sources that contain prejudices, offensive content, or discriminatory elements, perpetuating harmful stereotypes or narratives	use of biased data that resulted in great physical and emotional harm for a vulnerable population (low-income community). A local government office used an algorithmic model to predict the risk of child abuse, but the algorithm was penalizing low-income families as much more data existed on them compared to high and middle-income families. Little data exists for wealthier parents because they have easier access to private health care and mental health services, while low-income parents need to rely on public services which collect data on drug and alcohol treatment programs, child welfare services, jail records, mental health services, health insurance histories, etc. Thus, low-income families were oversampled in the model, and disproportionately targeted for intervention (Eubanks, 2018)

This dissertation does not treat the harm-centered definition as wrong so much as incomplete. A harm-centered account of misuse holds, but only from a situated and explicit position, because what one community identifies as harm another may not. Every claim of harm is made from some position, whether or not the claimant recognizes or acknowledges it. My condition falls not on the harm claimant but on the actor who adjudicates whether an instance of data reuse can be considered harmful and therefore, misuse. An actor deciding whether a given data reuse is misuse should name the standpoint from which it judges rather than present that judgement as neutral or universal. The requirement is not that communities share a definition of harm, but that whoever rules on harm makes explicit the position from which they rule.

Rather than adjudicate harm or misuse in general, the dissertation treats accusations of misuse analytically and reserves the label of misuse for the narrow cases specified below. I frame misuse as a site of *infrastructural friction*: a point where different epistemological communities come into conflict over what data are for and how they should be reused.

The term friction has a history in information science. Edwards describes data frictions the costs in time, energy, and attention required simply to collect, check, store, move, receive, and access data, costs incurred whenever data travel between places, machines, or media (Edwards, 2010, pp. 84–85). In social systems, he notes, that friction produces conflicts, disagreements, as well as delays. Edwards and colleagues extend this concept to "science friction," the difficulty two communities encounter when they work with the same data while holding different vocabularies, methods, and standards. Bates observes that this literature has largely treated friction as something to be overcome, and argues instead that friction is deeply political, produced by actors with unequal power to shape how data move, and in some cases

worth fostering rather than eliminating (Bates, 2017). I follow Bates in treating friction as political rather than merely obstructive, focusing on communities that may not be focused on working through the friction to establish common ground. Edwards et al. (2011) show an example of this in their discussion of Climategate. The authors observe that the released emails had been informal communication within a small community, but once public they were read as fixed statements about the data. Common ground could not be re-established. The two communities held incompatible standards for what counts as an adequate account of how data are made.

Viewing data misuse through the lens of friction reveals that misuse is not only a problem but a signal that the values embedded in open data systems diverge from the practices of the communities that use them. In this way, an accusation of misuse is a moment of infrastructural inversion, where the assumptions and standards that normally sit in the background of data reuse become visible. Chapter 7 develops this position more fully as a theoretical contribution in light of the study's findings.

2.4.2 Implications for this study

This dissertation does not attempt to adjudicate whether particular instances of data reuse constitute misuse in any technical or disciplinary sense. Rather than interrogating the scientific validity of climate claims, the analysis attends to how misuse is defined, contested, and operationalized within and across knowledge infrastructures, and what those processes reveal. Two analytic approaches follow from this stance.

First, claims of misuse are treated as objects of study. In the ACC skeptic corpus, these claims run in more than one direction: ACC skeptics routinely accuse mainstream climate scientists and repositories of manipulating, selectively reporting, or misrepresenting climate data, while mainstream scientists and data curators characterize many skeptic reuse practices as cherry-picking, misrepresentation, and reputational harm. Both sets of claims are treated here not as verdicts but as evidence of the infrastructural tensions they surface and the competing epistemic authorities they enact. Second, I identify instances of ACC skeptic data reuse that would be recognized as misuse both by the mainstream climate science community and by the community's own stated epistemic standards. Where those conditions are not met, the analysis attends to the accusation without resolving it. The coding procedure is described fully in Chapter 3.

Together, these approaches treat misuse not as an aberration of openness but as one of the most visible places where the tensions inherent in open infrastructures surface. Rather than asking how misuse can be eliminated, this dissertation asks what it reveals about the systems in which it occurs.

2.5 Understanding knowledge infrastructures: Foundations of data practices

As the preceding sections make clear, data reuse does not occur in isolation. Accordingly, I now discuss knowledge infrastructures, which provide the foundational arrangement for enabling, or constraining, the flow and interpretation of data. Variations in data practices across communities are rooted in different socio-technical networks that can include processes, disciplines, organizational and social structures, socioeconomic factors, organizational rules, discourses, artifacts, and other elements (Choudry, 2013, 2014; Lämmerhirt et al., 2024; Palmer & Cragin, 2008). Data practices, in turn, shape how these socio-technical networks function (Bowker & Star, 1999). These networks that enable data practices thus consist of dynamics interactions between human actors, technological artifacts, and the contexts in which they are embedded.

I consider these socio-technical networks knowledge infrastructures, "robust networks of people, artifacts, and institutions that generate, share, and maintain specific knowledge about the human and natural worlds" (Edwards, 2010, p. 17). Connecting these data practices and knowledge infrastructures aligns explicitly with previous research (Borgman et al., 2014, 2015, 2019; Borgman & Wofford, 2021), with many others implicitly connecting data practices with infrastructure in various ways. However, in some of these works, the meaning and theoretical underpinning of data practices is not clearly articulated. Within this dissertation I aim to articulate both clearly.

Edwards developed the knowledge infrastructure framework while studying the complex and distributed nature of knowledge production in climate science. He deploys the framework of knowledge infrastructures not simply to define an object of study but to make visible the sociotechnical conditions that render global climate knowledge possible. He used the framework to explain how localized data about the climate are integrated into global models, thereby managing the tension between local observations and global knowledge claims. In this way, climate science "becomes" a global knowledge infrastructure (vs. just a socio-technical system) (Star & Ruhleder, 1996). He conducted an historical analysis tracing how mundane data

practices, like collecting, calibrating, circulating, and reconciling data, gradually produced trust and coherence across scales. The framework operates as both an analytic and narrative strategy: it directs attention to the infrastructural work that stabilizes knowledge and to the historical processes that make certain ways of knowing appear natural, authoritative, and durable.

To operationalize or make infrastructure visible, Edwards employs infrastructural inversion. Drawing on Bowker and Star (1999), he turns infrastructure upside down to expose the socio-technical systems that usually remain invisible. He uses infrastructural inversion as a historiographical and analytic tool: by tracing how scientists revisit, repair, and reconcile their own data systems, he reveals the hidden work that allows climate science to appear stable and objective. Edwards shows that climate scientists themselves engage in infrastructural inversion when they scrutinize and reanalyze historical data, correcting for instrument changes, digitizing paper records, and reconciling inconsistencies across sites and decades, to produce coherent global datasets. Through this lens, climate knowledge is not a straightforward accumulation of observations but an ongoing process of infrastructural maintenance. The misuse accusations that ACC skeptics level at mainstream climate science, and that the mainstream science community levels at ACC skeptics, function as a form of informal infrastructural inversion: moments when the assumptions and values built into climate data infrastructure are exposed, contested, and made visible.

Since Edwards' original formulation, a substantial body of research has elaborated and applied the concept of knowledge infrastructures across STS, information science, and related fields. Work in STS and information science has examined knowledge infrastructures' historical development, their maintenance and repair, their materiality and organizational labor, and their role in shaping epistemic authority (Bates, 2017; Borgman, 2015; Borgman et al., 2016, 2019; Borgman & Wofford, 2021; Edwards, 2019; Hoeppe, 2019; Jackson, 2014, 2016; Karasti et al., 2016; Leonelli, 2013a; Plantin et al., 2016, 2018; Rayburn & Thomer, 2022; Shankar et al., 2016; Wofford et al., 2025). Across these works, knowledge infrastructures emerge as sites of negotiation rather than neutral technical arrangements, encoding values and shaping epistemic authority.

For my purposes, examining data practices within the broader framework of knowledge infrastructures enables two key moves. First, it foregrounds how data practices are not only situated but also dependent on infrastructural arrangements that may otherwise remain invisible.

Second, it highlights how practices themselves shape, maintain, or destabilize infrastructures over time. In the case of ACC skeptics, this combined perspective allows me to analyze both the everyday activities of data reuse and the socio-technical systems that make those activities possible, from repositories and platforms to norms, rules, and values. This dual focus provides a more comprehensive understanding of how the community produces and disseminates knowledge claims, illuminating the interplay between local practices and global infrastructure.

Applying this lens to the ACC skeptic data community, the community defined in Section 2.1 by shared data needs, methods, and epistemic commitments rather than disciplinary boundaries, the dissertation documents how their knowledge infrastructure shares technical dependencies with mainstream climate science while being reconfigured to support alternative interpretations and dissemination strategies. In the following subsections, I develop key concepts underlying knowledge infrastructures, including open data infrastructure and the challenges posed when data travel across communities and contexts.

2.5.1 Climate science knowledge infrastructure and the persistent presence of skeptics

Anthropogenic Climate Change, defined as "a statistically significant variation in either the mean state of the climate or in its variability, persisting for an extended period" caused by human activity (VijayaVenkataRaman et al., 2012), has reached an extraordinary level of consensus among publishing scientists, with 99.99% validating its existence (Powell, 2015). This consensus is not self-sustaining; it is produced and maintained by a knowledge infrastructure of extraordinary complexity and scale.

Edwards (2010a) describes the climate science knowledge infrastructure as a "vast machine:" a sociotechnical system that functions through three interlocking pillars: monitoring, modeling, and memory. Monitoring encompasses the global system of sensors and instrument platforms, including satellites, weather stations, ships, buoys, and weather balloons, that collect hundreds of terabytes of data each year. Memory refers to the distributed archives and data repositories, such as the National Snow and Ice Data Center, NASA Earthdata, and NOAA's National Centers for Environmental Information, that preserve this data as scientific memory, enabling researchers to track long-term variations across decades and centuries. Modeling is the core mechanism that makes the system function as a unified whole: because the Earth system is too large and complex for controlled laboratory experiments, computer models serve as virtual laboratories. These three pillars are supported by what Edwards calls a knowledge infrastructure

made up of robust networks of people, artifacts, and institutions, including the World Meteorological Organization, which sets international standards for data exchange, and the Intergovernmental Panel on Climate Change (IPCC), which synthesizes knowledge across disparate fields into periodic assessment reports.

A central feature of this infrastructure is what Edwards (2010a) calls the model-data symbiosis: models and data exist in a mutually dependent relationship in which neither can support a living understanding of the climate in isolation. On one side, models are data-laden, they incorporate empirical parameters derived from observations to represent processes that cannot be simulated from first principles, such as cloud formation. On the other, data become global through computer processing that filter noise, interpolate values for data-sparse regions, and reconcile measurements from different platforms, such as satellites and weather balloons. Edwards argues the following:

…without models, there are no data. I'm not talking about the difference between "raw" and "cooked" data. I mean this literally. Today, no collection of signals or observations—even from satellites, which can "see" the whole planet—becomes global in time and space without first passing through a series of data models. (Edwards, 2010, p. xiii).

This symbiosis has a direct consequence for historical understanding: because both observing systems and data models are constantly evolving, the climate record is never a single definitive account. Edwards calls this shimmering data where multiple global data images that continuously evolve and converge around a central warming trend rather than resolving into a fixed historical record (Edwards, 2010).

Managing uncertainty is a core activity of this KI rather than an incidental challenge. Uncertainty in climate science arises from several sources: observational inconsistency caused by changes in instrumentation, station locations, and observing practices over more than 160 years of records; computational limits that require the use of parameterization, mathematical approximations for processes like cloud formation, which remain the largest source of disagreement between models; and the chaotic nature of the atmosphere itself, which places a hard limit on long-term prediction (Edwards, 2010). Scientists address these uncertainties through several practices. Infrastructural inversion, in this case, involves probing every element of historical records, linking changes in data to technical changes in the infrastructure to correct for non-climatic factors (Edwards, 2010). Reanalysis reprocesses decades of raw historical

weather records through a single frozen modern model, creating physically self-consistent global datasets (Edwards, 2010). Together, these practices produce what Edwards (2010a) describes as "infrastructural globalism" the permanent institutional-technological complex designed to monitor, model, and preserve knowledge of the global climate.

Data management within climate science is complex, depending on extensive calibration routines, harmonization processes, and domain-specific metadata conventions (Edwards, 2010). Metadata fields often include instrument calibration, quality flags, uncertainty estimates, spatial referencing, units, and temporal aggregation all of which presume familiarity with domain standards and tacit knowledge developed through disciplinary training. Climate data repositories produce documentation that describes sensor drift, methodological changes, quality-control procedures, lineage, and processing histories. These practices collectively create a thick documentation culture that supports sophisticated reuse within expert communities (Edwards, 2010).

Skeptics take advantage of climate science knowledge infrastructure's characteristics to discredit the very knowledge it produces (Akin et al., 2020; Edwards, 2010, 2019; Prasad, 2019). ACC skeptics often exploit the climate science KI's reliance on models to dismiss the discipline and its findings wholesale as imaginary and politically motivated (Edwards, 2019). The shimmering nature of climate data, its continuously revised character, provides further opportunity for skeptic critique, as each revision of the historical record can be framed as evidence of manipulation rather than normal scientific refinement. In less extreme cases, they discover and report errors in climate data (Gramling, 2007) or use open climate data in direct service of their positions and beliefs (Lewandowsky, Mann, et al., 2016). Lewandowsky et al. performed a blind expert test on multiple contrarian claims about climate data and found them all to be misleading (Lewandowsky, Ballard, et al., 2016).

Additionally, because of the infrastructure's failure to publicly explain the practical uncertainty and ambiguity of scientific research, controversies abound when communications between scientists are released to the public. Oreskes and Conway (2011, p. 34) highlight how ACC skeptics use this "normal scientific uncertainty…to undermine the status of actual scientific knowledge. As in jujitsu, you could use science against itself." This is exemplified in the "Climategate" controversy, which led to a measurable decrease in public trust in climate science by using scientifically common discussions over uncertainty and data processing to suggest

scientists purposely manipulated data (A. A. Leiserowitz et al., 2013). Climate scientists and other environmental groups have offered alternatives to the status quo of science communication to allow for uncertainty and openness without the leveling of all knowledge claims (Berkhout et al., 2013; Dillon et al., 2019).

One prominent example of skeptic engagement with the knowledge infrastructure involved the University of Alabama in Huntsville (UAH) satellite-based temperature record. Dr. Roy Spencer and Dr. John Christy, climate scientists skeptical of the magnitude of projected warming, led the group that interpreted the microwave sounding unit (MSU) data, and in the early 1990s developed a synthetic channel to resolve lower-troposphere temperatures (Edwards, 2010). Their analysis showed a slight cooling of the lower troposphere, which mattered because models predicted that layer should warm somewhat faster than the surface (Edwards, 2010; E. A. Lloyd, 2012). Christy, Spencer, and other skeptics presented the record as empirical evidence against climate models (Edwards, 2010). The MSU does not measure temperature directly, however; it reads radiances at the top of the atmosphere, which are converted to temperatures through complex data models, so the record is itself heavily model-mediated (Edwards, 2010). Over the following years, independent groups, reprocessed the same instrument readings and made different processing choices, producing warmer trends compatible with the models. A series of corrections was also applied to the UAH analysis, including for orbital decay, inter-satellite bias, diurnal drift, and stratospheric radiances overlapping the measurement channel (Edwards, 2010). As these were incorporated, the satellite records converged toward one another and toward the surface trend, though the UAH record remained the coolest of them (Edwards, 2010). All of these datasets were derived from the same raw readings; the differences among them were differences between their data models (Edwards, 2010).

Ramírez-i-Ollé (Meritxell Ramírez-i-Ollé, 2019) documented a productive consequence of skeptic engagement with the climate science knowledge infrastructure. She describes a practice she calls “inside-out skepticism,” in which climate scientists internalize the critical perspective of skeptics during their own analysis, ensuring their methodological choices are defensible in public venues. Rather than simply reacting to external critique, researchers preemptively rationalize every step of their work, avoiding the black boxing of complex methods and moving toward more statistically rigorous and transparent standards. This practice has pushed some areas of climate science toward more explicit documentation, outcomes that serve

the broader knowledge infrastructure. In this sense, the relationship between ACC skeptics and the climate science knowledge infrastructure is not simply one of exploitation; it is entangled in ways that have, in some instances, strengthened the climate science KI.

2.5.2 The open data infrastructure

While definitions of open data vary across the literature, I follow the Open Knowledge Foundation's widely adopted formulation throughout this dissertation: data are considered open if they are "free to access, use, modify, and share it – subject, at most, to measures that preserve provenance and openness" (Open Knowledge Foundation, n.d.). Open data infrastructures encompass not only the data and repositories but also the people, policies, technical standards, and socio-political contexts that shape data creation, dissemination, and reuse. Open data has been lauded for its potential to facilitate reproducibility, make public assets available to the public, leverage investments in research, and advance collaboration and innovation in science (Margolis et al., 2014; Pasquetto et al., 2017; Wilkinson et al., 2016). Open data have fueled groundbreaking discoveries, supported transparency in governance, and created opportunities for new research, businesses, and even market sectors (Anderson, 2008; Arzberger et al., 2004; Boulton et al., 2015; Cukier, 2010; Field et al., 2009; Guillou, 2020; Hanson et al., 2011; Shankar et al., 2021). In response to these promises and benefits, governments, funding agencies, and even publishers have increasingly expected and mandated researchers to release data publicly. In the United States, the federal government has issued memoranda (Holdren, 2013; OSTP, 2022), executive orders (Obama, 2013), agency policies (Office of The Director, National Institutes of Health, 2023), and legislation (Ryan, 2019) that require federally funded researchers to make non-sensitive data publicly available.

Open data infrastructures form the foundation upon which climate science research, and much of its critique, now depends. The majority of climate scientists (Edwards, 2010, 2019) and ACC skeptics are reliant on open data and its infrastructure for their data practices (Wofford, 2022; Wofford & Thomer, 2023). Within the Earth and environmental sciences, these infrastructures include extensive public repositories such as NASA's Earthdata, Met Office's Climate Data Portal, NOAA's National Centers for Environmental Information, and the U.S. Geological Survey's open data portals, systems that epitomize open science ideals and provide the empirical foundation for climate research.

The open data movement rests on the assumption that data are fungible commodities that can be repurposed in any context. The "hypertransparency [of] open data, open code, commodity software tools, and alternative publication venues" allows diverse and distant reusers to engage in the research process (Edwards, 2019, p. 21). Yet, as scholars have argued, data cannot be meaningfully separated from the socio-political, technological, and epistemological contexts in which they were created (Bowker & Star, 1999; D'Ignazio & Klein, 2020; Gitelman, 2013). Data are created in a particular context by particular people with particular worldviews. Climate data, discussed further in the background chapter, are produced through situated practices, instrument calibration, model selection, and quality control, rooted in specific disciplinary traditions. When such data are detached from these contexts, reusers like ACC skeptics interpret them through different epistemic frameworks, drawing conclusions that diverge sharply from those of the original data producers. More understanding is required about what it takes for data to be successfully or unsuccessfully transferred from one setting to another (Leonelli, 2020).

The same features that make open data a powerful tool, like its accessibility, interoperability, and reusability, also introduce challenges and risks. The design of repositories and their metadata standards reflects assumptions about who the intended users are and what constitutes legitimate reuse (Bates et al., 2016; Bietz & Lee, 2009; Hine, 2006; Leonelli, 2014; Thomer & Wickett, 2020). The material forms of data repositories and curation activities profoundly impact how "these systems play in the social settings in which they are used" (Thomer & Wickett, 2020, p. 3). Studies show that repository design decisions are often guided by the needs of particular disciplinary communities (Borgman et al., 2015; Faniel et al., 2013; Faniel & Yakel, 2017; Weber et al., 2013). Thus, the open data infrastructure is not neutral or universally accessible but reflects the epistemic priorities and practices of its primary intended users. For communities like ACC skeptics, who operate at the margins of mainstream science, this creates both opportunities and friction. Repositories provide the raw materials for their analyses, yet disciplinary assumptions embedded in repository design may obscure meaning and complicate accurate interpretation. For ACC skeptics, the portability of open data enables practices of selective reuse, such as re-plotting, recontextualizing, or challenging official datasets, to question the credibility of climate models.

Furthermore, once the data are reused by ACC skeptics they travel to new spaces. In an era where "the power of big data" permeates public discourse as a potential force for good, it is

crucial to confront the challenges and limitations of data travel in the open data infrastructure (Leonelli, 2020, p. 9). While data travel may appear to democratize scientific participation, they also reveal how infrastructural openness can unintentionally facilitate epistemic conflict and fuel misuse (Pasquetto et al., 2024; Wofford & Thomer, 2023). Once data travel beyond their original communities, they become susceptible to reinterpretation, distortion, or strategic reframing. As Leonelli (2020) notes, successful "data travel" requires careful attention to the infrastructures and social processes that maintain meaning across contexts. When these are absent, data can lose their grounding. Without adequate attention to the social, technical, and epistemic contexts of data creation and reuse, the promise of open data remains unevenly realized, potentially privileging certain groups while marginalizing others.

Taken together, these studies underscore that the open data infrastructure is not a uniform or universally accessible system but a complex sociotechnical assemblage that mediates whose data practices are supported, constrained, or excluded. Understanding how these infrastructures shape, and are reshaped by, their reusers is critical for analyzing both the opportunities and the risks of openness. The data reuse practices of ACC skeptics offer a particularly revealing vantage point: they highlight how infrastructures designed for transparency and trust can be reinterpreted through alternative epistemic frameworks, and how infrastructural openness may simultaneously enable participation and contestation.

2.5.2.1 U.S. investment in making socio-environmental data open and usable

The United States government has a long, bipartisan history of investing in open socio-environmental data, data pertaining to the Earth's climate, atmosphere, ecosystems, and the human communities that interact with them, openly accessible to the public. These investments span decades and administrations, reflecting a fragile consensus that federal data collected in the public interest should be available for public use. The pivotal shift came in 1994, when NASA and NOAA adopted a "full and open" data policy for civil Earth observation satellites, establishing the principle that federally collected socio-environmental data belongs to the public (Committee on Best Practices for a Future Open Code Policy for NASA Space Science et al., 2018; Earth Science Data Systems, 2021). Subsequent administrations built on this foundation through memoranda (Holdren, 2013), executive orders (Obama, 2013), and more binding legislation (Ryan, 2019) requiring all federally-funded research to publish non-sensitive data publicly, progressively expanding the infrastructure of public repositories at NASA, NOAA, the

EPA, and USGS. The Biden administration extended this infrastructure with an explicit equity framing, directing that federally funded research and its underlying data be made immediately and freely available (OSTP, 2022) and broadening the intended user base to include communities previously marginal to federal data ecosystems (Earth Science Data Systems, 2022; Executive Office of the President, 2023).

This infrastructure has been substantially altered by the second Trump administration, beginning in January 2025. Federal agencies removed or modified thousands of web pages and datasets, with climate and public health content disproportionately targeted (Mosbergen et al., 2025; O'Leary, 2025; Singer, 2025; Stone, 2025). Tools central to climate and environmental research, including NCEI's Billion-Dollar Weather and Climate Disasters database, the EPA's EJScreen, and the congressionally mandated National Climate Assessment were removed from public access or ceased receiving updates (EDGI_Team, 2025; James & Haggerty, 2025; Press, 2025). The scale of these actions prompted an urgent preservation response from the research community, with the Data Rescue Project and academic institutions moving to archive federal datasets before their permanent deletion (*Data Rescue Project*, 2026).

In May 2025, President Trump issued Executive Order 14303, "Restoring Gold Standard Science" (*Restoring Gold Standard Science*, 2025), establishing federal science policy around tenets including reproducibility, transparency, falsifiability, and skepticism of findings. Readers familiar with ACC skeptic discourse may recognize this vocabulary, it has long been the rhetoric of skeptic communities (Edwards, 2010, 2019; Oreskes & Conway, 2011). The order, in a way, institutionalizes the epistemic posture of ACC skeptics. This EO may seem contradictory when considering the data deletions the Trump administration conducted. However, the administration reconciles this by emphasizing that some data and webpages must be cleansed of "politicized ideology" as Secretary of the Interior Doug Burgum states (*Interior Department Restores Gold Standard Science Across Federal Decision-Making | U.S. Department of the Interior*, 2025). Simultaneously, the order introduces aggressive new proactive disclosure requirements, including mandating that agencies publish the underlying data, analyses, and source code for models used to set public policy, and restricting agencies from using FOIA exemptions to withhold scientific models and analyses. It also concentrates authority over perceived integrity violations in senior political appointees, who are empowered to correct scientific information deemed inconsistent with the administration's standards and recommend disciplinary action

against personnel involved in integrity disputes. The open data infrastructure ACC skeptics engage with has been partially dismantled at precisely the moment their community's foundational epistemic framework has been elevated into federal policy.

2.6 Studying data practices and knowledge infrastructures through data journeys

The frameworks established in the preceding sections — data communities, data practices, data misuse, and knowledge infrastructures — together describe a landscape in which data move across boundaries, accumulate new meanings, and generate conflict. What they do not yet provide is a method for tracing those movements empirically. Data journeys fill that gap. Rather than studying communities, practices, or infrastructures in isolation, data journeys follow the data themselves, attending to how their form, interpretation, and value shift as they travel across sites of production, curation, dissemination, and reuse. This approach is particularly well suited to studying ACC skeptics, whose data practices are defined precisely by the movement of data across epistemic boundaries, from the repositories of mainstream climate science into the alternative knowledge infrastructures of the skeptic community.

To study these dynamics empirically, this dissertation adopts the concept of data journeys (Bates et al., 2016; Leonelli, 2020). Leonelli (Leonelli, 2020) describes data journeys as an adaptable framework that questions fixed locations, foregrounds non-linear narratives, draws on case studies, engages directly with data practitioners, and attends to reflexivity throughout the research process. Bates, on the other hand, offers more prescriptive steps to conduct a data journey focusing on mapping the data journey and taking steps to reveal the infrastructure (Bates et al., 2016)**.**Data journeys trace how data travel across sites of production, curation, dissemination, and reuse, attending to transformations in form, interpretation, and value as data circulate through heterogeneous communities and infrastructures. This approach foregrounds the social, technical, and institutional work required to make data reusable, while also revealing how assumptions about appropriate use are challenged as data move beyond their original contexts.

As a tool, data journeys take inspiration from infrastructural inversion, with its spotlight firmly on the data themselves and the implications that infrastructures have on their interpretation (Bowker & Star, 1999). In this case, infrastructural inversion involves bringing to light the often-invisible infrastructures — such as databases, platforms, algorithms, power dynamics, technical constraints, institutional policies, and more — that facilitate the movement and interpretation of data.

By employing infrastructural inversion as a conceptual-analytic strategy, I shift the focus from the knowledge-making activities invisibly supported by the infrastructure to the foundational elements that enable the infrastructure to function (Bowker, 1994; Bowker & Star, 1999). Applying infrastructural inversion involves a "gestalt shift" of figure-ground relations examining the usually hidden aspects of the systems, technologies, activities, material conditions, power asymmetries, and social arrangements that support ACC skeptics' data practices. This method is intended to bring infrastructure to the foreground by "learning to look closely at technologies and arrangements that, by design or habit, tend to fade into the woodwork" (Bowker & Star, 1999, p. 9). In other words, infrastructural inversion allows us to interrogate the foundational elements that typically operate unnoticed but are essential for functioning knowledge infrastructures. Within a data journeys framework, infrastructural inversion is the analytic move that makes the journey visible. It is what allows us to see not just where data travel, but what infrastructural conditions enable, constrain, and shape that travel at each step.

Infrastructural inversion has been particularly insightful in the realm of climate science. As Edwards (2010a) notes, climate scientists have engaged in infrastructural inversion in their own work for decades to address gaps in historical climate data. The global climate knowledge infrastructure is built on various datasets, many originally localized and inconsistent. In the 1970s, scientists meticulously re-examined historical data, reconfiguring records to form a coherent global dataset. This process, a form of infrastructural inversion, involved digitizing old weather records, correcting errors, and compensating for gaps or inconsistencies. Through this process, the climate science community produced robust models and datasets that are now foundational to the Intergovernmental Panel on Climate Change (IPCC), the most visible layer of climate knowledge infrastructure. This reexamination of historical data highlights how the climate science infrastructure continually opens itself up, reanalyzing data and metadata to ensure accuracy, reliability, and historical coherence (Edwards, 2010). For this dissertation, this example illustrates what data journeys can reveal methodologically: by tracing the work scientists did to reconcile and stabilize historical records, we can see how infrastructure shapes what counts as reliable data and how meaning is maintained, or lost, as data move across time and context.

In contrast, ACC skeptics have also employed infrastructural inversion for different ends (Edwards, 2010). They dig deeply into the data and models produced by climate scientists, often to critique and challenge their findings. This skepticism manifests through questioning the assumptions and techniques underpinning climate models. By scrutinizing the infrastructure behind climate knowledge, they attempt to uncover inconsistencies or weaknesses, framing their critiques around the idea that models are inherently uncertain. Tracing these skeptic data journeys using infrastructural inversion as an analytic tool allows me to examine not only what skeptics do with data, but what the infrastructure makes possible, and what it forecloses, as data move from mainstream repositories into skeptic communities and back again. This dichotomy shows how infrastructural inversion can be used to achieve very different ends by more than just social science researchers.

2.7 Conclusion

Taken together, the theoretical frameworks outlined in this chapter, data communities, data practices, knowledge infrastructures, data misuse, and data journeys, provide an integrated multifaceted lens for examining how ACC skeptics reuse open data. Rather than treating data reuse as a linear process, this study approaches reuse as a set of situated data practices that emerge within communities and are shaped by the knowledge infrastructures, they rely upon. This framework guides both what this study attends to and what it does not attempt to explain.

First, this framework specifies the primary analytic focus of the study. The unit of analysis is not individual belief or attitude toward climate change, but observable data reuse practices as they unfold within ACC skeptic data communities. Drawing on a practice-theoretical perspective, the analysis centers on what actors do with data, how they search for, select, process, analyze, represent, and share data, and how these activities are shaped by sociotechnical infrastructures such as repositories, platforms, metadata standards, and documentation practices. Data reuse activities outlined in Section 2.3 function as categories rather than as a prescriptive lifecycle; they provide a vocabulary for identifying data practices.

Second, this framework clarifies which dimensions of data reuse are examined most closely in this study. While all high-level data reuse activities described in Section 2.3 inform the analysis, the empirical chapters focus especially on practices that are most visible in ACC skeptics' engagements with open climate data. These include data discovery and selection, data processing and transformation, analytic and interpretive practices, and practices of quality

assessment and critique. Attention is given to how trust and distrust in the data are developed through these practices, as well as how research products, such as graphs, blog posts, and reports, are shared within and beyond skeptic communities. Attention is also given to how instances of data misuse and accusations of misuse surface conflicts over epistemic authority, credibility, and infrastructural norms as data travel across communities. Other activities, such as the management and re-representation of data and documentation, formal data publishing, and standardized data citation, appear more unevenly.

Third, the framework describes the boundaries of the study. This theoretical approach does not position the analysis to decide the scientific validity or correctness of ACC skeptics' climate claims. Nor does it seek to infer individual motivations, internal beliefs, or private reasoning processes beyond what can be observed through public-facing data practices, interviews, and their traces. The study does not attempt to reconstruct original data production in full, nor does it capture all offline or private forms of data work. Instead, the focus remains on how data are mobilized, transformed, and made meaningful as they travel across open infrastructures and online communities.

This framework grounds chapters that follow. Chapter 3 details how this theoretical framework is operationalized through a digital ethnography consisting of content analysis, participant observation, and semi-structured interviews. Chapter 4 traces in detail the data journeys of four ACC skeptics across the open data infrastructures described in this chapter. Chapter 5 moves beyond these journeys to the data practices observed across the full corpus of the study. Chapter 6 turns to the knowledge infrastructures ACC skeptics rely on and, in some cases, build. Chapter 7 draws these findings together in a discussion organized around three theoretical contributions and a set of practical implications. The discussion addresses the audit orientation that organizes the skeptic knowledge infrastructure, the community's radicalization of mainstream scientific values, and the entanglement of the ACC skeptic and mainstream climate science knowledge infrastructures**.** The chapter then turns to practical implications for data intermediaries and repositories, addressing reframing open data governance, designing for trust across epistemic differences, designing for and responding to potential data misuse, and the role of information professionals in these situations.

Chapter 3 Methods

Within information science, researchers primarily employ qualitative methods to study data reuse practices. Specific methods include direct interaction with data users and reusers, such as surveys (Curty et al., 2017; Faniel et al., 2016; K. Gregory et al., 2020; Tenopir, Allard, et al., 2011; Tenopir et al., 2015; Yoon & Kim, 2017), interview studies (Faniel et al., 2012, 2012, 2016; Muller et al., 2019; Yoon, 2017; Zimmerman, 2008), case studies (Kassen, 2021; Thomer, 2022), and ethnographies (Borgman et al., 2020; Hoeppe, 2021; Pasquetto et al., 2019; Wallis et al., 2012). Using these methods, researchers directly ask data reusers about their practices. These methods are especially valuable for investigating sociotechnical questions and for understanding the lived realities of data reuse.

A smaller number of quantitative studies take a more indirect approach, examining the products of scholarly work. Bibliometric methods, for instance, trace reuse through data access requests (Coady et al., 2017; Federer, 2019) or citations in published literature (Huang et al., 2015; Khan et al., 2019; Park et al., 2018; Peters et al., 2017; Pienta et al., 2010; Piwowar & Vision, 2013; Robinson-García et al., 2016). I highlight these approaches to show both the range of tools available and their different affordances for studying reuse. While indirect methods are useful for identifying broad trends, they often capture only what researchers intend to do—or what they choose to disclose in formal outputs.

Each of these methodological approaches offers insights yet also has limitations. Prospective studies using methods such as surveys, data access request analyses, and grant applications often reveal researchers' intentions or expectations around data reuse, rather than their actual practices. Similarly, studies relying on citation metrics capture only reuse that results in formal publication, thereby excluding informal, unpublished, or unconventional reuses and reflecting only what reusers choose to disclose. Interview-based studies provide greater depth on reuse practices but remain reliant on self-reporting and memory. Ethnographies and many qualitative case studies allow for richer, more contextualized insights into real-world data reuse practices, even as they typically focus on small, often institutionally embedded populations,

limiting broader generalizability. There are also emerging methods, such as workflow capture tools, that offer new avenues for direct observation of data handling and analysis (Mash et al., 2021).

In what follows, I explain why digital ethnography is particularly suited to the community I study: anthropogenic climate crisis (ACC) skeptics. Many existing studies risk overlooking this community entirely or mischaracterizing it as a low-uptake outlier. For example, much of the community's data reuse does not appear in published academic literature and would therefore be invisible if I relied on citation metrics. Similarly, a survey-based study would likely generate few responses due to the community's deep distrust of academia. For the same reason, I chose not to depend on access solely through interviews or traditional ethnography.

Because ACC skeptics are especially active on social media, I hypothesized that much of their data reuse practices would be visible publicly online. This proved correct. Digital ethnography allowed me to observe these publicly available practices and supplement them with semi-structured interviews from community members willing to engage with me. This approach also allowed me to map the infrastructure of skepticism: the discourses, tools, and social networks through which data are mobilized.

This study is situated in the interpretive tradition of qualitative research, which emphasizes an idiographic and inductive process. In this approach, researchers provide a "thick description" of a specific context (Geertz, 1973), allowing observations to guide the generation of theory (Neuman, 2014). Rather than testing hypotheses or generating generalizable results, interpretive research prioritizes depth, context, and complexity. My aim is not to produce a representative account of all data reusers, but to engage deeply with one group, anthropogenic climate crisis skeptics, to understand how they create meaning from data. As Neuman notes (Neuman, 2014, p. 106) many interpretivists believe "laws sought by positivists may be found only after the scientific community understands how people create and use meaning systems, how common sense develops, and how people apply their common sense to situations." In this spirit, I argue that deeply understanding users of data repositories is essential when designing systems that they will use. At the same time, this digital ethnography is not only a source of "implications for design"; it also reveals the conceptual logics underlying skeptics' data practices and the infrastructures they rely upon (Dourish, 2006).

The remainder of this chapter outlines the methodological approach of this study, detailing how I operationalize digital ethnography. It also describes how I analyze data, address my positionality and reflexivity, consider ethical issues, and present the study's limitations.

3.1 Digital Ethnography

Ethnography is a qualitative research method centered on long-term, immersive engagement with a community in order to understand how members create meaning in their everyday practices (Geertz, 1973; Hammersley & Atkinson, 2019). Digital ethnography extends these principles to online environments, emphasizing how social interactions, practices, and infrastructures are mediated through digital technologies (Hine, 2020; Pink et al., 2015). This approach allows researchers to trace practices across platforms and to attend to the fluid boundaries between online and offline contexts in modern life. Digital ethnography is not without limitations: it can be difficult to observe private exchanges, platform affordances shape what is visible to the researcher, field sites lack natural boundaries, new ethical challenges are introduced, and interpreting behavior without in-person context carries risks of misinterpretation (Cera, 2023; Hine, 2020; Pink et al., 2015; Seta, 2020).

For anthropogenic climate crisis (ACC) skeptics, much of the data reuse and meaning making occurs or is presented on digital platforms. Digital ethnography is particularly well suited to studying hard-to-reach communities: participants often disclose more in online settings than they would in person, and the method allows for deep immersion into subcultures whose logics and practices would be difficult to access through conventional means (Cera, 2023). Digital ethnography thus provides a methodological fit, enabling observation of publicly visible practices while also allowing for deeper insights through semi-structured interviews with community members willing to participate.

Specifically, this study integrates three ethnographic techniques - content analysis, participant observation, and semi-structured interviews - adapted to a digital field (Table 3.1), following boyd's (boyd, 2015) model of digital ethnography as reflecting traditional ethnography's core components. No single method was sufficient for studying this community: content analysis could not capture subjective experience; interviews could not capture the breadth of public practice; participant observation alone could not reconstruct the technical details of specific data reuse events. Content analysis traces how data and knowledge artifacts circulate (Altheide, 1987; Krippendorff, 2018; Lacy et al., 2015; Schreier, 2012); participant

observation provides insight into norms, routines, and infrastructural engagements (Hammersley & Atkinson, 2019; Spradley, 1980); and interviews capture subjective perspectives that may not surface in public discussions (Galletta & Cross, 2013; Kvale & Brinkmann, 2009; Rubin & Rubin, 2011; C. A. B. Warren, 2001). Used together, these methods enable triangulation, helping to capture multiple dimensions of complex socio-technical practices (Neuman, 2014).

As a reminder, this study is guided by the following research questions:

- What are the data practices of anthropogenic climate crisis (ACC) skeptics?
- How does this community develop trust or distrust in the data they reuse?
- What are the knowledge infrastructures that support data reuse within this community, and how do they operate?

Table 3.1 Summarizes how each research activity contributes to addressing these questions.

Table 3.1: Overview of Research Activities

Method	**Data Collected**	**Contribution to Research Question 1: Data Practices**	**Contribution to Research Question 2: Infrastructure**
Content Analysis	Posts mentioning or reusing data; associated comments	Documents publicly available data practices and knowledge artifacts; traces how members interact with them	Maps of data journeys; reveals how infrastructures shape practices as data move across socio-technical contexts
Participant Observation	Memos from online monitoring	Identifies routines, values, and practices; surfaces trust-related dynamics.	Reveals infrastructures (technologies, values, norms, expectations, routines, power relations) that enable or constrain practices at each stage of the data journey.
Semi-structured Interviews	Transcripts; memos	Fills gaps in observed practices; captures perspectives not publicly available; member-checking of preliminary findings	Walkthroughs of specific data reuse cases; reveals hidden infrastructural components (e.g., software environments)

To operationalize this approach, I established boundaries for the digital ethnography. Defining the scope of platforms, actors, and artifacts was essential both for methodological rigor and for ensuring that the study remained manageable while still capturing the fluid and networked nature of the online community. The following section describes how I scoped the

digital field site, drawing on existing models of investigative digital ethnography and data journeys to create the monitoring environment.

3.1.1 Boundaries of Digital Ethnography

Digital field sites lack natural boundaries which requires researchers to make decisions about which platforms, actors, and artifacts to include and which to set aside (Seta, 2020). If those scoping decisions are not made carefully, the result is a skeletal representation of the community's social world rather than a rich one (Seta, 2020). On the other hand, because digital networks can proliferate uncontrollably, researchers must also engage in deliberate pruning if only for the sake of completing the research (Seta, 2020). This section describes how those decisions were made for this study.

The ACC skeptic data community is geographically distributed across the world and thus understandably organized through digital platforms (as discussed in Section 2.1.1). Early mapping of the community identified the blogosphere as its central infrastructure, with *Watts Up With That*, *Climate Audit*, and *JoNova* serving as the most linked-to and visited sites, functioning as alternative public sites of expertise where community members reinterpret climate data and challenge mainstream scientific claims (Sharman, 2014). These blog sites still function as such as this study shows. By the time this study commenced in September 2021, the community had expanded across mainstream platforms including X, YouTube, Reddit, and Facebook, as well as alt-tech platforms such as Telegram, Truth Social, and Gab (Essien, 2025; Singh, 2025; Treen et al., 2022).

The monitoring environment was scoped to X, Reddit, and YouTube for multiple reasons. First, previous research had shown this is where many prominent data reusers were active (Essien, 2025; Singh, 2025; Treen et al., 2022). Second, during the monitoring period, social media platforms functioned predominantly as distribution channels for content produced elsewhere. The same data reuse blog posts appeared repeatedly across X, and Reddit, suggesting that most data reuse activity originated in blogs and was subsequently amplified across social media. Additional social media monitoring was therefore unlikely to surface substantively different data reuse practices and was not pursued. YouTube was included selectively, primarily where channels hosted substantive data reuse content.

Alt-tech platforms such as Telegram, Truth Social, and Gab tend toward conspiratorial and politically extremist content, which was not what I was interested in for this study (Tucci et

al., 2025). Entering these spaces would also have raised additional ethical and safety concerns for the researcher, discussed further in Section 3.7. Meta platforms were excluded for a related reason: Facebook's group architecture, particularly for smaller communities, required revealing my identity to gain access, raising ethical concerns given the community's documented distrust of academic researchers and potential risks to researcher safety.

The boundaries of the digital ethnography are inspired by Friedberg's investigative digital ethnography (Friedberg, 2020), Bates et al.'s data journeys (Bates et al., 2016), and Leonelli's data journeys (Leonelli, 2020). Following Friedberg's preliminary steps, I established the monitoring environment by identifying relevant topics and artifacts, ascertaining the media ecosystem and influencers, and creating a monitoring strategy. This study focuses on data reuse by the community's members and the knowledge artifacts the members produce from that data. These knowledge artifacts include graphs, statistics, tables, charts, maps, computational models, code, and more. I also examine how members evaluate and discuss the data and artifacts created. I collected data with an added focus on the infrastructure that supports this data reuse, both the open science infrastructure and the community's own infrastructures (Pink et al., 2015). Boundary determination also involved selections regarding community privacy and research safety: whether to follow data into closed groups or more private messaging spaces. For this study, I limited the monitoring environment to publicly accessible spaces, a decision informed both by ethical considerations discussed in Section 3.7 and by the community's extensive public-facing activity.

To create the monitoring environment, I used Google Chrome as my browser and created a separate Gmail and Google Voice account for the digital ethnography. I created accounts on social media sites that the community is active on, including Reddit, X, and YouTube, and seeded those accounts over fifteen days following popular organizations and influencers. I engaged with platform affordances, such as viewing, liking, and reposting content, to influence recommendation algorithms and expand the monitoring environment. The recommendation algorithm offered new accounts to follow through what Friedberg (2020) calls algorithmic-assisted snowball sampling: with monitoring accounts, algorithmic suggestions of topics and accounts to follow become a discovery tool, allowing the ecosystem itself to surface relevant actors and communities at its edges. Platform recommendation systems drove the chain rather

than human referrals, but the underlying logic is the same: engagement with known actors surfaces connections to less visible ones. (Biernacki & Waldorf, 1981).

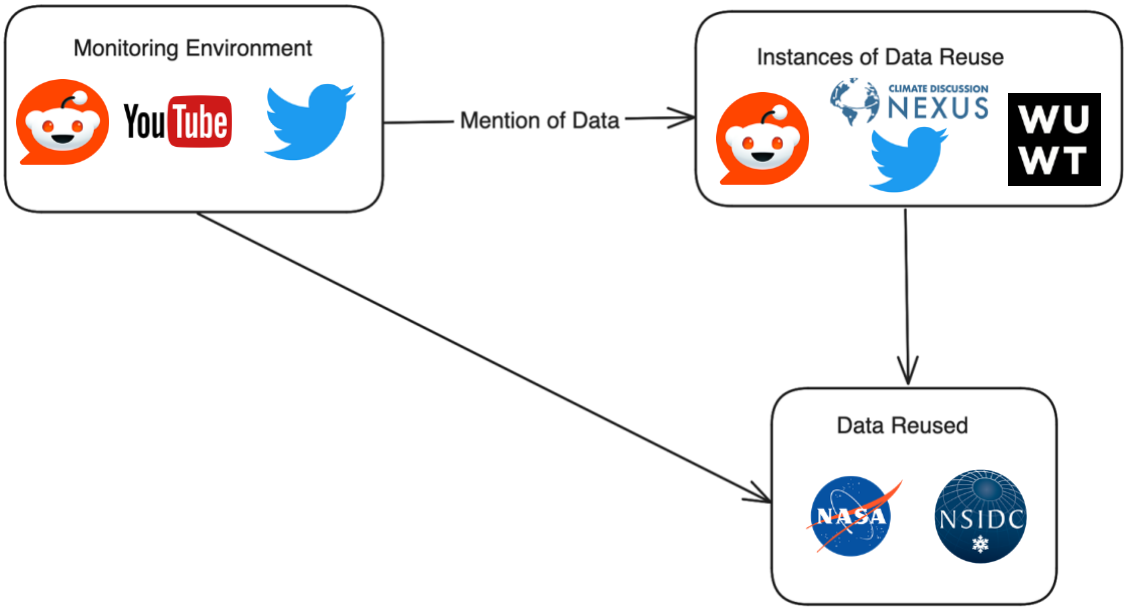
Algorithmic snowball sampling, like its human referral counterpart, may yield a sample biased toward the portion of the community most visible within platform networks. Accounts active on Reddit, X, and YouTube were well represented; communities operating primarily through private channels, email lists, or less algorithmically prominent platforms may be underrepresented. The monitoring environment therefore captures the community's public-facing activity more fully than its private deliberations.

At the same time, to minimize the impact of this limitation, I focused on background research to understand the community more fully, and incorporated the well-established ACC skeptic blogosphere in the monitoring environment (Lay, 2022; Sharman, 2014). Sharman (2014) mapped 171 ACC skeptic blogs using social network analysis, identifying the most central and influential nodes in the blogosphere and categorizing them by thematic focus. The DeSmog Climate Disinformation Database (Lay, 2022) is a searchable, regularly updated tool that documents individuals and organizations involved in efforts to delay public and legislative action on climate change, drawing on investigative research rather than network analysis. These mappings provided a starting point for identifying key community members and platforms. Both mappings were designed with different purposes than this study: Sharman's work aimed to characterize the blogosphere's structure and influence networks, while the DeSmog database focuses on documenting ideological and organizational affiliations. Neither was designed to identify data reuse practices specifically. This study used them as entry points for identifying prominent community members and key platforms, then narrowed the monitoring environment around actors whose work centrally involved the reuse of open climate data. The monitoring environment ultimately comprised 40 Twitter accounts, two subreddits, and five YouTube channels.

Although the monitoring environment was established, I did not restrict myself to just these accounts; instead, I allowed the data to guide my investigation. I followed the data's journey in this community focusing on the socio-technical conditions that shaped their movement and transformation (Figure 3.1). As discussed in Leonelli's data journeys, I traced the data back to "their production site [and] to the sites in which they are processed, mobilized, and re-purposed" in this community (Leonelli, 2020, p. 9). Drawing from both Bates and Leonelli's

framework, this process allowed me not only to observe the reuse of data by the community but also critically examine the knowledge infrastructures that facilitate and constrain these practices through infrastructural inversion.

Figure 3.1: Workflow during a monitoring session for ACC skeptics following the data and questioning fixed locations.



3.2 Content analysis

The content analysis component of this project is grounded in two complementary methodological frameworks: ethnographic content analysis and trace ethnography. The content analysis draws on what Altheide (1987) calls ethnographic content analysis (ECA): the reflexive analysis of documents, oriented toward discovery and understanding rather than verification, that is common in more quantitatively-based content analyses. ECA moves reflexively and circularly between concept development, data collection, coding, and interpretation, with categories and emphases expected to emerge during the study rather than being fixed in advance. For this dissertation, that means attending to how ACC skeptics themselves describe, evaluate, and produce data reuse, rather than coding their practices into pre-established analytical categories at the start of coding.

Trace ethnography extends traditional ethnography using documentary traces, produced within sociotechnical systems (Geiger & Ribes, 2011). Unlike purely observational or interview-based methods, trace ethnography allows researchers to reconstruct interactional routines and

infrastructural dependencies retroactively, offering a fine-grained understanding of practices and information flow across distributed contexts. In a digital field site like this one, where community activity is asynchronous and distributed across platforms and time, trace ethnography is particularly valuable: it allows the researcher to follow a knowledge artifact backward through its documented history and forward through its subsequent circulation, rather than being limited to what is observable in real time.

The approach and what documentary evidence are analyzed is operationalized here through the data journeys framework. Again, data journeys follow data where they take you from their production site to the sites in which they are processed, mobilized, and repurposed (Bates et al., 2016; Leonelli, 2020). Data journeys guide what posts and documents are analyzed as well as the higher-level movement and transformation of data through the community. Infrastructural inversion (Bowker & Star, 1999; Edwards, 2010) guides what to attend to within each journey: the friction points, breakdowns, and moments of contestation where the infrastructure that supports the journey becomes visible.

Sampling was purposive and theoretical in Altheide's (Altheide, 1987) sense: not drawn randomly from a pre-fixed corpus but guided by the data journeys themselves. The starting points were posts in the monitoring environment that mentioned or reused data; from there, I followed embedded links, traced the origins of knowledge artifacts, and returned to cited repositories and datasets. This involved detailed examination of online posts that mention data and the knowledge artifacts produced from that data, including graphs, statistics, tables, charts, maps, computational models, and code. I followed where the data led, not fixing myself to one location, to comprehensively trace how data travel and transform within the community (Figure 3.1).

I began observation periods by browsing through the monitoring environment, focusing on Reddit, YouTube, or X/Twitter, for real-time discussions about data. When real-time monitoring was not fruitful, I drew on trace ethnography to examine past instances of data reuse by analyzing existing documentary traces. (Geiger & Ribes, 2011). In all instances where data were discussed or reused, I archived the post and followed any embedded links, creating a trail. Where data were not explicitly mentioned, but the links seemed likely to discuss them, I followed and read through those links. Those links often led to blogs or organizations that discussed data reuse in more depth, and those websites were also archived. I also returned to the

cited data repository or dataset to make notes on documentation, affordances, and any other relevant information. Through this process, I archived the following documents as shown in Table 3.2 from September 2021 through December 2025. All were entered into NVivo for qualitative coding. Not all archived content was coded, as described in Section 3.5. Table 3.2 reports both the number of documents archived, and the number coded for each document type.

Table 3.2: Documents archived and coded during content analysis, September 2021 through December 2025.

Source Type	Number Archived	Number Coded
Blog Posts + Comments	120	71
Reddit Posts + Comments	81	45
Website Pages	59	30
X/Twitter Posts + Comments	19	13
YouTube Transcripts	13	9
Preprints	6	5
Thinktank Reports	4	1
Journal Articles	4	2
Government Reports	1	1
Total Documents	307	177

Several limitations of this content analysis are worth acknowledging. Digital content is ephemeral: posts may be deleted, blogs taken offline, and links broken over time (Krippendorff, 2018; Lacy et al., 2015). Archiving at the point of collection mitigated this issue as well as documenting connections between sites in my metadata (ex: Reddit post led to this blog post). Content analysis of publicly visible material cannot directly access author intent or motivation, making certain inferences difficult to validate (Krippendorff, 2018). I addressed this concern through triangulation across content analysis, participant observation, and semi-structured interviews addresses. Finally, this content analysis was conducted by a single coder, meaning intercoder reliability measures standard were not applicable. In keeping with ECA practice, however, reliability was addressed instead through prolonged engagement with the data, reflexive memoing, and member checking (Altheide, 1987).

3.3 Participant observation

Participant observation is the primary tool for generating what Geertz (1973) calls "thick description" which requires a sustained, contextually grounded account of life that reveals not just what people do but what it means to be them. By immersing themselves in a setting over time, researchers can make visible the taken-for-granted norms, routines, and tacit rules that structure everyday practice, which self-report methods like surveys and interviews often miss (Hammersley & Atkinson, 2019). In the context of this study, participant observation was well suited to the community: ACC skeptics conduct much of their activity publicly online, making their practices directly observable, and their deep distrust of academic researchers made approaches that depended on self-report alone unreliable as primary methods.

Participant observation in digital ethnographies can encompass a spectrum of activities, from passive browsing to active engagement (Seta, 2020). These includes following links, navigating between platforms, commenting, liking or disliking content, posting, messaging users, and more (Hine, 2007). In digital spaces, physical co-location is replaced by co-presence, which means being online and attentive across platforms rather than physically situated in a bounded field site (Beaulieu, 2010) Extended immersion in a community through what Geertz calls "deep hanging out" (Geertz, 1998) remains a core practice even in digital contexts (Marwick & Partin, 2022).

The project employed participant observation to understand the ACC skeptic community's values, shared norms, routines, practices, and interactions with infrastructure. The data collected during this process consist of memos written after online observation periods and email conversations with potential interview subjects. By employing infrastructural inversion during participant observation, I made visible the often-invisible knowledge infrastructures that ACC skeptics interact with and create, uncovering the socio-technical networks that enable their data practices by directing attention not only to their routine actions but also to the underlying infrastructures that support them.

Infrastructural inversion is mobilized both as a conceptual-analytic strategy and an empirical-ethnographic strategy within my methodology (Simonsen et al., 2020). First, as a conceptual-analytic strategy, I use infrastructural inversion to examine how infrastructure supports ACC skeptics' data practices with a figure-ground reversal highlighting the underlying mechanisms, standards, and routines hidden during smooth operation. I follow how the data

travels, tracing connections between technologies, organizational routes, and social practices that form the backbone of the infrastructure. This includes focusing on the mundane, "singularly unexciting" details of these infrastructures, such as technical standards, or administrative tasks that ACC skeptics complete or interact with (Bowker & Star, 1999, p. 377). These small details are crucial to understanding how infrastructure works in practice. The analysis made visible the often-overlooked work that keeps infrastructures running smoothly I remained cognizant of potential disruptions, disturbances, and breakdowns in the infrastructures that skeptics rely on, which "elicit naturally occurring inversions in which various exigencies make infrastructural operations abundantly visible" and easier to analyze and interrogate (Harvey, 2016, p. 4).

Second, I also use infrastructural inversion as an empirical-ethnographic strategy (Dagiral & Peerbaye, 2016; Kaltenbrunner, 2015; Parmiggiani, 2015; Parmiggiani et al., 2015) because of ACC skeptics' history of infrastructural inversion (Edwards, 2010). ACC skeptics engage in infrastructural inversion to uncover inconsistencies or weaknesses in the climate knowledge infrastructure. This reflects the often seen reality that before inversion becomes an analytic tool for researchers, it is first encountered as an empirical experience by those who work with the infrastructure (Harvey, 2016; Jensen, 2014). Breakdowns present valuable opportunities for both researchers and those engaged with infrastructures, as they expose "entire swaths of infrastructural activity (even those that didn't fail such as properly functioning back-ups or routinized repair activities) are revealed" (Ribes & Lee, 2010, p. 238). Breakdowns perceived by ACC skeptics, real or not, served as entry points for infrastructural inversion. By identifying and following their instances of infrastructural inversion, I uncovered which aspects of the climate science knowledge infrastructure they interact with and how they use inversion as a tool.

I take inspiration from both types of data journeys (Bates et al., 2016; Leonelli, 2020) in my own methodology, I pay attention to the following during participant observation to reveal more about these knowledge infrastructures:

- Frictions and Barriers to Reuse: I analyzed the friction points - places where data movement was slowed or altered due to socio-technical factors (Bates et al., 2016) which revealed the infrastructure more easily (Harvey, 2016; Ribes & Lee, 2010). Understanding these points of friction allowed me to also assess how knowledge infrastructures shaped not only the flow of information but also the credibility and authority of knowledge claims within the ACC skeptic community.

- Acts of Infrastructural Inversion by ACC Skeptics: I identified instances of infrastructural inversion conducted by ACC skeptics, which revealed both the data practices of ACC skeptics as well as infrastructural components ACC skeptics were particularly interested in (Edwards, 2010; Simonsen et al., 2020).

My objective was to engage authentically, necessitating an understanding of the norms and practices unique to each platform. Predominantly, I immersed myself in the community through "deep hanging out" spending extended periods within a group, mainly (Geertz, 1998; Marwick & Partin, 2022). I did not engage in explicit content creation of my own posts as a significant portion of daily interactions on websites, apps, and online services involved activities like reading, watching, and searching that do not explicitly require active participation (Seta, 2020), and posting risked misrepresenting my presence in the community.

However, I did engage in conversations with ACC skeptics, often through the semi-structured interview recruitment process. In my initial email, I identified myself as a researcher and the purpose of my project. This often prompted conversations back and forth, even when individuals were not interested in interviews. I engaged in participant observation for 220 hours, from September 2021 through December 2025, writing 77 memos concerning the community's norms, data practices, interactions with infrastructure, and more.

3.4 Semi-structured interviews

Semi-structured interviews with data reusers were important to understanding the subjective experience of using open research data that were not fully captured through online content analysis or participant observation. Semi-structured interviews are not only tools for gathering information but also interpretive encounters, in which meaning is co-constructed between researcher and participant (Kvale, 1996; Brinkmann, 2014). This framing emphasizes interviews as opportunities to elicit participants' own accounts of data practices, values, and challenges, rather than as neutral question-and-answer sessions. As Spradley (1979) argues, interviews within ethnography are especially valuable for eliciting "narratives of practice," complementing observation by capturing the ways participants themselves describe their world.

In this project, interviews were designed to fill gaps by engaging directly with participants to discuss details such as data interpretation techniques, software environments used for analysis, and specific challenges they encountered with data reuse. Additionally, where applicable, I conducted walkthroughs of specific instances of data reuse identified during

participant observation and content analysis. Such walkthroughs situated the discussion in concrete practices, allowing participants to reconstruct their interactions with datasets, tools, platforms, and people. These walkthroughs contributed to the mapping of data journeys, highlighting key infrastructural components at different stages of data reuse.

The semi-structured format allowed for flexibility while still addressing key themes that had emerged from other branches of the ethnography (DiCicco-Bloom & Crabtree, 2006). The protocol relied primarily on open-ended questions to elicit participants' own accounts rather than directing them toward predetermined answers, with more structured probes used during walkthroughs of specific data reuse instances or at the end of the interview (Galletta & Cross, 2013). Introducing specific or theory-laden questions too early also risks shortchanging the emergence of grounded data or causing participants to feel guarded before the conversation has developed (Galletta & Cross, 2013). Participant observation and content analysis informed the specific questions in the semi-structured interview protocols. For instance, questions about trust development and the credibility of datasets, which emerged as significant themes during the preliminary phase of research, were incorporated. Drawing from Sands et al. (2012), interview protocols encouraged participants to discuss their most recent data reuse activities, essentially "following the data" through its transformations.

Memory is a recognized limitation of interview data: participants reconstruct past events rather than recall them directly, meaning that accounts of data practices are shaped by how participants currently understand and narrate their work (Galletta & Cross, 2013; C. A. B. Warren, 2001). Walkthroughs of specific documented reuse instances were used in part to anchor accounts in concrete artifacts rather than relying on recollection alone.

Aspects of the interview protocol varied depending on the interviewee's role in the community. For example, when speaking with the leader of an ACC-skeptical organization that publishes instances of data reuse, I focused on how their organization approaches data reuse more broadly, since they had not conducted such work themselves in recent years. In contrast, when interviewing the author of a journal article who had conducted data reuse, I carried out a detailed walkthrough of the specific instance of data reuse to fill gaps left out in the published account.

To further document the data practices and knowledge infrastructure, I produced textual, and sometimes visual maps, of data journeys, a thick description of each journey that traces the

organizations, actors, data, and knowledge artifacts involved at each stage and how they evolved (Bates et al., 2016). This in-depth mapping was conducted for four cases, selected because they represented the most intensive instances of first-person data reuse in the interview corpus. Other interviewees were not selected for in-depth journey mapping for the following reasons: some described their organization's data reuse practices rather than their own, some served primarily as interlocutors with data reusers rather than reusers themselves, and others engaged primarily with knowledge artifacts produced from others' data reuse rather than with raw data directly. These four in-depth data journeys are documented in Chapter 4.

3.4.1 Recruitment and sampling

Recruitment for the interviews was based on identifying data reusers during participant observation and content analysis. Like other studies on data reuse practices, my sampling focused on data reusers identified through online records and publications (Faniel et al., 2016; Yoon, 2017; Zimmerman, 2008). At the time I stopped recruiting interviews, I had identified 86 data reusers, and I reached out to 72 of these practitioners for interviews. For the other 14 data reusers, I could not find publicly available contact information online. I also contacted two larger organizations that work with a variety of data reusers.

General appeals in large community forums were not pursued, as direct individual contact is generally more effective in digital ethnography (Hine, 2020). I reached out directly to individuals whose data reuse practices I had identified during participant observation and content analysis, explaining specifically why their perspective was valuable to the study and emphasizing the study's focus on understanding their data practices rather than critiquing their beliefs, adopting an open and non-judgmental approach. Six of the interviews were conducted virtually on Zoom. The interview times ranged from 45 minutes to 1.5 hours and were recorded with participant consent for transcription and further analysis. I offered a $50 incentive to participate although no participants requested the incentive.

Given the community's documented distrust of academic researchers, I framed the study as an attempt to critically engage with their perspectives on data reuse, presenting myself as willing to learn and listen. Recruitment, proved challenging, even with strategies in place. For instance, I engaged in snowball sampling, where initial interviewees recommend others within their community (Biernacki & Waldorf, 1981). Potential interviewees were provided but the process did not garner any additional interviews.

I tentatively aimed to conduct 20 to 25 semi-structured interviews, a range consistent with recommendations for qualitative research, that are focused on depth rather than breadth (Malterud et al., 2016). However, I was only able to conduct eight, as well as three pilot interviews with data reusers from information science (1) and earth science (2). Six of the eight interviews were conducted virtually on Zoom. Two were conducted via written questions, rather than synchronously, due to technology difficulties, time zone differences, language considerations, and scheduling constraints. These are noted in table 3.3 below. I use interviewee's real names here as they all wanted their interviews attributed to them. Job titles were given by the interviewees or identified through their websites.

Table 3.3: Overview of semi-structured interview participants, including interview format and role in the ACC skeptic community.

#	Name	Job Title	Interview Format	Date	Chapter 4 Case Study (Y/N)
1	Dave Burton	Software Engineer	Zoom	4/16/2025	Y
2	James Taylor	President of The Heartland Institute	Zoom	6/12/2025	N
3	H. Sterling Burnett	Director of the Arthur B. Robinson Center on Climate and Environmental Policy at The Heartland Institute.	Zoom	6/12/2025	Y
4	Tom Nelson	Podcaster	Zoom	6/13/2025	N
5	Ross McKitrick	Professor of Economics	Written	N/A	Y
6	Ferdinand Engelbeen	Retired Chemical Process Automation Engineer	Written	N/A	Y
7	Jim Steele	Retired Ecologist	Zoom	8/19/2025	N
8	Jonahtan Cohler	Physicist, Climate and AI researcher	Zoom	8/29/2025	N

Given the difficulty of securing interviews, I supplemented this method with additional data from the participant observation and content analysis conducted between September and December 2025. To further address the limited number of interviews, I incorporated publicly

available transcripts from the Tom Nelson Podcast, with permission from the interviewer, whom I also interviewed directly. These transcripts are not treated as equivalent to the semi-structured interviews conducted for this study. Nelson's questions differed substantially from the research protocol, and discussions emerged organically rather than in response to structured prompts. Several individuals I was unable to interview personally had participated in interviews with Nelson, and because discussions of data reuse frequently arose in those conversations, the transcripts provided useful supplementary material. These transcripts are public, available through YouTube and Apple Podcasts.

To reduce the transcripts to a manageable volume, I first narrowed the set from approximately 400 to 86 using a keyword search for "data." I then uploaded those 86 publicly available transcripts to Google NotebookLM through the University of Michigan and asked the tool to identify the most relevant remaining episodes. Although the transcripts were publicly available, I used a university-supported version of the tool and reviewed its status in the University of Michigan Sensitive Data Guide to IT Services before using it. The tool identified 19 community members across 24 episodes as especially relevant. I added these transcripts to the data corpus.

3.4.2 Member checking

As part of the interview process, I also incorporated member checking to strengthen the credibility of my findings and ensure that participants' perspectives were represented accurately (Birt et al., 2016; Candela, 2019). During the interviews, participants were sometimes asked to review and comment on preliminary interpretations developed from earlier phases of the research. For example, in one interview and another recruiting email, I asked whether the term anthropogenic climate skeptics, which I had used in previous publications, accurately described their community. Several participants expressed a preference for another term, such as anthropogenic climate crisis skeptics, which more closely reflected how they describe themselves. In response, I adopted this terminology to better represent participants' own framing. Although many participants might prefer the term climate realists, I *use* anthropogenic climate crisis skeptics* *because it more accurately captures the range of beliefs encompassed by this data community.

For similar reasons, I use the term mainstream climate science rather than alternatives such as consensus climate science, or the preferred ACC skeptic term, climate alarmists.

Participants often dispute the level of consensus around anthropogenic climate change, or exactly what consensus is measuring. No term is perfect, and members of the ACC skeptic data community would likely object to any terminology I use. Still, mainstream climate science more accurately captures the artifacts, institutions, and people against which this community defines much of its work.

Following interviews, I again engaged in member checking. In one instance, a participant explicitly requested to review how I planned to use his quotes in the dissertation, which provided an opportunity to test my findings. Other participants suggested I reach out if I had any questions, or wanted to run things by them, opening additional avenues for clarifying meaning and maximizing validity. Second, I conducted follow-up correspondence with select participants to share visual maps of data journeys and solicit feedback on whether the representations reflected their practices. These processes allowed participants to validate or correct my analysis, ensuring that their experiences and perspectives were accurately represented. Member-checking provided a reflective space for participants to reconsider their perspectives on data reuse, reinforcing the co-constructed nature of qualitative research, where both researcher and participant contribute to the meaning-making process (Lincoln & Guba, 1985). Moreover, as Saunders (2018) suggests, this iterative approach ensured that saturation was properly assessed, and participants' voices were faithfully represented. Member checking was targeted; I returned the factual accounts of the data journeys in Chapter 4 to the four data reusers, asking them to confirm that my reconstruction of what they did with the data was accurate. I did not share my analysis or interpretation of those journeys. Three of the four responded. Separately, one interviewee asked to review how his quotes would be used in the dissertation, and I shared the relevant passages with him. I did not return findings to the full set of interviewee participants.

At the same time, scholars have noted important limitations to member checking. Participants may misrepresent themselves, intentionally (Dean & Whyte, 1958; Y. Wang, 2024) or unintentionally, due to things like memory lapses, subjective perception, or miscommunication (Koelsch, 2013; Madill & Sullivan, 2018). We all tell stories about ourselves, and those stories may not always reflect actual practices. In such instances, I triangulated claims made during member checking with other data sources (Schwartz-Shea, 2020). When contradictions emerged, I treated them as analytically productive rather than as flaws, asking what these divergences reveal about how ACC skeptics want to be perceived versus how their

practices unfold in situ. This approach acknowledges both the performative dimension of self-reporting and the situated realities revealed through ethnographic observation. Rather than discarding inconsistencies, I analyzed them as part of the data, considering what they reveal about the politics of representation, credibility, and identity within the ACC skeptic community (Madill & Sullivan, 2018).

The small number of completed interviews, eight, supplemented by Tom Nelson transcripts, limits the range of perspectives captured through this method. Additionally, Interview accounts are retrospective and subject to reconstruction, selective emphasis, and the desire to present oneself favorably to a researcher (Galletta & Cross, 2013). The community's distrust of academic researchers may also have shaped what participants were willing to share. These limitations are why interviews function as one component of a triangulated design rather than the sole data source, with content analysis and participant observation providing independent evidence against which interview accounts were checked.

3.5 Data analysis

Data from the content analysis, participant observation, and semi-structured interviews were entered into NVivo and qualitatively coded using a unified codebook. During the initial phase of my project in 2021, I coded data inductively, drawing on grounded theory (Glaser & Strauss, 1967). I synthesized themes across different data sources and these themes were organized into overarching categories, gradually building theories grounded in the data (Charmaz, 1983). Six initial higher-level categories were derived from my first research question: What are the data practices of ACC skeptics?

- How do ACC skeptics discuss open data?
- How do ACC skeptics evaluate open data?
- How do ACC skeptics reuse data?
- How do ACC skeptics share data?
- How do ACC skeptics cite data?
- How do other members of the ACC skeptic community interact with the artifacts skeptics produce from the data?

A significant and initially unexpected finding from the first round of coding was that ACC skeptics both relied on mainstream climate science knowledge infrastructure and actively constructed alternative infrastructures of their own. This observation was not part of the original

pilot study design but emerged from the data in combination with my existing familiarity with knowledge infrastructure theory. It shaped how I extended the codebook: I added high-level codes for platforms, tools, data repositories, trust development, and epistemic values to document the knowledge infrastructures that enabled and constrained ACC skeptic data practices. These codes informed the second research question and the analytical framework developed in Chapters 5 and 6. The full codebook is included in Appendix B.

As the first round of inductive coding progressed, I moved into a second round of focused coding (Emerson et al., 2011). Where open coding is expansive and divergent, focused coding is targeted and convergent, using a restricted set of prioritized categories to conduct fine-grained analysis of topics identified as analytically significant (Emerson et al., 2011). This progression from open to focused coding is consistent with grounded theory's emphasis on moving from broad inductive exploration toward increasingly refined conceptual categories grounded in the data (Charmaz, 1983; Glaser & Strauss, 1967). In this study, the open coding phase generated the broad "How do they X" categories and the knowledge infrastructure codes described above. Focused coding then worked through those categories systematically, delineating subthemes, connecting pieces of data that initially appeared unrelated, and actively looking for exceptions and variations within each theme. In this round, I refined emerging practices using concepts from existing scholarship on data practices, data reuse, and other data communities. Foundational literature on grounded theory suggests using previous work and codes as a "source of questions and comparisons rather than as a measure of truth" (Charmaz, 1983, p. 117). I drew on Chao et al.'s Data Practices and Curation Vocabulary, or DPCVocab, (Chao et al., 2015), and the USGS data lifecycle model (Faundeen et al., 2014). When inductive codes did not align with the framework, I developed new terms. I applied a similar strategy for coding types of data reuse, drawing on papers like Pasquetto et al. (2019), and Thomer (2022). If instances of data reuse did not fit within these categories, new codes were created.

Where data practices were found to depend on or contest specific infrastructural arrangements, those arrangements were coded separately using the high-level infrastructure codes described above. This dual attention to practices and their infrastructural conditions allows the analysis to move between the data practices documented in Chapter 5 and the knowledge infrastructures examined in Chapter 6.

I applied this coding strategy across semi-structured interviews, online content, participant observation materials, and analytic memos using a unified codebook. The mechanics of coding varied slightly by source. Interviews and online content were coded line by line, while memos were coded incident by incident, attending to patterns across observation sessions rather than individual lines of text (Charmaz, 2006). Memos were also cross-linked with specific digital evidence from interviews, content analysis, and participant observation (Hine, 2020). Using a shared codebook allowed me to compare themes across data types while still attending to differences in how each type of evidence was produced. Semi-structured interviews filled gaps in the data journeys, supplementing publicly available information from content analysis and participant observation. This integrative approach under one codebook ensured a cohesive understanding of ACC skeptics' data practices and the knowledge infrastructures they engage with, enabling constant comparison and contrast across data types even if the mechanics of coding varied slightly (Charmaz, 2006).

This iterative process of coding continued as I integrated new data, returning to the analysis as needed to ensure saturation of the emergent themes. Saturation is the point at which no new information is emerging from additional data collection or analysis (Saunders et al., 2018). Archiving during content analysis followed embedded links and traced data wherever they led, which produced more material than proved analytically relevant or feasible to code, a recognized challenge in digital ethnography (Seta, 2020). I therefore coded a subset of the archived corpus, continuing until themes were saturated, rather than coding the full archive. The coded subset prioritized materials directly connected to the focal data journeys, frequently referenced artifacts, and episodes or posts where data practices were especially visible.

Throughout the coding process, I tried not to evaluate normatively whether participants' conclusions from data reuse were scientifically valid. The aim was to describe data practices in locally relevant contexts, situating them within the community's own epistemic frameworks (Haraway, 1988; Harding, 1992). A note on mainstream science applies throughout the finding's chapters that follow: where mainstream science appears, it is intended to orient readers to the infrastructure that ACC skeptics interact with and respond to, not to serve as a comparison case. This dissertation studied one community. Claims about how ACC skeptic practices differ from those of mainstream scientists are therefore beyond the scope of what the evidence supports.

The code “Data mistakes they make” was used in cases where practices diverged from the community’s own stated standards and epistemic norms, as well as those of mainstream climate science. The code “accusations of data misuse” was used where the community’s stated standards came into tension with the epistemic norms of mainstream climate science knowledge infrastructure. This is consistent with the misuse-as-infrastructural-friction framework developed in Chapter 2, which treats such designations as empirical objects rather than objective categories. Critique in this analysis emerges from juxtaposing practices with their consequences and contradictions, not from external normative judgment.

This integrated approach has limitations. There is a well-established tradition of analyzing different forms of qualitative data through distinct analytical frameworks. For example, qualitative interview analysis often emphasizes the subjective meaning-making co-constructed by interviewer and interviewee (DiCicco-Bloom & Crabtree, 2006). By contrast, content analysis and digital ethnographic observation may foreground patterns of public discourse, artifact circulation, and interactional context. A limitation of my integrated approach is therefore the potential loss of nuance that might come from analyzing each data format through its own distinct analytical lens. For instance, the situated context of semi-structured interviews may be partially diluted when considered alongside broader patterns identified through content analysis. However, applying a consistent coding strategy across data types allowed me to address the same core questions through each method, supporting thematic integration while maintaining attention to the distinct contexts in which each form of data was produced (Charmaz, 2006; Hine, 2020).

3.6 Addressing Positionality and Epistemology

My positionality, including my belief in the anthropogenic climate crisis, is a crucial aspect to my ethnography. Interpretive research argues that researchers should reflect on, reexamine, and analyze personal points of view and feelings as a part of the process of studying others, making values explicit within the research (Neuman, 2014). Looking at ethnography, in particular, following Sunder Rajan’s multisituated ethnography (Rajan, 2021), I understand ethnography as a situated practice—one shaped by my biographical, institutional, and political inheritances as a researcher. Rajan’s orientation, rooted in feminist and postcolonial frameworks, resists problems common to ethnography, both idealized intimacy with subjects and extractive objectification. Within my own work I aim for both responsibility to relationships I have

established within this community, but also for the potential impact of my work on broader society.

Feminist epistemology provided a useful framework for navigating and operationalizing these commitments investigating ACC's data practices in locally relevant contexts without leveling all knowledge claims. Haraway (1988) and Harding (1992, 1995) advocate for feminist objectivity and strong objectivity, respectively, both of which require "critical reexamination, decoding, deconstruction, and interpretation" of all knowledge claims (Haraway, 1988, p. 584). Feminist objectivity is centered around "limited location and situated knowledge…[as] only partial perspective promises objective vision" (Haraway, 1988, p. 583). Situated knowledges ground knowledge in specific and limited contexts. Strong objectivity is an inclusion of all standpoints, thus revealing those that are "less false" (Harding, 1995, p. 346). Harding suggests avoiding epistemic relativism by recognizing the social location of knowledge claims, using standpoint theory to transform social situatedness into a resource for objectivity, and employing reflexivity to critically evaluate both the subject and object of knowledge. Both perspectives avoid relativism by emphasizing accountability: knowledge is credible when its standpoint is made explicit and critically examined. Additionally, I rely on the idea of the social construction of knowledge (Latour & Woolgar, 2013): commonly accepted beliefs represent our best current knowledge but can shift when new information emerges. From this perspective, climate crisis skepticism, can be understood as having different probabilities of being true. How people reuse data to support their beliefs therefore remains an important topic of study, regardless of the factual accuracy of those beliefs. My goal is not to achieve impossible neutrality but to acknowledge how my positionality shapes the production and interpretation of knowledge.

Reflexivity is key to operationalizing this type of partiality. Holland (1999, p. 464) states that reflexivity involves "…methods that take into consideration the effect of the personality or presence of the researcher on the investigation." This requires continuous reflection on the choices made throughout the research process and how these decisions shape the results. Following the methodological commitments of data journeys, investigative digital ethnography, and strong objectivity, I practiced reflexivity. To put this into practice, I engaged in assumption audits throughout the research process (Friedberg, 2020). This involves:

- Documenting all assumptions that guide the initial research process.

- Revisiting these assumptions, original research questions, and topics of interest to ensure they remain relevant in light of ongoing discoveries.
- Adjusting initial assumptions when they no longer align with observations in the field.
- Engaging with divergent cases where special attention is given to outliers or cases that challenge initial assumptions, ensuring the analysis captures the full complexity of the data.
- Cross-referencing observations with other researchers to ensure the robustness of my analysis.
- Engaging in member-checking, where participants review and comment on my interpretations to ensure their perspectives are accurately reflected, providing an external check on my assumptions.

As Wagner et al. (2021) suggest, reflexivity involves reflecting not only on what is being studied but also considering the ethical implications of those decisions. This becomes crucial when analyzing politically charged data practices such as those of ACC skeptics, where the relationship between the data and its interpretation can carry social consequences. Reflexivity, in this case, includes taking responsibility for the outcomes of decisions made during this project. By maintaining a reflexive stance, I aim to balance the need for critical engagement with the commitment to represent ACC skeptics' data practices in locally relevant contexts.

3.7 Ethical considerations

Ethical practice in this project involves balancing multiple obligations: safeguarding participants' privacy, representing them as fairly as possible, protecting myself as a researcher, and ensuring that the findings are communicated responsibly. Examining ACC skeptic data practices is necessary for understanding the consequences of open science and the infrastructures that sustain it, as well as for creating more informed data services, reducing misuse, promoting trust, and fostering critiques that advance scientific progress. Given the risks of amplifying harmful narratives, I am attentive to how I represent ACC skeptics' practices, quoting selectively and contextualizing interpretations to avoid inadvertently legitimizing false claims. This research is informed by ethical guidelines specific to digital ethnography, emphasizing the complexity of the public-private spectrum and the importance of context-sensitive data handling (Cera, 2023; Gilbert et al., 2021; Luka & Millette, 2018; Reilly & Trevisan, 2016; Tiidenberg, 2018; Vitak et al., 2016; Zhao, 2024).

This study received Institutional Review Board (IRB) approval from the University of Michigan (HUM00205597) under exempt status. Consent was obtained from all interview participants, including preferences regarding anonymization, recording, and how their contributions would be attributed.

I adopted a framework that acknowledges the fluidity of the public-private divide, focusing on large digital spaces with low barriers to entry such as YouTube, Reddit, X, and blog sites (Cera, 2023). In these large public spaces, I chose not to announce my presence, although if I had entered a space with more barriers to entry or less than 50 members I would have done so. This dissertation follows three distinct attribution practices depending on data source. All interview participants consented to the use of their real names, which are used throughout. Second, a small number of individuals are named without having been interviewed. These are explicitly public figures and scholars whose relevant activity is a matter of published record and whose role in the history of the community is documented repeatedly in scholarly literature. Third, for individuals observed through content analysis and participant observation, codes are used even where those individuals are publicly known figures**.**I use two code series: ACC#### for the identified data reusers whose practices this study examines, including bloggers, podcasters, tool builders, and authors of skeptic publications, and COMM### for commenters who appear in blog, forum, and social-media threads, usually under screen names. Both series are anonymized; the distinction marks a participant's role and visibility in the community. This decision reflects both a commitment to privacy and precautionary measures for researcher safety, given the contentious nature of climate skepticism and the documented risk of online harassment for researchers in this area. Because these participants are active in open spaces and some are publicly known, coding cannot guarantee anonymity. A determined reader could, in some cases, identify an individual from publicly available material.

These attribution decisions reflect the character of the corpus. Of the 99 reusers I observed, a large majority are public figures and professional skeptics: people who publish under their own names, run organizations, maintain websites, host podcasts, or in some cases publish peer-reviewed research. Anonymous commenters and social media users were certainly present in my field site, but they made up a much smaller proportion of the active and prolific data reusers I observed, and I deliberately did not attempt to identify them for privacy reasons. The people who agreed to be interviewed were largely those with substantive and detailed reuse

practices they were willing to walk me through, and people willing to engage with an academic researcher, which is not common in this community and is why I supplemented interviews with other data collection methods.

One of the most ethically fraught aspects of this work is the challenge of representing a community whose epistemic claims I reject without either uncritically validating those claims or treating the community as unworthy of serious analytical attention (Cera, 2023; Pasquetto et al., 2026). The risk is collapsing into uncritical empathy on one side or dismissive detachment on the other. I resist both. Treating participants as fully human — as people who are curious, critical, and sometimes misled or misleading — is itself an ethical commitment that strengthens rather than weakens the integrity of the research. How I operationalize this stance is addressed in the positionality section above.

Aligning one's research with community expectations is an important part of understanding community-level impacts of digital research (Zent et al., 2025). However, in this project, striving for this alignment with the data community risked reinforcing potentially harmful narratives. I am not the first researcher to confront this dilemma. In one previous study, CSCW researchers, Jonas and Hanrahan propose a Designing for Shared Values (DSV) approach, as a conflict-avoidant method for values negotiation that consists of " (1) recognition of participant and researcher values, (2) consideration of impacts on indirect stakeholders, (3) evaluation of points of commonality between values, and (4) evaluation of depth of values" (Jonas & Hanrahan, 2022, p. 14). I attempted to operationalize this within the ACC community by focusing on aspects of the work where our values intertwined over things like transparency, documentation, and fostering constructive critique. At the same time, some of my design implications and recommendations, may still contradict community goals.

Balancing decision-making with my safety and privacy was also paramount. Given the contentious nature of climate change skepticism, I took measures to protect myself from potential online harassment or doxxing. This included using pseudonyms for my social media accounts during participant observation and content analysis and employing strict privacy settings. In addition, I consulted the Researcher Support Consortium for proactive measures to safeguard my research and well-being (Researcher Support Consortium, 2024b, 2024a). This included actions like using a Google Voice phone number and separate email to interact with participants, creating a list of people that would put out a statement of support if needed, and

removing personal information from websites, search engines, and data brokers. I also consulted PEN America's Online Harassment Field Manual and Social Media Audit for cybersecurity best practices (PEN America, 2019, 2023). These measures were essential to conducting the research safely.

3.8 Limitations and Risks

This study encompasses several limitations and risks that must be acknowledged to allow for mitigation, contextualize the findings, and guide future research. As a qualitative research project, it is highly situated to the data community of ACC skeptics and is not intended to be externally generalizable. Additionally, the reliance on semi-structured interviews, participant observation, and content analysis inevitably leaves gaps in understanding ACC skeptics' data practices. While these methods capture self-reported accounts and observable public interactions, they miss more private or less visible exchanges within the community. As such, the findings cannot claim representativeness of all possible variations in skepticism or data reuse.

To address validity, I draw on Guba and Lincoln's (1982, 1989) criteria for qualitative research, which emphasizes understanding, credibility, dependability, and confirmability. In practice, this includes prolonged engagement in the field, triangulation and structural relationships across data sources, member checking, peer debriefing, maintaining an audit trail, and attending to negative cases or outliers. These strategies align with different types of validity in qualitative research: descriptive validity (being truthful about what was observed), interpretive validity (accurately capturing what practices mean to participants), and theoretical validity (ensuring that conceptual categories and relationships are well grounded in the data) (Huberman & Miles, 2002). Because multiple data collection techniques were used, "the weaknesses of each can be counterbalanced by the strengths of others" through triangulation and structural relationships between data types (Newman & Benz, 1998, p. 66)

As a researcher who believes in the anthropogenic climate crisis, my positionality introduces potential bias. Concomitantly, there is also a risk of being manipulated by ACC skeptics during the semi-structured interviews. Skeptics may present convincing but biased or exaggerated information. To mitigate these risks, I rely on the practices such as assumption audits, member checking with participants, and triangulation across multiple data sources. These strategies do not eliminate the role of sociality on my knowledge production but instead render it explicit and accountable, strengthening the credibility and validity of the analysis.

The study also risks amplifying misleading and potentially harmful theories regarding climate change. I believe it is important to study what open science leads to in practice to actualize the promises and minimize the perils of open data, even if it is complicated. However, the magnification of these views, even through a critical examination, could inadvertently lend them legitimacy or visibility (Phillips, 2018). Nonetheless, given the prominence of climate crisis skepticism, ethnographic analysis of their data practices is essential to understand how they oppose scientific consensus through the reuse of open data, as well as accurately evaluate existing or proposed management and mitigation techniques. Reuse of open research data in ways that repositories, intermediaries, or science may not agree with cannot be avoided. However, this reality should not be used as an excuse not to release critical data or to avoid taking the potential repercussions caused by this reuse seriously.

Taken together, these limitations and risks highlight the difficulty of studying politically charged data practices. Yet they also underscore the importance of such research: to critically examine how open science infrastructures are used in ways that both fulfill and complicate their promises.

Chapter 4 The Data Community and Four Data Journeys

This chapter traces four data reuse journeys drawn from the ACC skeptic community, documenting how individual community members interact with open data. Before introducing the four subjects, it is useful to situate them within the broader community this dissertation observed.

4.1 The Data Community

Across my monitoring environment, I observed approximately 99 data reusers. The demographics described below therefore reflect the publicly identifiable segment of the community. Most identifiable public figures were conservative in their political orientation, based on their public statements and organizational affiliations. Based on visual perception and publicly available information rather than direct self-identification, the large majority appeared to be male, white, and over 50. They were predominantly based in the United States, Canada, Australia, and the UK, partly reflecting the limits of my English-language research, but climate skepticism is particularly visible in the anglosphere.

Generally, reusers were highly educated, with many stating they have PhDs. Their training was concentrated rather than evenly spread: the heaviest clusters fell in atmospheric and earth sciences, physics, and engineering, with chemistry, economics, political science, and law also well represented. These backgrounds were embedded in a recurring set of advocacy, policy, and research organizations. Many reusers held positions at bodies such as the Heartland Institute, the Competitive Enterprise Institute, the Cato Institute, the CO2 Coalition, Clintel, and the international Climate Science Coalition. Some also had professional ties to industries historically resistant to climate regulation, including oil, gas, coal, and in some cases tobacco. And importantly, their data literacy was not limited to data related to the climate. Many of the same actors pivoted to reanalyzing COVID-19 datasets, 2020 U.S. presidential election data, and other public datasets during my ethnography. They often developed and honed these data skills

through this ACC skeptic community but applied them to other issues. They were not only climate skeptics; they were broadly oriented to audit public data and institutional claims.

Some participants initially pushed back on the demographic patterns I described. As conversations unfolded, however, several offered their own explanations. As one participant put it:

> The only people who are not afraid to tell the truth are people who are older, and that's because they're either retired or close to retirement or untouchable in their position, have already done their hundreds of papers of publishing, they don't need a postdoc, they don't need another job, et cetera (Cohler, Interview, 2025).

Many within this community saw their own experts not as anti-science outsiders, but as experienced insiders who feel protected enough, by tenure, retirement, or career status, to challenge mainstream science precisely because they have less to lose professionally.

4.1.1 Pathways into ACC skepticism and data reuse

For some, engagement with climate data grew out of professional expertise. A retired engineer with decades of work in optics and spectroscopy described beginning his own "independent research into climate change in 2007," focused on "the calculation of the surface temperature from first principles" (ACC049, Nelson Podcast, 2025). Jonathan Cohler, who holds a BA in physics from Harvard, mentioned that his physics and software backgrounds go back nearly 50 years to when he worked developing "finite element analysis" software, which, according to Cohler, is "actually a precursor" to climate modeling software (Cohler, Interview, 2025). For others, their discipline supplied a standard of rigor they brought to climate work: one physicist recalls finding climate and environmental studies even in the early 1990s, "weren't normal science," and questioned why they "should have different standards [from] high energy physics or fusion," fields he describes as "fairly rigorous" and "open to questions" (ACC050, Nelson Podcast, 2022). This pattern is not confined to the physical sciences. ACC051, trained in law and criminology, recounted that a discrepancy between tide-gauge readings and reports of rising temperatures "clicked" his "legal mind…into action" (ACC051, Nelson Podcast, 2023). One investigative journalist described entering ACC skepticism "as a journalist…an investigative, inquiring journalist," drawn by what he saw as "fantastic stories" that the mainstream press ignored "under the settled science narrative" (ACC037, Nelson Podcast, 2025).

For others, the entry point into ACC skepticism research, were specific observations that mainstream climate science could not account for in their opinion. Jim Steele, an ecologist who directed a university field station for decades, recounted a meadow ecological system that collapsed during his tenure (Steele, Interview, 2025). Others attributed the changes to climate change, but as he:

> Walk[ed] that whole system several times, talking to different people, what became obvious it was a hydrology problem created by a railroad track a hundred years earlier that was downgrading and draining the water table. And that's why that meadow system was collapsing. So [he] got grants and [they] fixed it, and birds came back better vegetation came back better than ever. (Steele, Interview, 2025).

He came to question how averaging minimum and maximum temperatures could, in his account, produce "a false sense of what's happening," and described the experience as having "prompted [him] to look at a bunch of other stuff" (Steele, Interview, 2025). He asserted "the whole purpose of [his work] is to make a more sound environment. Humans screw it up in different ways, but it wasn't climate change, it was the way we degraded some of the landscapes" (Steele, Interview, 2025). Similar accounts recur: one blogger said he accepted global warming until 2003, when heavy snowfalls struck him as at odds with forecasts that snow was becoming rare (ACC017, Nelson Podcast, 2022), and a data management professional described auditing official temperature records after a headline proclaimed a record-hot weekend that was recorded as cold (ACC016, Nelson Podcast, 2025).

Many discussed a standing disposition to test public claims against data. Tom Nelson described "checking into each one of these claims on their own, one at a time, concluding, in his characterization, that "not one single scare story holds up if you look at the real data" (Nelson, Interview, 2025). Many of podcast guests described the same orientation. For instance, ACC052 described himself as a "citizen scientist because [he] dig[s] into the problems" (ACC052, Nelson Podcast, 2025).

Many also described disillusionments with institutions or movements they had once trusted. Some narrated this shift from inside the environment movement. ACC053 a co-founder of an environmental organization, is invoked as someone who left the organization after concluding that it had moved away from science and toward ideology. ACC054 similarly framed his trajectory describing himself as involved in the environmental movement "back in the late

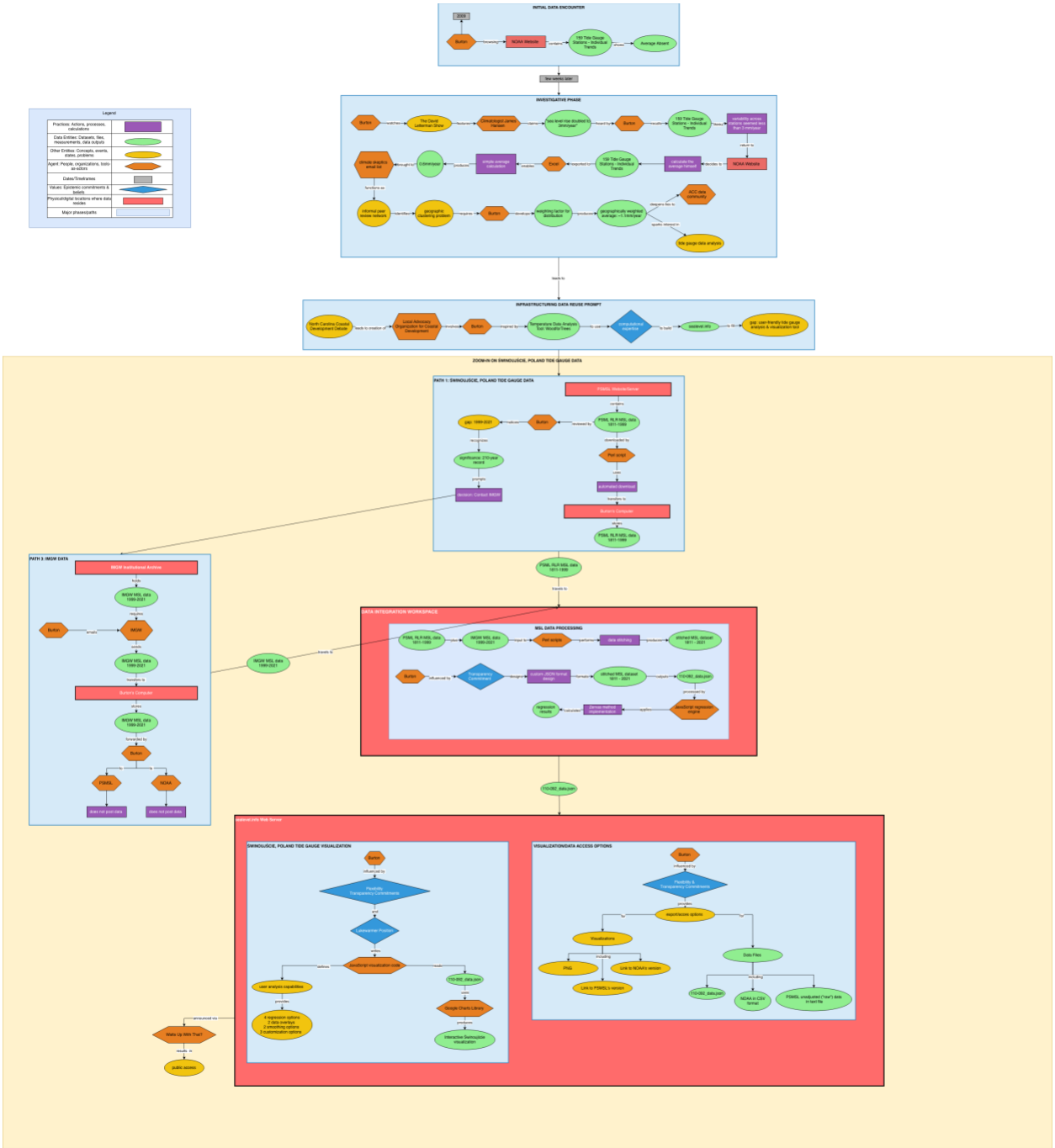
70s," where he had "seen it all," including "every single scare you can imagine" (ACC054, Nelson Podcast, 2025). Over time, he said he came to believe that some of those scares were "totally exaggerated" and that he had been "exposed to exaggerated fear mongering" (ACC054, Nelson Podcast, 2025). Others described disillusionment from inside the classroom. One skeptic recounted a school environment in which anthropogenic climate change "was snuck in all over the place" and "all that we ever talked about" (ACC055, Nelson Podcast, 2023). These background and pathways converge on a shared disposition to examine climate data directly, which the four journeys that follow document in depth.

This chapter documents four data reuse journeys in detail, following each from the initial prompt that set the reuse in motion through acquisition to publication. I chose these journeys to detail because they were the clearest data reuse instances that came out of the content analysis and were supplemented by semi-structured interviews to create the most complete and accurate journeys possible. I first reconstruct each journey as its participants would describe it, then offer my own analysis. Where mainstream climate science appears, it orients readers to the infrastructure these reusers engage with and push against, this dissertation studied one community and claims about how ACC skeptic data practices differ from those of mainstream climate scientists are beyond the scope of what the evidence here supports.

The names that follow are the real names of the data reuse journey subjects, as each wanted to be identified publicly. Dave Burton is an IT professional whose engagement with sea level data led him to build and maintain sealevel.info, a parallel data aggregation and analysis tool that reuses NOAA and PSMSL tide-gauge records, for over fifteen years. H. Sterling Burnett is a policy director, editor, and podcaster at the Heartland Institute, whose data reuse journey is rapid and reactive, oriented toward quickly contesting a specific newspaper article. Ferdinand Engelbeen is a retired process chemist and decades-long participant in ACC skeptic online communities, whose data reuse journey takes the form of intra-community quality assessment, applying skeptic epistemological standards to claims made by fellow skeptics. Ross McKitrick is a tenured economics professor at the University of Guelph and Senior Fellow at the Fraser Institute, whose data reuse involves quality assessment and replication directed at policy-relevant empirical claims. Together, these journeys reveal patterns across the full arc of data reuse practices, which I analyze within each section and develop further across my full data corpus in Chapters 5 and 6.

4.2 Infrastructuring: Building Parallel Data Systems (Dave Burton — sealevel.info)

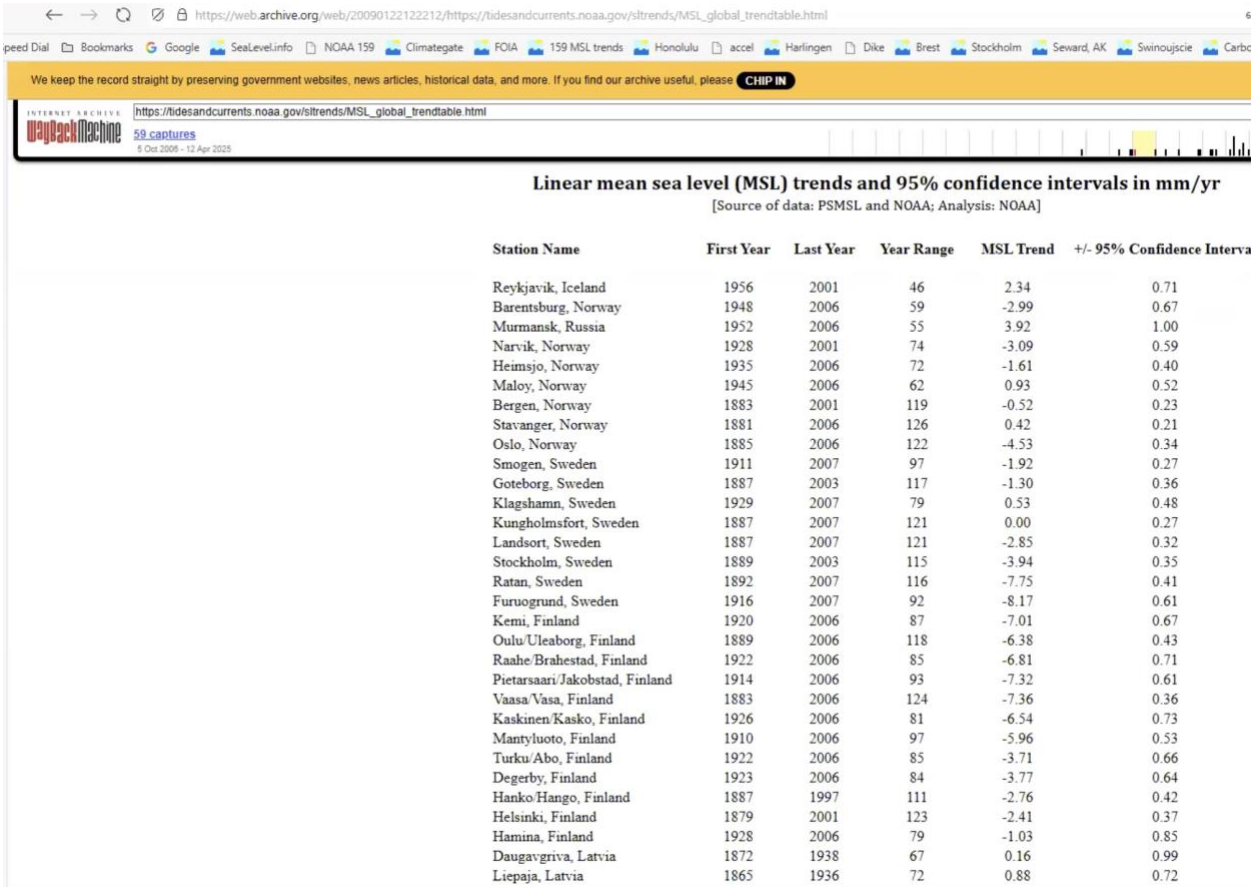
Figure 4.1: The data reuse journey map of Dave Burton's SeaLevel.info.



4.2.1 Data reuse prompt and project design

In 2009, Dave Burton, an IT professional with an undergraduate degree in systems science and a master's degree in computer science, stumbled upon a NOAA website listing 159 long-term tide gauge stations as shown in image 4.1 below. At first, Dave did not think much of it and explored the site casually, noticing some trends at individual stations. Curious, he "scrolled to the bottom" of the page, "looking for the average just out of curiosity," but found "there's just no average" (Burton, Interview, 2025). He wondered, "Why no average?" but soon "just kind of went on with whatever [he] was doing" (Burton, Interview, 2025).

Image 4.1: Screenshot of NOAA's linear mean sea level (MSL) trends page, as archived by the Wayback Machine. Burton went to this site during his semi-structured interview to illustrate the page he first encountered in 2009.



Linear mean sea level (MSL) trends and 95% confidence intervals in mm/yr

[Source of data: PSMSL and NOAA; Analysis: NOAA]

Station Name	First Year	Last Year	Year Range	MSL Trend	+/- 95% Confidence Interva
Reykjavik, Iceland	1956	2001	46	2.34	0.71
Barentsburg, Norway	1948	2006	59	-2.99	0.67
Murmansk, Russia	1952	2006	55	3.92	1.00
Narvik, Norway	1928	2001	74	-3.09	0.59
Heimsjo, Norway	1935	2006	72	-1.61	0.40
Maloy, Norway	1945	2006	62	0.93	0.52
Bergen, Norway	1883	2001	119	-0.52	0.23
Stavanger, Norway	1881	2006	126	0.42	0.21
Oslo, Norway	1885	2006	122	-4.53	0.34
Smogen, Sweden	1911	2007	97	-1.92	0.27
Goteborg, Sweden	1887	2003	117	-1.30	0.36
Klagshamn, Sweden	1929	2007	79	0.53	0.48
Kungholmsfort, Sweden	1887	2007	121	0.00	0.27
Landsort, Sweden	1887	2007	121	-2.85	0.32
Stockholm, Sweden	1889	2003	115	-3.94	0.35
Ratan, Sweden	1892	2007	116	-7.75	0.41
Furuogrund, Sweden	1916	2007	92	-8.17	0.61
Kemi, Finland	1920	2006	87	-7.01	0.67
Oulu/Uleaborg, Finland	1889	2006	118	-6.38	0.43
Raahe/Brahestad, Finland	1922	2006	85	-6.81	0.71
Pietarsaari/Jakobstad, Finland	1914	2006	93	-7.32	0.61
Vaasa/Vasa, Finland	1883	2006	124	-7.36	0.36
Kaskinen/Kasko, Finland	1926	2006	81	-6.54	0.73
Mantyluoto, Finland	1910	2006	97	-5.96	0.53
Turku/Abo, Finland	1922	2006	85	-3.71	0.66
Degerby, Finland	1923	2006	84	-3.77	0.64
Hanko/Hango, Finland	1887	1997	111	-2.76	0.42
Helsinki, Finland	1879	2001	123	-2.41	0.37
Hamina, Finland	1928	2006	79	-1.03	0.85
Daugavgriva, Latvia	1872	1938	67	0.16	0.99
Liepaja, Latvia	1865	1936	72	0.88	0.72

A few weeks later, a television appearance by the climatologist James Hansen prompted him to return to the data. During an episode of *The Late Show with David Letterman*, Hansen stated that sea level rise had doubled to 3 mm/year. Dave thought it was peculiar how Hansen "talk[ed] like this one number, that it's the same everywhere" (Burton, Interview, 2025). Reflecting, he recalled "all these different numbers" from NOAA's website, thinking the

3mm/year “seems like the high end of the range of those numbers [he] was seeing” (Burton, Interview, 2025). This discrepancy prompted him to revisit NOAA’s website and conduct further investigation.

The account that follows describes sealevel.info as I observed it over the course of this study, from September 2021 through December 2025, with the site's specific holdings, links, and interface features last verified in February 2026. sealevel.info is under continuous revision by Burton, and later states of the site may differ from what is described here.

4.2.1.1 Initial analysis and community engagement

Dave exported the station trend data to Excel and found the simple average calculation “trivial.” He “just collected the column and did the average…and it was 0.6[mm]/year,” much smaller than Hansen’s claim (Burton, Interview, 2025). He brought his finding to a climate skeptics email list - it is unclear from the interview whether Burton was already a member of this list or joined around this time - curated by a correspondent in Denmark, asking: “Anybody have any ideas what’s going on here?” (Burton, Interview, 2025). A community member immediately identified a methodological problem: “‘These tide stations are not uniformly distributed,’” and thus a simple average would not work (Burton, Interview, 2025).

Recognizing that clustered gauges would be overrepresented in any unweighted mean, Dave devised a more sophisticated approach. He “did some analysis on the distance that tide gauges were from one another and looked for how well the trends correlated between them,” ultimately developing a weighting factor to produce a geographically weighted average of the tide stations, “to try to get rid of that problem of clustered tide stations getting multiplied” (Burton, Interview, 2025). The resulting weighted average was approximately 1.1mm/year, still considerably lower than Hansen’s number.

Beyond the weighted average, Dave also examined NOAA’s graphs for individual tide stations directly. He found that they “didn’t look like there was any of them where the trend had doubled, maybe one or two,” and concluded that the sea level trend simply did not appear to be doubling at these stations (Burton, Interview, 2025). This combination of a recalculated average and a visual survey of station-level records formed the basis of his skepticism about Hansen’s claim.

4.2.1.2 From Investigation to Infrastructure: NC-20 and the Decision to Build

Dave's initial investigation might have remained a private exercise had external events not given it new stakes. North Carolina, his home state, became the site of a political debate over proposed coastal development regulations premised on accelerated sea level rise projections. Dave became involved with NC-20, an organization opposing those regulations, and found that the questions he had been asking informally suddenly had direct policy relevance. As he describes it: "We were off to the races. It's been ever since. And I'm sort of like a dog with a bone. I just don't let things go" (Burton, Interview, 2025). This involvement transformed a fleeting curiosity into a sustained commitment, and it was out of this context that the decision to build sealevel.info emerged.

His technical background made the construction feasible. As he explained, "I'm an IT guy. My undergrad degree was in systems science, and then my graduate degree was in computer science. And so, coding is, it's my bailiwick. And so, I wrote a bunch of code" (Burton, Interview, 2025). The site he ultimately built provides "interactive regression analysis (line/curve fitting) and visualization (graphing) tools available for mean sea level (MSL) measurements from over 1200 tide gauges, plus spreadsheets which combine various subsets of that data" (Burton, 2017). Figure 4.1 traces this full data journey, from Burton's initial encounter with the NOAA tide gauge page through the data acquisition, integration, and publication phases, with particular focus on the Świnoujście tide gauge as an illustrative case.

4.2.2 Seeking and selecting data

Burton integrates multiple data sources on the backend of his tool. His primary sources are the Permanent Service for Mean Sea Level data (Holgate et al., 2012; *PSMSL*, 2026) and the NOAA Center for Operational Oceanographic Products and Services' unadjusted (NOAA CO-OPS, 2026) and seasonally adjusted data (NOAA Center for Operational Oceanographic Products and Services, 2026). These citations link to what is provided by Dave Burton on sealevel.info. He described PSMSL as "the number one go-to for sea level data," noting that it aggregates records from national authorities, including NOAA (Burton, Interview, 2025). However, PSMSL, according to Burton, updates only approximately once per year, meaning its holdings are somewhat out of date. NOAA, by contrast, provides more frequently updated records for many U.S. stations and offers direct API access, which Dave uses to supplement PSMSL data when recency matters. These two repositories are not perfectly aligned: they cover

overlapping but not identical geographic ranges, with PSMSL drawing on a broader global network of national authorities while NOAA focuses on U.S. stations**.** Dave therefore sources from both, trying to stich together the most complete and current record available for each station.

Addressing data gaps requires going beyond these standard repositories. The Świnoujście, Poland tide gauge illustrates this most clearly. PSMSL holds Świnoujście records stretching back to 1811, but its holdings for the station end around 1999, leaving over two decades of data absent from their record. Dave subsequently reached out to the Polish Institute of Meteorology and Water Management directly by email (*IMGW-PIB – **Instytut Meteorologii i Gospodarki** Wodnej*, n.d.). The record's temporal length made the gap too consequential to leave unfilled. As he explained:

> I actually reached out and contacted them directly and said, 'Could you possibly send me some data?' And I did that. I've done that a couple of times. Last time was two, three years ago, and I need to do it again because we're getting a little out of date (Burton, Interview, 2025).

IMGW provided data covering approximately 1999 to 2021, filling the gap left by PSMSL. The data he received were not incorporated by PSMSL or NOAA, though he had "forwarded the data to [them] and…thought they would probably use it," but they never did.

Beyond sea-level data, the tool incorporates atmospheric measurements to enable users to visually compare tide-gauge trends with greenhouse gas concentrations. The data sources and their citations are as follows. Current CO_2 levels come from the Mauna Loa Observatory (Lan & Keeling, 2026); historical CO_2 data prior to 1958 are drawn from NASA ice core records (NASA GISS, 2007); and data from 1800 to 1850 come from the Law Dome ice core series archived at NOAA's National Center for Environmental Information (NCEI) (D. M. Etheridge et al., 1996, 2001). CH_4 measurements follow a similar structure, with ice core records from the Carbon Dioxide Information Analysis Center (CDIAC) data archive covering 1840 to 1992, but the link Burton provided as of February 2026 for CDIAC (CDIAC, n.d.) was no longer available because the CDIAC data archive had transitioned to ESS-DIVE. The dataset is now archived at ESS-DIVE (D. Etheridge et al., 2002), although Burton has copied over the interpolated dataset to sealevel.info with the original Etheridge citation (Burton, 2026a). The Mauna Loa Observatory CH_4 measurements cover 1984 to the present (Lan, 2026).

Dave also provides subsets of tide gauge data organized according to lists compiled by other scientific organizations and individuals. In total, he provides 15 different subsets of tide gauge data. This includes those who "have made their collections of what they consider the important tide gauges" as well as subsets of NOAA data (Burton, Interview, 2025). The subsets are as follows.

1. NOAA's 2015/2016 list of 375 Long Term Trend (LTT) tide stations
2. All 1269 available tide stations
3. Dr. Jerry Mitrovica's 2001 list of 23 LTT stations w/ minimal vertical land motion
4. Dr. Simon Holgate's list† of nine excellent tide gauge records of sea-level measurements
5. NOAA's 2013 list of 285 LTT tide stations
6. NOAA's 2013 list, excluding stations with less than 50 years data
7. NOAA's 2013 list, excluding stations with less than 60 years data
8. NOAA's 2012 list of 239 LTT tide stations
9. NOAA's 2010 list of 159 LTT tide stations
10. Combined NOAA 2010 & 2012 lists of 239+16=255 LTT tide stations
11. GLOSS 1997 list of 187 GLOSS-LTT tide stations
12. NOAA's 2010 list of 45 U.S. LTT tide stations
13. NOAA's 2010 list of 42 U.S. LTT tide stations w/ data to 2011
14. NOAA's 2010 list of 31 U.S. LTT tide stations in the contiguous USA
15. NOAA's "sweet 16" tide stations, dropped in 2012

4.2.3 Processing and transforming data

A feature of the confidence interval calculations is the handling of serial autocorrelation. Monthly sea-level measurements are correlated with those of adjacent months, meaning a measurement in one month is related to the months immediately preceding and following it. Subsequently, "a naive confidence interval calculation" would "underestimate the width of confidence intervals" by treating each month as an independent observation (Burton, 2026d).

> The code "follows the method of Zervas 2009 (Zervas & Center for Operational Oceanographic Products and Services (U.S.), 2009), "Sea Level Variations of the United States 1854-2006,"NOAA Technical Report NOS CO-OPS 053 (or here [link to NOAA report archived at sealevel.info]), p. 15-24, to account for autocorrelation, when calculating confidence intervals and prediction intervals" (Burton, 2026d).

Dave notes in sealevel.info documentation that the Zervas method to handling autocorrelation is not the only valid approach and exploring alternatives is on his to-do list (Burton, 2026d)

For the Świnoujście gauge specifically, processing required stitching together two data paths before any analysis could occur. PSMSL data covering 1811 to approximately 1999 and IMGW data covering approximately 1999 to 2021 were integrated through Perl scripts that handle data stitching across sources. To my knowledge, these scripts are not publicly shared. The combined record, stored as a JSON file, provides a continuous sea level time series for Świnoujście spanning over two centuries, as shown in Image 4.2.

4.2.4 Managing, documenting, and preserving data

Data management within sealevel.info is handled through a combination of automated Perl scripts and manual curation. He has "scripts running automatically that update the PSMSL and NOAA stuff" (Burton, Interview, 2025). Version control is managed by TLIB, software created by Burton under his company Burton Systems Software, which tracks modifications and updates the "last modified" date displayed on web pages (Burton, 2021). Dave makes sea level data available to download in three primary formats, each serving different analytical needs and preserving different levels of contextual information.

4.2.4.1 JSON data file

The JSON files are Dave's own creation, integrating data from multiple sources with extensive metadata (Burton, 2023). Each JSON file (e.g., 110-092_data.json for Świnoujście, Poland) functions as a self-contained data package designed to support both programmatic analysis and transparent documentation of data provenance (Burton, 2026b, 2026c). Station-level metadata includes multiple identifier systems (NOAA ID, PSMSL ID, and PSMSLS's legacy coast-station code), geographic coordinates (latitude and longitude), human-readable locations, data availability flags (Boolean indicators for whether data exist in each repository), temporal coverage, and data source credits. Dave also includes NOAA's published trend analysis, with a caveat in the readme noting that these values are "often obsolete" because NOAA doesn't update them frequently (Burton, 2023).

There is also explicit documentation of seasonal cycle information, with each file including an array representing the average monthly deviation from the annual MSL. The JSON also documents how the seasonal signal was derived, the seasonal_cycle_source field. The value

can be a “NOAA (file)” for signals read directly from NOAA, “NOAA (implied)” when Dave calculated the season cycle by comparing NOAA’s raw and seasonally-adjusted data, “NOAA (consistent)” when the implied and file-based signals match, NOAA (implied*)” meaning Dave computed it by comparing NOAA’s raw and seasonally-adjusted data, and the asterisk indicates his computed result differs slightly from NOAA’s own published seasonal cycle file, and “Computed” when Dave calculated the seasonal cycle himself because NOAA had not provided one (Burton, 2023). For the Świnoujście gauge, which draws on IMGW data not held by NOAA, the seasonal cycle derivation falls into the latter category, computed by Burton directly rather than drawn from an official source.

The time series data itself is structured as an array (msl_data) where each monthly record contains 14 fields (Burton, 2023). These include the following:

1. Year (AD)
2. Month (1–12)
3. NOAA_adjusted_MSL: NOAA’s seasonally-adjusted monthly mean sea level in meters (or null if missing)
4. NOAA_unverified: Flag for unverified NOAA data (appears to be null in current practice, possibly deprecated)
5. NOAA_linear_trend: NOAA’s linear regression value for this date
6. NOAA_high_conf: Upper 95% confidence interval
7. NOAA_low_conf: Lower 95% confidence interval
8. PSMSL_rlr_MSL: PSMSL’s unadjusted monthly MSL in millimeters (currently -99999 for missing data, though Dave notes this may change to null)
9. PSMSL_num_missing_days: Number of days missing from PSMSL’s monthly record
10. PSMSL_dataflag: PSMSL’s 3-character quality flag (e.g., “000”)
11. NOAA_raw_MSL: NOAA’s unadjusted monthly MSL in meters
12. NOAA_interpolated: Flag indicating whether NOAA data is interpolated (appears as null when not interpolated)
13. msl_raw: Dave’s combined unadjusted MSL from all sources, in meters
14. msl_adj: Dave’s combined seasonally-adjusted MSL, in meters — the field users typically want for analysis (00readme.txt)

Dave emphasizes in the documentation that fields 13 and 14 represent his integration work: "where we can, we use NOAA_adj_MSL to fill in gaps in NOAA_raw_MSL, and vice-versa, by adding or subtracting the seasonal signal, respectively" (Burton, 2023). He notes with some surprise: "You wouldn't think there would be such inconsistencies, but there often are" (Burton, 2023). Missing months are represented with year and month values but null for all data fields, ensuring there are never gaps in the monthly sequence. Dave explained this design choice: "We include null/undefined values because if we don't include the null values then Google Charts will interpolate them, which would make a misleading graph" (Burton, 2023).

4.2.4.2 NOAA seasonally-adjusted CSV

The second format is NOAA's official seasonally-adjusted data file (e.g., 110-092_meantrend.csv for Świnoujście, Poland) (NOAA CO-OPS, n.d.). Rather than hosting a copy of this file, sealevel.info provides a viewer script that fetches the file directly from NOAA's tidesandcurrents.noaa.gov server and displays it in the browser. The data itself remains on NOAA's infrastructure. The file begins with a four-line plain-text header explaining that it contains "Monthly mean sea levels with the average seasonal cycle removed," that values are "in meters relative to the most recent Mean Sea Level datum established by CO-OPS," and that columns represent "Year; Month; Monthly Mean; Relative Sea Level Trend Line; Higher 95% Confidence Interval; and Lower 95% Confidence Interval." Attribution is provided: "Product of NOAAs National Ocean Service / Center for Operational Oceanographic Products and Services (CO-OPS)" (CSV file). What is notably absent: there is no station metadata in the file itself, no coordinates, no station name, and no identifier cross-references (CSV file).

4.2.4.3 Unadjusted/raw data files

The third format is the raw (unadjusted) monthly measurements file (e.g., 2.rlrdata for Świnoujście, Poland) (PSMSL, n.d.-c). These files can be from NOAA or PSMSL depending on the tide gauge and contain different metadata depending on the source. In the case of Świnoujście, Poland, this file is a plain delimited text (semi-colon separated). Each row is one month of sea level data. However, the fields are not labeled.

Beyond the data files themselves, each station page includes direct links to NOAA's station-level resources so users can access instrumentation details, photographs of the gauge,

precise coordinates, elevation data, and measurement history (Burton, Interview, 2025). As Dave explained:

> If you're interested in what's going on with the sea level at San Francisco, you might want to go see NOAA's page about San Francisco… If you go there, then you can go to their station info and there's a picture of it, and there's the latitude and longitude and the elevation and et cetera (Burton, Interview, 2025).

4.2.4.4 Note on the differences between dataset documentation:

The choice of documentation for each dataset has significant implications for what contextual information travels with the data. NOAA's seasonally adjusted CSV file does not have any station-level context (NOAA CO-OPS, n.d.). A user analyzing 110-092_meantrend.csv must already know this is Świnoujście, Poland, or know that 110-092 is the station ID, and must obtain coordinates, data provenance, and quality information from an external source. The same is the case for 2.rlrdata (PSMSL, n.d.-c). In fact, although the .rlrdata file looks self-contained, it is not and usually is downloaded in a folder with various documentation files. For both the NOAA CSV and .rlrdata file, if a user is navigating from NOAA's Tide and Currents site, PSMSL's database, or Dave's sealevel.info, they would be able to find this information. But, if they were given either file without context, it would be difficult to determine this information.

Dave's JSON files, by contrast, are designed to be self-documenting, embedding attribution, multiple identifier systems (allowing cross-referencing with NOAA and PSMSL), seasonal cycle derivation methods, data availability flags, and more. This design aligns with his stated principle that users who are "interested in the data" are presumably "interested in that stuff," the infrastructural context that makes the data interpretable (Burton, Interview, 2025). The JSON format makes Burton's own processing adjustment decisions easily reviewable, allowing users to see both the raw and adjusted values alongside the exact seasonal correction that was applied.

This multi-format strategy reflects Dave's dual commitments: preserving the authority of NOAA's official data (by linking directly to their CSVs without modification) while augmenting it with the contextual metadata and cross-repository integration. This linking strategy treats the JSON metadata and station pages as gateways to deeper infrastructural context rather than as self-contained data products.

4.2.4.5 Limitations on managing, documenting, and preserving data

This commitment is not without complications. Due to the extensive nature of the tool, and the limitations of Burton as the single maintainer of the site, Burton acknowledges that for many of the tide gauges, he has "never even looked at the graphs" and that "doing regression calculations and generating graphs without manual 'sanity checks' has pitfalls" (Burton, 2017). He gives the following examples of what this means in practice:

> For instance, some of the MSL data at some locations has been identified by PSMSL and/or NOAA as being of questionable accuracy, and these tools do not currently take that into account. (It's on the to-do list.)
>
> Also earthquakes can cause sometimes-extreme distortions in the trend. Consider the case of Seward, AK: March 27, 1964 was the date of the Great Alaska Earthquake (magnitude 9.2!). Seward got a full meter of sea-level rise, all at once. So if you're going to calculate the MSL trend for Seward, you should start with the data after that date. like this. Note that NOAA did that on their web page for Seward (Burton, 2017).

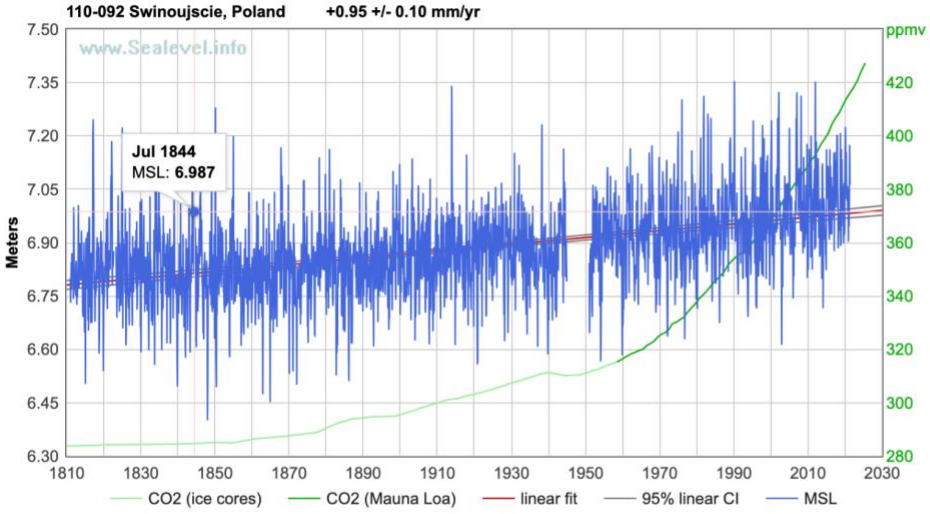
In response to this potential missing information within this tool, Dave recommended the following:

> So, when using these tools to analyze sea-level data, I recommend that you also click on the NOAA and PSMSL links, to view their pages about the tide gauges of interest, and I recommend that you look in particular for warnings about questionable data, apparent datum shifts, and earthquakes (Burton, 2017).

4.2.5 Analyzing, interpreting, and abstracting data

Image 4.2: Screenshot of sealevel.info's mean sea level visualization and analysis tool for the Świnoujście, Poland tide gauge station, showing the site's regression analysis and display options (Burton, 2026b).

Mean Sea Level at Swinoujscie, Poland (NOAA 110-092, PSMSL 2)



The mean sea level (MSL) trend at Swinoujscie, Poland is +0.95 mm/year with a 95% confidence interval of ±0.10 mm/year, based on monthly mean sea level data from 1811/1 to 2021/5. That is equivalent to a change of 0.31 feet in 100 years. (R-squared = 0.221)

The plot shows the monthly mean sea level without the regular seasonal fluctuations due to coastal ocean temperatures, salinities, winds, atmospheric pressures, and ocean currents. By default, the long-term linear trend is also shown, in red, along with its 95% confidence interval. The plotted values are relative to the most recent Mean Sea Level datum established by NOAA CO-OPS.

Options

- [x] Show calculated linear trend (linear regression)
  - [x] Show linear trend 95% confidence intervals†
  - [ ] Show linear trend 95% prediction intervals†
- [ ] Show quadratic regression
  - [ ] Show quadratic trend 95% confidence intervals†
  - [ ] Show quadratic trend 95% prediction intervals†
- [ ] Show NOAA's linear trend
  - [ ] Show NOAA's linear trend 95% confidence intervals
- [x] Show CO2 levels

Boxcar smoothing: [ ] Enable Width: 3 Repeat: 1

Line weight: ○ thin ◉ medium ○ thick color scheme: 1

Graph from:	start date	1811	1	thru end date	2029	12
Calculate using:	start date	1811	1	thru end date	2029	12
Show MSL from:	start date	1811	1	thru end date	2029	12

Data source: ◉ All ○ NOAA ○ PSMSL

Draw projection: ◉ none ○ linear ○ const accel

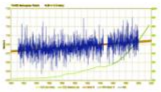
From: start date 2026 3 MSL (meters) 6.987 Slope (mm/yr) 0.953

To: end date 2099 12 MSL (meters) 7.057

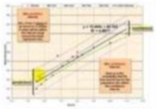
Height: 450 pixels (default 450).

Width: 806 pixels (default 806 is determined by window width).

Or, to change graph width, resize browser window and click: Update



Click here for a downloadable, bookmarkable image (except in Internet Explorer or Microsoft Edge, use PrtScn, or see NOAA's version, or PSMSL's version.)



For a simple explanation of the meaning of the 95% confidence & prediction interval traces shown in the graph, see this two-minute instructional video, by Dr. Gerard Verschuuren.

Regressions

Linear:

$y = B + M \cdot x$

$= 6882.783 + 0.953 \cdot x$ mm

Quadratic:

$y = B' + M \cdot x + A \cdot x^2$

$= 6860.451 + 0.953 \cdot x + 0.00595 \cdot x^2$ mm

where:

Date range = 1811/1 to 2021/5

x = (*date* - 1916.71) (i.e., 1916/9)

slope = M = 0.953 ±0.103 mm/yr

acceleration = 2·A = 2×0.00595 = 0.01189 ±0.00376 mm/yr²

Pasteable

linear trend = 0.953 ±0.103 mm/yr

acceleration = 0.01189 ±0.00376 mm/yr²

Source: 0 months from NOAA + 2188 months from PSMSL + 257 months from IMGW.

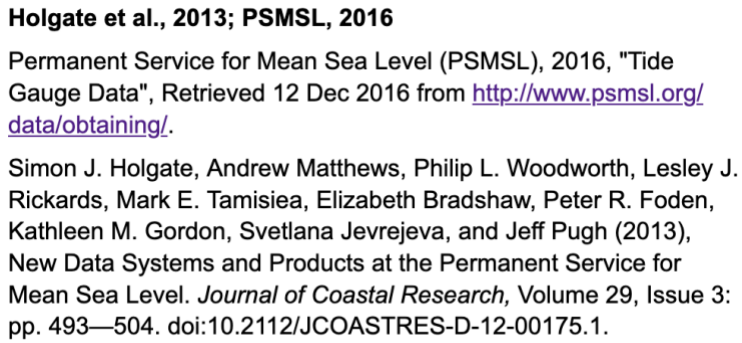
The analytical affordances of the tool are numerous. Users can run linear and quadratic regressions, compare the two fits, and examine whether detected acceleration exceeds the width of the confidence interval. They can overlay CO_2 and methane measurements on sea level graphs to visually assess whether atmospheric greenhouse gas concentrations correlate with local tide gauge trends or not. Users can also build their own weighted sea level indices from whichever stations they consider most reliable. As he states directly on the site: "The data shows what the data shows, regardless of my opinions…or anyone else's. I hope these tools are useful to everyone doing sea-level work, regardless of their views on climate issues" (Burton, 2026d).

Most of the users' analyses within the interactive tool occur through JavaScript running in the user's browser. No data are downloaded or modified server-side during a user session. Users interact with the data by selecting station locations, choosing date ranges for regression calculations, toggling between data sources, and adjusting display parameters. The analysis is therefore distributed: the tool provides the data and statistical machinery, while the specific analysis is configured by the user. Although as mentioned, aspects of the dataset can be downloaded in various formats.

4.2.6 Attributing and Citing Data Sources

Burton is consistent in attribution across sealevel.info, though not uniform in citation format. On the tool's homepage, he solely links to NOAA Center for Operational Oceanographic Products and Services with the following link: https://tidesandcurrents.noaa.gov/. However, for the PSMSL dataset, he provides two full citations as shown below in Image 4.3 (Burton, 2026d). This difference is probably because PSMSL has clear instructions on citation practices on their site (PSMSL, n.d.-b).

Image 4.3: Screenshot of sealevel.info's citation of the Permanent Service for Mean Sea Level (PSMSL) dataset, illustrating Burton's attribution practices.



On datasets he has archived on sealevel.info, Dave shows a subset of attribution metadata at the top of associated text files that include information like dataset title, creators of the data, and more as shown below in Image 4.4.

Image 4.4: Screenshot of attribution metadata displayed at the top of sealevel.info's archived text file for historical CH_4 mixing ratios from Law Dome (Antarctica) and Summit (Greenland) ice cores.

```
Historical CH4 mixing ratios from Law Dome (Antarctica) and Summit (Greenland) ice cores.
Source:  D.M. Etheridge
              L.P. Steele
              R.J. Francey
              R.L. Langenfelds
Historical CH4 record from DE08, DE08-2 and DSS ice cores (Antarctica).
Smoothed line through the ice core data (75 year smoothing spline).  From Figure 2.
```

For each individual tide station graph, he cites sources used for that specific tide gauge, which is usually a combination of PSMSL and NOAA. The Świnoujście station is an exception: because its most recent data comes from IMGW rather than PSMSL or NOAA, Burton cites IMGW directly in the station's data citations, as shown in Image 4.5, one of the few instances on the site where an institutional source outside the two primary sea level data repositories receives explicit attribution.

Image 4.5: Screenshot of sealevel.info's data citations for the Świnoujście, Poland tide gauge station, illustrating Burton's attribution of IMGW as the data source.

References ↑

I am grateful to the following data providers:

1. IMGW data (all data since 1999) is from the **Polish Institute of Meteorology and Water Management**, imgw.pl/en (see also mhwm.pl).

2. PSMSL data is from the **Permanent Service for Mean Sea Level**, retrieved 26 Jan 2026 from https://www.psmsl.org/data/obtaining/. For documentation see: Simon J. Holgate, Andrew Matthews, Philip L. Woodworth, Lesley J. Rickards, Mark E. Tamisiea, Elizabeth Bradshaw, Peter R. Foden, Kathleen M. Gordon, Svetlana Jevrejeva, and Jeff Pugh (2013), "New Data Systems and Products at the Permanent Service for Mean Sea Level." *Journal of Coastal Research* **29**(3), pp. 493-504. doi:10.2112/JCOASTRES-D-12-00175.1.

3. NOAA data is from the **NOAA Center for Operational Oceanographic Products and Services**, https://tidesandcurrents.noaa.gov/sltrends/sltrends.html (for seasonally adjusted data) and/or https://tidesandcurrents.noaa.gov/api/ (for unadjusted data).

4.2.7 Publishing and sharing data

Data are made available to users in multiple ways: JSON files with extensive metadata, and links to NOAA Seasonally-Adjusted CSV, and Unadjusted/Raw data files. Additionally, images of the resulting graphs can be downloaded as PNG files. Updates have been documented publicly on the site's changelog, which records both technical changes and Dave's commentary on related news and policy events. Data are updated continuously via automated scripts for NOAA and PSMSL holdings, with manual updates for edge cases like the Polish data.

4.2.7.1 Ongoing Maintenance of Tool and Datasets

The tool requires ongoing maintenance. Burton monitors for changes in upstream data structures that can break his automated scripts. In one instance, according to Burton, a Unicode character rather than an ASCII character introduced into a NOAA station list file caused his parsing script to fail silently, and the NOAA data were not updated for a period before he noticed and fixed it (Burton, Interview, 2025). He also manually investigates apparent data shifts caused by events such as earthquakes, to assess whether discontinuities in the record reflect genuine sea level changes or instrument and benchmark issues. As of February 2026, NOAA link changes had introduced 404 errors into some sealevel.info spreadsheets, illustrating the challenge of maintaining a tool that depends on stable upstream infrastructure. For the underlying data records themselves, the actual data points stored in his JSON files rather than linked from upstream sources, he maintains local copies.

4.2.8 Analysis:

Burton's journey is a roughly fifteen-year commitment that began with a single missing statistic and ended with him building and maintaining a parallel database and analysis tool. The journey documents an infrastructural data reuse: the construction of a database and visualization tool. Rather than reusing data to answer a single research question, Burton built a publicly accessible tool that aggregates, processes, and presents open climate data in ways designed to support flexible analyses of sea level trends. This was a relatively common mode of data reuse among this data community, contributing to what I argue is an alternative knowledge infrastructure, a point developed further in Chapters 6 and 7.

The journey moved through three phases: an initial data encounter (no average on the NOAA page); an investigative phase structured around Hansen's 3 mm/year claim and mediated by community engagement (the email list identified geographic clustering as a methodological problem); and an infrastructuring phase prompted by political involvement with NC-20. The political stakes converted an analytical curiosity into a sustained commitment to building a publicly accessible tool. The decision to build rather than simply publish an analysis was not predetermined; it emerged from Burton's accumulated work and was made possible by his technical capabilities as an IT professional with graduate training in computer science.

The social infrastructure of the skeptical community played an important role in this journey. Burton's initial Excel calculation was an unweighted average of unevenly distributed

gauges. The email list Burton consulted served as a knowledge network, surfacing a methodological issue that caused an error in the analysis and prompted its correction. This interaction illustrates how the ACC skeptic data community organizes around shared data problems and generates methodological responses, and how community membership shapes individual data practices in ways that cannot be understood by examining individual behavior alone. This kind of interaction is not unique to Burton's journey; across my ethnography, comment threads frequently functioned as informal peer review where corrections were made and resources shared, sometimes across hundreds of comments.

Burton's commitment to transparency operates as both an epistemic value and a design principle throughout sealevel.info. The site provides extensive metadata, confidence intervals, and direct links to underlying sources. Burton stated this directly: "Measurements that are transparently obtained and transparently processed are better than measurements that aren't" (Burton, Interview, 2025). He described the tool as "ideologically neutral" but still chose to disclose his own position, describing himself as a "lukewarmer" who views anthropogenic warming as "real, modest and benign" (Burton, 2017).

The tool, like most tools, is not without weaknesses and complications. Burton acknowledged that the tool can produce regressions and graphs for over 1,200 tide gauges, many of which he has never personally verified, and that PSMSL and NOAA have identified some data as being of questionable accuracy, a limitation the tool did not surface clearly to users during my observation, although it was on Dave's to-do list (Burton, 2017 ; Burton, Interview, 2025). He recommended that users consult NOAA and PSMSL pages directly and look for warnings about questionable data, datum shifts, and earthquakes (Burton, 2017). The same flexibility that supports legitimate analysis — fifteen station subsets, multiple download formats, user-configurable weighted averages — also lets a user select the stations and date ranges that produce a preferred trend. The tool is transparent about which records underlie any given graph, but it does not constrain which stations or date ranges a user selects, so whether a given graph fairly reflects the broader record depends on the user's choices rather than the tool.

Sealevel.info is built on top of NOAA and PSMSL data, and its outputs were often used to contest claims made by mainstream climate science in my digital ethnography. But the relationship between Burton's tool and the data repositories it draws on is not simply one of dependency or opposition. For instance, the Polish tide gauge data complicates that framing.

Burton obtained Polish tide gauge data not held by global repositories, incorporated them into his tool, and forwarded them to PSMSL and NOAA, who did not use them. According to PSMSL's data documentation, that gap is "due to data policies in Poland" (PSMSL, n.d.-a), a governance constraint, not a quality judgment on Burton's data. Burton's willingness to contact IMGW directly and incorporate the records reflects an infrastructural flexibility that formal repositories, bound by national data-sharing agreements, do not always have. In this instance, the skeptic tool held a more complete record of a specific station than either of the two leading global data repositories. The relationship between sealevel.info and the repositories it draws on is therefore symbiotic: dependent on their data, sometimes extending their coverage, and oriented toward contesting the conclusions drawn from it.

That symbiosis runs deep in this data reuse journey. Sealevel.info is reliant on the continued access and collection of sea level tide gauge data from PSMSL, NOAA, and even IMGW. The tool built in response to gaps in interoperability and usability is vulnerable to the fragility of the infrastructure it was designed to supplement. When NOAA changes its link structures or a Unicode character breaks a parsing script, Burton's tool breaks too.

4.3 Rapid Reuse: Contesting Media Narratives (Burnett: *Climate Realism*)

4.3.1 Data Reuse Prompt and Design

H. Sterling Burnett is the director of The Heartland Institute's Arthur B. Robinson Center on Climate and Environmental Policy, editor of Heartland's *Climate Change Weekly*, host of *The Environment & Climate News Podcast*, and a panelist on *The Climate Realism Show*. Prior to joining the Heartland Institute in 2014, Burnett worked as senior fellow for environmental policy at the National Center for Policy Analysis ("H. Sterling Burnett," n.d.). He has a B.B.A. and a B.A. in cultural anthropology from Southern Methodist University and an M.A. and Ph.D. in applied philosophy with a specialization in environmental ethics ("H. Sterling Burnett," n.d.). The Heartland Institute is a conservative free-market think-tank ("About Us," n.d.-a). *Climate Realism* is a daily publication affiliated with Heartland that responds to media coverage of climate change. As Burnett described it, the project is "not, except indirectly, addressing the science of climate change. What it is, is addressing the media's presentation of the science of climate change" (Burnett, Interview, 2025). Each day, Burnett searches Google News for climate-related stories and selects one or more to respond to, typically by checking whether

claims in the media article are supported by what he calls “measured data:” observed records rather than model projections (Burnett, Interview, 2025). This data reuse journey begins, as most *Climate Realism* articles do, not with the data but with media coverage.

The account that follows describes one instance of Burnett's *Climate Realism* workflow as I observed it during this study (September 2021–December 2025), reconstructed from his blog post and a semi-structured interview. Both *Climate Realism* and the external data sources it draws on are subject to change, and the EPA source central to this journey was, as described below, removed in December 2025.

H. Sterling Burnett encounters a news article in *The Hill*, written by Sharon Udasin, titled “Climate change is hampering US apple quality and output: Study.” (Udasin, 2025). *The Hill* news article cites a single journal article (Preston et al., 2024), which examines changing climate risks for high-value tree fruit production across the United States. Burnett frames *The Hill* article as an instance of journalistic overreach, characterized by three features (Burnett, 2025):

- Reliance on a single peer-reviewed study,
- Insufficient engagement with “real-world data,” and
- Participation in a politicized science-to-media-pipeline

Burnett’s use of “real world data” in the blog post reflects a specific epistemic commitment that runs throughout his work and the broader ACC skeptic community (Burnett, 2025). For Burnett, models “are tools, nothing more. They don’t produce data. They produce forecasts and outcomes, but they’re projections, and that’s not data” (Burnett, Interview, 2025). “Real world data” means measured or observed records, production statistics, temperature readings, tide gauges, as distinct from model outputs or projections. This distinction shapes which evidence Burnett treats as relevant and which he sets aside. For Burnett, what apple orchards “produce and trends over years and decades” are more probative than what climate risk models suggest growing conditions might become (Burnett, Addition, 2026).

Rather than engaging specifically with the journal article, Burnett focuses on the news article as the primary object of correction. He views the more consequential problem as how scientific findings move through journalism into public discourse. His analysis targets the media claim and reframes the issue as a question of empirical verification: do measured data support claims that climate change is harming apple production? He pursues this through two subquestions: are heat waves increasing? Has apple production declined over time?

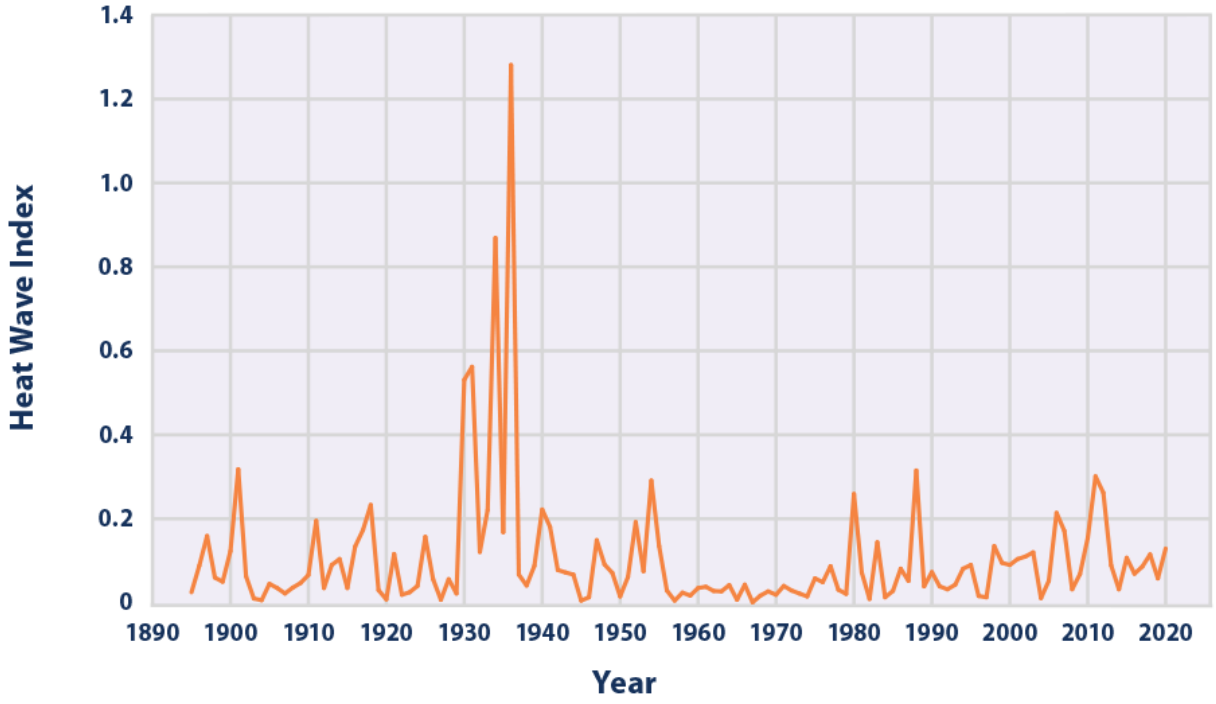
4.3.2 Seeking and Selecting Data

Burnett approaches his research questions with considerable experience writing similar articles responding to media claims that climate change was harming food production. He has well-established "go-to" information and data sources: the EPA's U.S. Heat Wave Index Graph in the image below, covering the contiguous 48 states from 1895 to 2020, and the Food and Agriculture Organization of the United Nations (FAO) (Burnett, Interview, 2025).

Within these types of blog posts, he prefers reusing graphs if he trusts the underlying data rather than repeating the work himself. The EPA's U.S Heat Wave Index Graph (shown below in Image 4.6) is reused across Heartland Institute's *Climate Realism* and *Climate at a Glance* articles because it shows the heat waves that occurred in the 1930s. Burnett uses this visualization to argue that heat waves are less frequent now than in the 1930s. The EPA's Climate Change Indicators page, including this graph, was removed in December 2025 and the original URL is no longer accessible (*Page Not Found | US EPA*, n.d.).

Burnett's trust in a data source rests on the institutional history of the producing agency, vetting by colleagues, and past reliability in his own experience. On the National Weather Service, for instance, he noted its longevity as evidence of credibility, it predates the government agencies he associates with more recent political pressures on data (Burnett, Interview, 2025). He also uses data sources that his colleagues at the Heartland Institute have already evaluated; their endorsement carries weight. And sources he has used before without encountering problems accumulate trust over time: "because I've found those sources to be trustworthy in the past, I feel okay trusting them on that" (Burnett, Interview, 2025). Where he has reservations about a source's limitations, as he does about FAO's reliance on national self-reporting, he uses it anyway on pragmatic grounds: "we can only work with the data that exists" (Burnett, Interview, 2025).

Image 4.6: The U.S. Heat Wave Index graph as embedded in Burnett's *Climate Realism* blog post, sourced from the EPA's Climate Change Indicators: Heat Waves page.



He treats the UN's FAOSTAT as an authoritative global source for agricultural data. Before navigating to FAO, he first performs Google searches (e.g., "apple production in the United States") to see whether pre-made visualizations already exist. If others have produced usable charts using data he already trusts, he prefers not to repeat the work himself. As he explained, the goal is not data analysis for its own sake but rather answering the question as efficiently as possible: "Usually they're using FAO data… but sometimes they already have nice graphics done and I don't have to do that" (Burnett Interview).

In this case, he does not find a suitable pre-made visualization and therefore navigates to FAO directly.

4.3.2.1 Data Infrastructure Context

The FAOSTAT data Burnett relies on originates from member countries of the United Nations. The collection methods may vary and come from things like Annual Production Questionnaires (APQs), National yearbooks, Administrative data or expert estimates (*FAOSTAT*, n.d.). Countries validate their own data. FAO conducts transmission error checks, and consistency and outlier detection. Corrections occur only in consultation with national correspondents. Missing data are imputed using ensemble models (crops, meat) or hierarchical linear models (milk, eggs). Imputed data are flagged in the flag metadata field with an IM. The

accuracy of the imputation is unknown, though the process is regularly reviewed. For the US apple production data Burnett analyzes, no flags appear, indicating that the values are reported rather than imputed.

There are stated limitations to these data. First, geographical comparability is limited due to national methodological differences (except among EU countries). Temporal comparability is reasonable in the short-term but weak over a long time series. Temporal comparability weakens over longer time series for two reasons: the series extend back to 1961, spanning multiple methodological transitions, and several major countries have revised historical production data as collection methods improved. Overall, the accuracy cannot be fully assessed due to reliance on national self-reporting. These limitations do not surface in Burnett's blog post itself, though Burnett acknowledged some during the interview:

> So could there be problems with the FAO data? Of course, they depend to a large degree on self-reporting from the nations that they're reporting on. My suspicion is that different of those nations have varying levels of capabilities and competencies partly tied to the funding, devoted to what they're doing. Some nations, as we know, are poorer than others, and so maybe their data quality [is lower], but we can only work with the data that exists. And so, we take the FAOs data as being, for our use, is fairly authoritative when we're working on agricultural stories (Burnett, Interview, 2025).

4.3.2.2 Note on Differences in Data Scope:

The journal article (Preston et al., 2024) and Burnett's blog post rely on substantively different data and analytical approaches. Preston et al. (2024) analyze climate risk to apple production using gridMET (Abatzoglou, 2013), a high-resolution climate dataset, county-level spatial analysis, and six phenological climate metrics (chill portions, cold degree days, frost risks, growing degree days, extreme heat days, and warm nights). Critically, the study explicitly does not analyze apple production or yield data, stating "we do not quantify impacts on yields or quality because of unavailability of suitable data" (Preston et al., 2024, p. 2), because "data on apple yields are limited at the county resolution in the USDA NASS census and survey. Higher spatiotemporal resolution data on apples and other tree fruit are necessary for analyzing the magnitude of climate impacts" (Preston et al., 2024, p. 9). Burnett's blog post, by contrast, analyzes production quantity, area harvested, and yield from FAO's agricultural statistics.

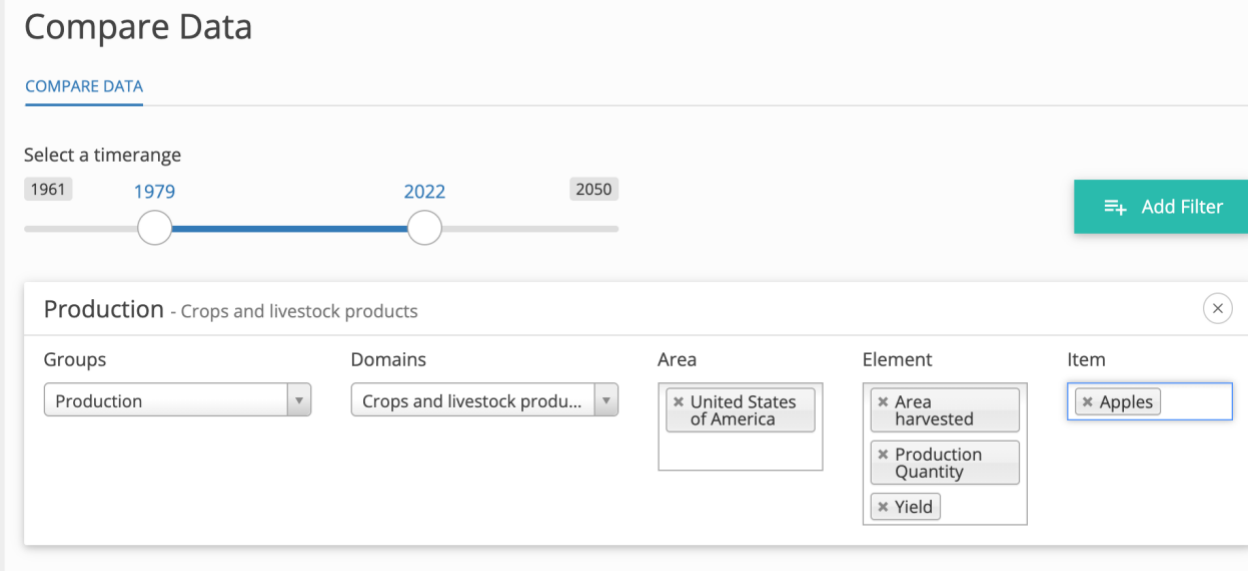
4.3.3 Processing and transforming data

Within the FAOSTAT interface, Burnett selects apple production data for the United States. In most *Climate Realism* articles, he prefers a 30–40-year window, typically from around 1990 to the present, which he views as a meaningful period for evaluating climate claims. He said:

> It's a 30-year trend line. One year means nothing. So we started. Sometimes we go back to '88, when Hansen first claimed before Congress that climate was really bad, really bad. Humans were causing dangerous climate change. That was '88. Usually now we'd go back to '90. So that's 34 years now, right? I think that's a fair period of comparison. So you go to the select time range, and you go to, let's say 1990 (Burnett, Interview, 2025).

However, in this case, he selects the 1979–2022 time range in the United States in order to match the period and place analyzed in the journal article (Preston et al., 2024) cited by *The Hill* (Burnett, Interview, 2025). The variables selected include area harvested (ha), production quantity (tonnes), and yield (hg/a). All processing occurs within the FAOSTAT interactive interface. No data are downloaded at this stage.

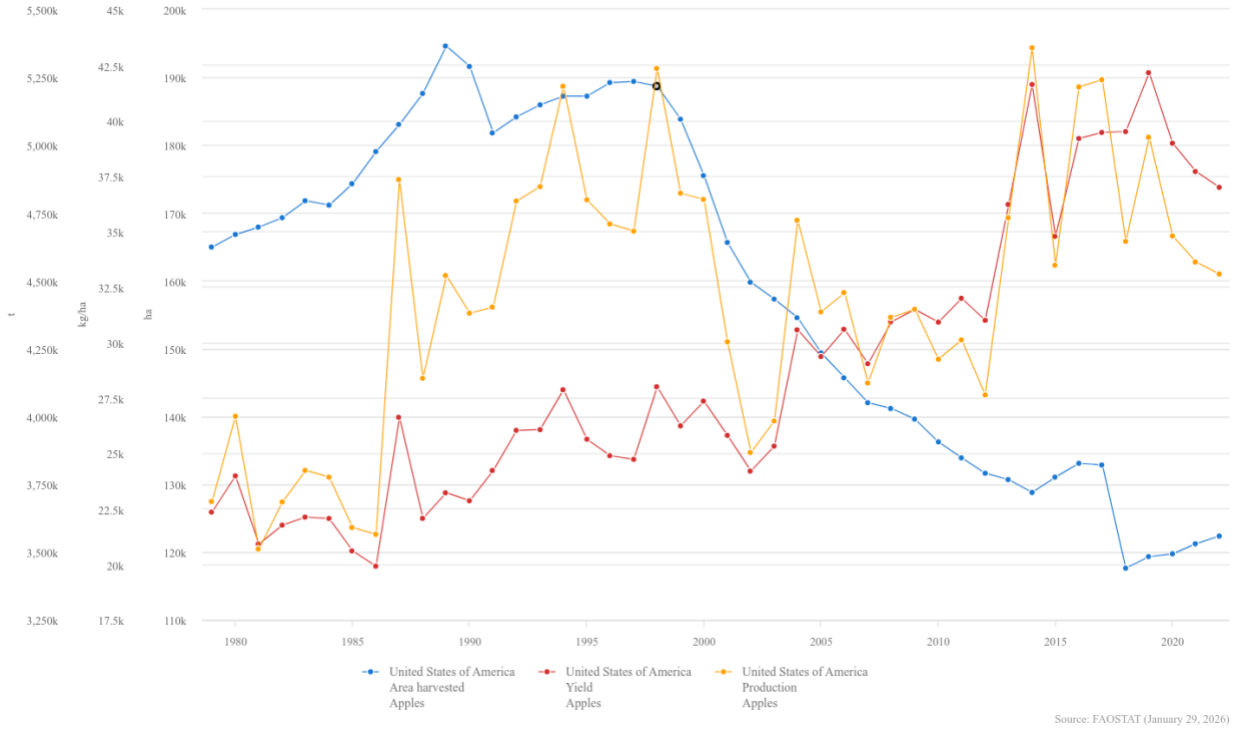
Image 4.7: Screenshot of the FAOSTAT Compare Data interface showing Burnett's variable selections for his apple production analysis: United States of America, 1979–2022, with area harvested, production quantity, and yield selected as elements, and apples as the item.



4.3.4 Analyzing and Interpreting Data

Burnett presses "compare data," and a visualization is generated (Image 4.8). Burnett then performs manual calculations using values displayed in the interactive graph. He uses the cursor-hover feature to read values for selected years and calculates percentage change by hand:

Image 4.8: Screenshot of Burnett's FAOSTAT graphs posted on blog post.



So if you go back now to the FAO on apples, you can do the calculations yourself. The FAO doesn't do it for me. I have to do it. So if you click on 2023, the last date on apples, you go to that chart and you wave your little pointer over it, you see apples. It tells you exactly how many were produced and what the yield was. For both the world… and then I go back to 1990 and I look at what they were, and I do the math. To produce a percentage. It could be a decline. I can't think of a situation where it was a decline. I think one we had from Colombia recently was a decline, but the state surrounding it weren't declines, and so it had to be explained by some other factors. But in any case, it's usually not a decline. So I just calculate the percentage increase. You take the large numbers, subtract the initial number, divide the new number by the initial number, multiply by 100 to and it tells you what the percentage growth is. So that's how we do it. I mean, like I said, the FAO doesn't do that. So if anyone said, well, where are you

> getting that? To be fair, when we copy the graphic, when you go up to the little bars where it says chart and you copy the PNG or the JPEG over, it shows the absolute number, not the percentage change. The percentage change doesn't show up in your Word file. So you actually have to compute that yourself in theirs, using their data (Burnett, Interview, 2025).

Through this process, Burnett determines that between 1979 and 2022: acreage devoted to apple production declined by approximately 26 percent; total apple production increased by about 23 percent; and yield increased by approximately 66 percent. He interprets these trends as indicating increased productivity over time and presents declining acreage alongside rising yield and production as evidence of increased efficiency.

4.3.5 Publishing, and sharing data

Burnett downloads the FAOSTAT visualization as a PNG or JPEG image and pastes it directly into the *Climate Realism* blog post. Although FAOSTAT allows users to download the underlying data and metadata as CSV files, these are not reused. Only the static image travels forward into publication. Once embedded in the blog post, the image is low resolution and difficult to read. Metadata, uncertainty indicators, and methodological context do not accompany the visualization.

4.3.6 Attributing and citing data

Burnett links to the FAOSTAT Compare tool (*FAOSTAT*, n.d.). The EPA Heat Waves Index Graph is also embedded in the blog post at a higher resolution than the FAOSTAT graph with the following caption:

> This figure shows the annual values of the U.S. Heat Wave Index, covering the contiguous 48 states from 1895 to 2020. Source: Graph from Environmental Protection Agency, "Climate Change Indicators: Heat Waves;" https://www.epa.gov/climate-indicators/climate-change-indicators-heat-waves#%20 (Burnett, 2025).

He also links to both the *Climate at a Glance* article on U.S. heatwaves that includes the graph with more analysis (*U.S. Heatwaves*, n.d.). That link is now unavailable: the EPA's Climate Change Indicators page was actively removed in December 2025 (Comms, 2025).

4.3.7 Publishing and sharing research

The visualizations are presented alongside the blog's textual interpretation, which uses long-term heatwaves, apple production, and yield trends to argue against claims that climate change is harming apple production in the United States. Burnett also analyzes the nature of science to media pipeline, critiquing the media article stating that:

> *The Hill*'s readers would be better served if, going forward in this age of highly politicized science, its writers sought out real world data on any crop being discussed, before referencing a single new report as proof of harm to a crop and quoting only its authors as if they were the sole arbiters of truth on the topic (Burnett, 2025).

The blog post is published on *Climate Realism* and disseminated through the website and associated social media channels.

4.3.8 Managing, documenting, and preserving data:

Burnett does not download or archive the underlying FAOSTAT data in this journey. The data exist within the FAOSTAT interface for the duration of the analysis and are not preserved beyond the downloaded visualization. No additional documentation or metadata are retained.

4.3.9 Analysis:

This data reuse journey documents a media-prompted instance of data reuse, using long-term measured datasets to evaluate a claim made in a news article. *Climate Realism* is affiliated with the Heartland Institute, Burnett's employer, whose organizational focus ties responses to media coverage to its broader goals around climate policy and markets. This is a framing the Institute's president, James Taylor, articulated directly in his own interview. Within this context, Burnett's rapid response to media coverage is not coincidental. It is the publication's designed mode of operation. During the interview, Burnett shifted from first person singular to first person plural when describing the analytical method: "so that's how we do it" (Burnett, Interview, 2025). This shift perhaps reflects that what looks like an individual data reuse journey is embedded in an organizational content production workflow, the method he described is a shared practice across the *Climate Realism* team, not only his personal approach.

Burnett directs his response at the media article rather than the journal article it references. This is a deliberate choice grounded in his stated purpose: *Climate Realism* addresses the media's presentation of science, not science itself. Burnett frames *The Hill*'s coverage as mischaracterization rather than simplification. Whether *The Hill* mischaracterized or simplified

Preston et al. (2024) is outside the scope of this study. What matters analytically is that Burnett's reuse is structured around this framing: the media article failed to check its claims against what he calls "real world data," meaning observed production records rather than projected climate risk factors that Preston analyzes.

The FAOSTAT Compare interface functions here as a harmonized global dataset that makes broad comparative data reuse easy. The infrastructure is designed in a way that supports rapid access and use by a broad user base. While the FAOSTAT Compare interface increases accessibility and usability, there are potential drawbacks. The interface produces a visualization without the statistical machinery to assess significance, confidence intervals, or confounding factors. It enables rapid comparison, thus reducing friction, but does not prompt the user toward methodological caution. It is optimized for accessibility over depth.

Contextual information is reduced as each stage of the data reuse journey. Not all of this loss is attributable to Burnett's choices; the tool's design plays a significant role. He begins with FAOSTAT, where data are structured, queryable, and accompanied by documentation. When he enters the Compare interface, the tool itself limits what travels forward: it presents only selected variables and time ranges, without surfacing imputation flags, methodological caveats, or uncertainty information in the visualization itself. When he downloads the visualization as an image, the interface provides no mechanism for carrying metadata or provenance information into the exported file. Once Burnett embeds the image in the blog post, the visualization circulates as a stabilized artifact alongside interpretive text. Burnett could work actively against this loss, by annotating the visualization, linking to the underlying data, or noting FAO's methodological caveats, but the tool's design does not prompt or facilitate any of these steps. The contextual loss is therefore co-produced: structured by the interface and compounded by the publication format.

What specifically is omitted matters for understanding this journey. Preston et al. (2024) examined six phenological climate risk metrics: chill portions, cold degree days, frost risks, growing degree days, extreme heat days, and warm nights. Preston et al. explicitly did not analyze production yield data, noting, that in their view, it was unavailable at sufficient resolution. Burnett substitutes a different question: did apple production go up? These are not the same question. Burnett's focus on production outcomes is internally consistent with his epistemic commitment to observed records over projected or modeled risk.

The active removal of the EPA Climate Change Indicators page in December 2025 illustrates how infrastructure dependencies make data reuse vulnerable to upstream political decisions (Comms, 2025). A visualization *Climate Realism* has reused across many articles is now unavailable at its source. The graph however lives on, copied across blog posts and social media, retaining its authority as an EPA-produced graph, while losing the contextual information that once accompanied it, including the documentation that would situate the 1930s heat wave peak within a broader view of climate variability. This is not a data management failure. It is a political intervention in the knowledge infrastructure, and it has consequences for how evidence circulates long after the removal. ACC skeptics can continue to deploy it as EPA-sourced evidence that heat waves were worse in the 1930s than today, using the graph to contest claims about worsening extreme heat events without the surrounding context that would facilitate a more careful reading of the index and its meaning.

4.4 The Human Contribution to Atmospheric Carbon Dioxide: Skeptical data practices and a mainstream-aligned conclusion (Engelbeen - CO_2 Coalition)

4.4.1 Data reuse prompt and design

Ferdinand Engelbeen "has a B.Sc. in process chemistry" and spent most of his career in "process engineering, primarily in manufacturing of paints and inks and later in a large chlorine/VCM/PVC factory" (Engelbeen et al., 2024, p. 65). Engelbeen's path to this research began in the 1990s when he was working in the chlorine and PVC industry, which "was heavily attacked by Greenpeace in that period" (Engelbeen, Interview, 2025). Convinced Greenpeace was wrong but lacking all the answers, he "started a search in the scientific literature to see what was right and wrong with their stance," culminating in founding the "Chlorophiles" organization and engaging in legal proceedings against Greenpeace in Hamburg, Germany (Engelbeen, Interview, 2025).

His interest in the climate emerged from reading about the influence of solar cycles on Earth's climate "some 50 years" ago (Engelbeen, Interview, 2025). In the late 1990s, he became absorbed in debates between Peter Dietze, who argued for an approximately 50-year CO_2 atmospheric decay rate, and Fortunat Joos of Bern University, whose model predicted the much longer CO_2 residence times used by the IPCC. With his "background in chemical processes," Engelbeen attended a course in climate modeling at the University of Oxford to better understand

climate model performance (Engelbeen, Interview, 2025). During the same period, “there were several discussions between climate realists [a common term used by ACC skeptics to describe members of their community] about the origin of the CO_2 increase, similar to today” (Engelbeen, Interview, 2025). Engelbeen created a webpage that summarized these discussions. Those debates were never fully resolved. Engelbeen thus formalized his response in a grey literature report co-authored with Dave Burton — whose sealevel.info journey is documented in Section 4.2 of this chapter — and one other researcher, for the CO_2 Coalition (Engelbeen et al., 2024).

Engelbeen describes the motivation for the report as “repeated publications of claims” within the ACC skeptical community “that the current increase of CO_2 in the atmosphere was (mainly) natural, based on wrong data, wrong interpretation of real world processes and ignoring the carbon mass balance” (Engelbeen, Interview, 2025). The report presents “multiple lines of scientific evidence demonstrating conclusively that the modern increase in CO_2 is mainly due to anthropogenic emissions” (Engelbeen et al., 2024). Its scope is narrow and deliberate: it addresses the attribution question only and does not address whether rising CO_2 is harmful or beneficial. The CO_2 Coalition’s organizational position that CO_2 is beneficial appears in the report’s subtitle (“How Human Emissions Are Restoring Vital Atmospheric CO_2”) but is not analyzed or defended scientifically within the report itself. On the specific attribution question the report addresses, its conclusions align with mainstream climate science (Engelbeen et al., 2024, p. 4).

This data journey focuses on Section 2.3 of the grey literature report, which documents Engelbeen’s quality assessment of a fellow skeptic’s research and data. In Section 2.3, Engelbeen audits a specific claim made by Ernst-Georg Beck that atmospheric CO_2 reached an approximately 70 ppmv peak around 1942, a claim that, if true, would suggest natural variability could produce large CO_2 swings and undermine the anthropogenic attribution argument.

The account that follows reconstructs Engelbeen's audit as presented in the CO_2 Coalition report and elaborated in a semi-structured interview conducted during this study (September 2021–December 2025). The report was published in December 2024; the analysis described here reflects that published version and the data sources as they stood at the time of my review.

4.4.2 Seeking and selecting data

Data discovery was straightforward; if Engelbeen did not already know of a source, he relied on Google searches (Engelbeen, Interview, 2025). His data selection strategy reflects both

pragmatic constraints and epistemic commitments about quality and reliability. He explained: "As far as possible from reliable sources, in many cases there is no choice at all, because there is only one source: e.g., NOAA for ocean pH and pCO_2 data" (Engelbeen, Interview, 2025). Where competing datasets exist, such as between Scripps Institute and NOAA for atmospheric CO_2, he evaluated consistency: "The two data sets don't differ with more than 0.2 ppmv [parts per million by volume], which proves that the CO_2 measurements worldwide are very accurate" (Engelbeen, Interview, 2025). He maintained a preference for direct over proxy measurements: "Direct measurements, if available, are almost always preferable to proxy measurements" (Engelbeen, Interview, 2025). He preferred CSV files, "as that is easily imported into Excel and easily converted into other data formats" (Engelbeen, Interview, 2025).

Engelbeen compares the Modern Hessisches Landesamt für Naturschutz, Umwelt und Geologie (HLNUG) continuous monitoring data from Linden/Giessen, Germany, measured at half-hour intervals via gas chromatography (*Linden Air Gauge | Measurement Data Portal*, n.d.) to Beck's historical measurements at the same location. NOAA background station data from Barrow, Mauna Loa, and the South Pole provide a stable atmospheric baseline for assessing how much local bias exists at Linden/Giessen (Lan et al., 2023). And NOAA NCEI Law Dome ice core data serve as an independent check on Beck's claimed CO_2 peak (D. Etheridge et al., 1998). Beck's own compilation of historical CO_2 measurements (Beck, 2007, 2022)[3] is the primary object of evaluation rather than a trusted data source.

4.4.2.1 Assessing Data and Repository Trustworthiness

Engelbeen articulated criteria for evaluating data trustworthiness that apply directly to his assessment of Beck's historical measurements:

> If there were different datasets, looking at the differences and why there were differences and how the differences were explained (if at all). Further the reputation of the data provider for integrity of data and eventual corrections of the data in the past and external critiques on these corrections from third parties (Engelbeen, Interview, 2025).

[3] Beck (2022) was published in *Science of Climate Change*, a journal associated with the ACC skeptic community, as in the "start-up phase" it was "hosted and strongly supported by the Norwegian Climate Realists" https://scienceofclimatechange.org/.

He questioned data trustworthiness “when there are huge differences between similar datasets” (Engelbeen, Interview, 2025).

His preference for “direct measurements” over proxies reflects a distinction between data that directly measures the phenomenon of interest and data that reconstruct it through calibration against an intermediary (Engelbeen, Interview, 2025). He illustrated this with ice cores versus plant stomata.

> That is e.g., the case between direct CO_2 measurements from ice cores and stomata index data. The latter is a proxy from leaves from land plants, where there is always a local bias in CO_2 levels. The stomata data are calibrated against ice core CO_2 over the past century, but nobody knows how the local bias evolved over the centuries in the past, due to changes in local vegetation or vegetation in the main wind direction (Engelbeen, Interview, 2025).

Ice core gas bubbles, which enclose ancient atmospheric air, are treated as more reliable than stomata or tree rings. The paper itself reflects this hierarchy, describing ice core CO_2 data as “direct measurements of ancient air” and treating them as the gold standard for paleo-atmospheric CO_2 reconstruction (Engelbeen et al., 2024, p. 37). This is the reason why Law Dome ice cores serve as the primary independent check on Beck’s claimed 1942 peak.

Beck’s data do not meet Engelbeen’s trustworthiness criteria. The specific failures are: lack of description of methods, sampling, handling, and calibration; measurements taken at locations subject to strong local CO_2 sources and sinks; and implausible variability, with a range of 240–680 ppmv at a single location. By contrast, HLNUG modern data and NOAA CO_2 stations meet his criteria as continuous, calibrated monitoring stations. Law Dome ice cores meet his criteria as direct measurements with a known resolution of 10–14 years.

4.4.3 Processing and transforming data

Engelbeen downloaded continuous half-hour interval data from the HLNUG monitoring station at Linden/Giessen and NOAA background station data from Barrow, Mauna Loa, and the South Pole for the same time periods. No custom programming is described for this section. Processing was primarily manual comparison of modern and historical measurement ranges.

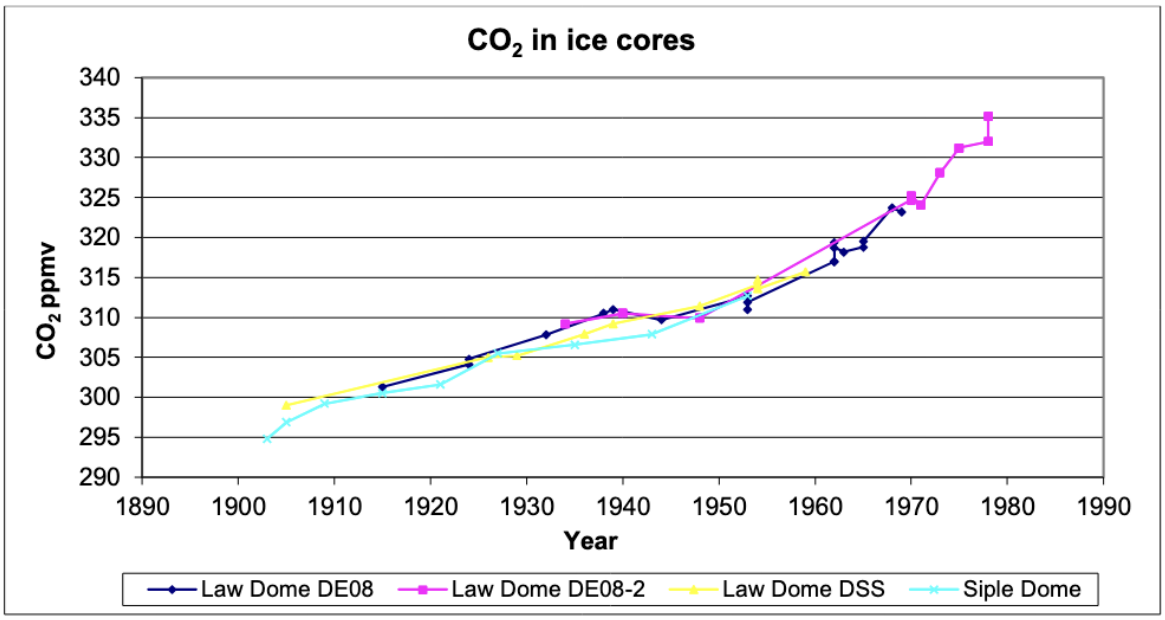
4.4.4 Analyzing, Interpreting, and Abstracting Data

Engelbeen's analysis evaluates Beck's claim of a 70 ppmv CO_2 spike around 1942 through four comparisons (Beck, 2007, 2022). As a reminder if a 70 ppmv atmospheric CO_2 peak occurred around 1942, it would demonstrate that natural processes alone can produce large CO_2 swings, directly undermining the anthropogenic attribution argument. Notably, the HLNUG Linden/Giessen data Engelbeen uses to evaluate Beck's claim come from the same semi-rural place Beck used to validate his background estimation methodology. Where Beck used the historical station data to support his approach, Engelbeen uses the modern data to demonstrate that the Linden/Giessen location is subject to severe local contamination. "Besides more traffic in modern times, Giessen did remain semi-rural with surrounding forests, which have a huge day/night contrast on CO2 levels, especially visible under inversion conditions" (Engelbeen, Addition, 2026). This directly undermines the historical Giessen measurements that form the basis of Beck's claimed 1942 CO_2 peak.

The first comparison pits Beck's claimed peak against the ice core record. Law Dome ice cores show only a 9 ppmv increase from 1930–1960 with no peak in 1942, directly contradicting Beck's claimed 70 ppmv spike. Engelbeen concludes:

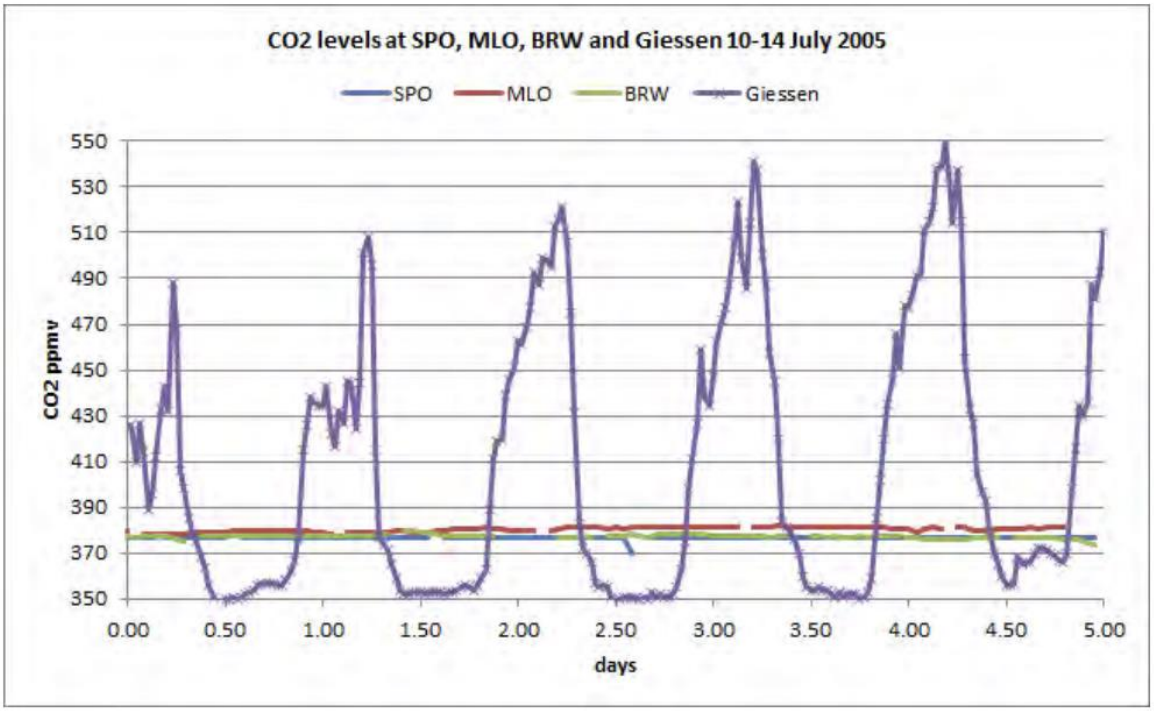
> Any real peak of some 70 ppmv over a period of nearly a decade would have been detected by the high-resolution (10–14 year) Law Dome ice cores, but it is not visible in the ice core record. That means that Beck's "peak" is questionable and is probably an artifact of local biases in the historical CO_2 measurements (Engelbeen et al., 2024, pp. 50–51).

Image 4.9: Figure 2.3.2 from Engelbeen et al. (2024, p. 51) showing "Ice core CO_2 in four ice cores, Law Dome DE08 and DE08-2 have a resolution of eight years and Law Dome DSS and Siple Dome a resolution of around 20 years. The Law Dome ice cores have an overlap with the South Pole CO2 data (1958–1978) within a few ppmv of each other (NOAA ongoing)."



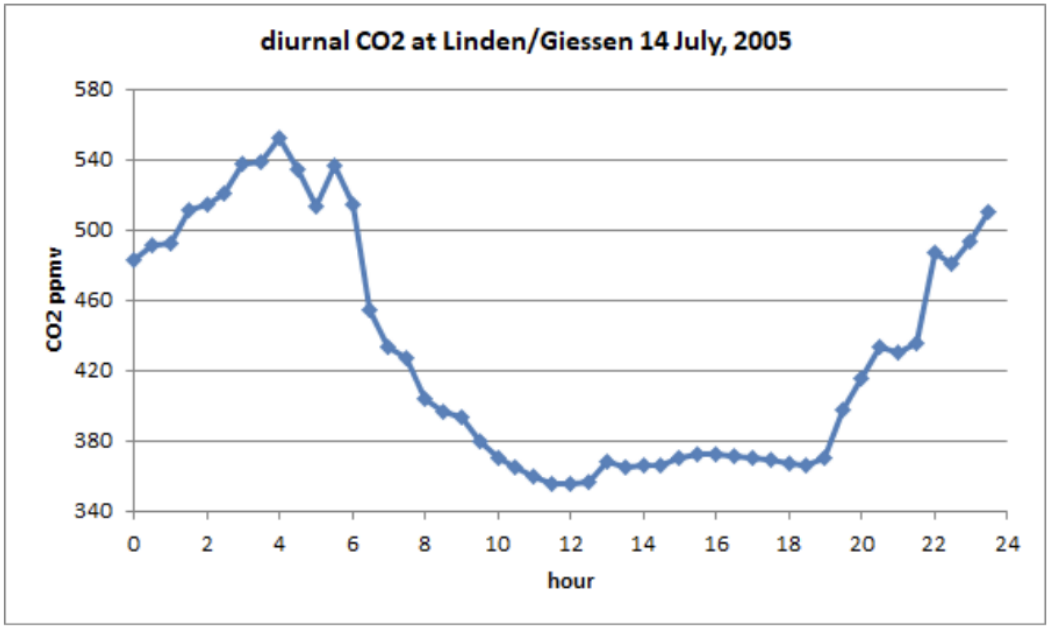
The second comparison examines HLNUG modern monitoring data at Linden/Giessen against NOAA background station data from Barrow, Mauna Loa, and the South Pole. Engelbeen shows that CO_2 at Linden/Giessen fluctuates between 350 and over 500 ppm within a single 24-hour period under inversion conditions, while the NOAA background stations remain stable across the same period, as shown in Image 4.10 below. This demonstrates that the Linden/Giessen site is subject to severe local contamination that precludes its use as a reliable background measurement location.

Image 4.10: Figure 2.3.3 from Engelbeen et al. (Engelbeen et al., 2024, p. 52) showing "CO_2 levels as measured at Linden/Giessen under inversion (HLNUG, ongoing) compared to hourly averaged raw CO_2 data from Barrow, Mauna Loa and the South Pole for the same days, including all local outliers. Data from (NOAA ongoing)."



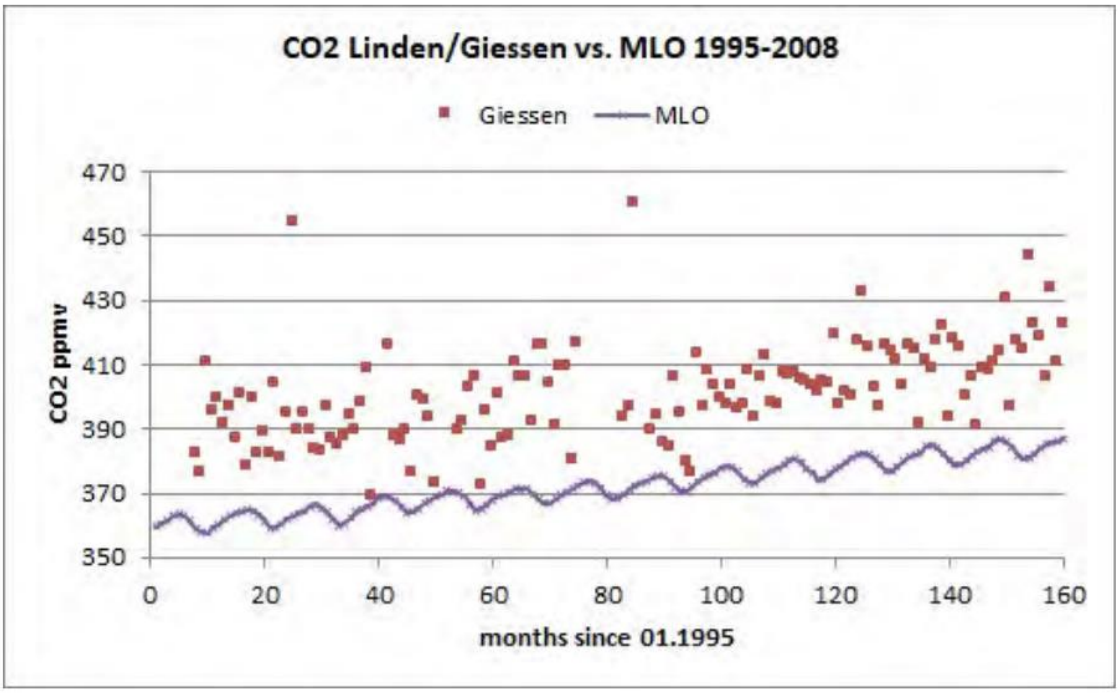
The third comparison narrows the focus to a single 24-hour period, examining HLNUG diurnal CO_2 variation at Linden/Giessen under inversion conditions. Engelbeen demonstrates that "CO_2 concentrations at ground level can vary dramatically over the course of a day" as shown in the Image 4.11 below (Engelbeen et al., 2024, p. 52), with implications for the reliability of Beck's measurements, which were taken only three times daily at fixed hours.

Image 4.11: Figure 2.3.4 from Engelbeen et al. (Engelbeen et al., 2024, p. 53) "Diurnal CO2 change at Linden/Giessen under inversion (HLNUG ongoing)."



The fourth comparison draws on monthly averaged HLNUG data at Linden/Giessen compared against Mauna Loa (NOAA) for 1995–2008. Engelbeen shows in Image 4.12 below that Linden/Giessen maintains a persistent local bias of 30–40 ppmv above Mauna Loa background levels even when averaged across full months, demonstrating that the contamination is not only an inversion-condition phenomenon but a structural feature of the site.

Image 4.12: Figure 2.3.5 from Engelbeen et al. (2024, p. 53) showing “Monthly averages at Linden/Giessen compared to Mauna Loa data.”



Taken together, the four comparisons support one conclusion: "A huge peak of +70 ppmv CO_2 (150 PgC) in the atmosphere around 1942 is physically impossible, and the claim is based on measurements at places contaminated by huge local CO_2 sources" (Engelbeen et al., 2024, p. 58).

4.4.5 Managing, documenting, and preserving data

Engelbeen explains: "The data and graphs are stored on our computers and backed with backup on hard drives and memory sticks. The most important data and graphs are also stored in clouds and Internet sites" (Engelbeen, Interview, 2025).

For visualizations, Engelbeen created most graphs in Excel, converting them to JPG format for older graphs (2000-2012) or PNG format for newer graphs using Paint: "Except for 'borrowed' graphs from different sources, most graphs were made in Excel and via 'Paint' transformed into .jpg files for the older graphs (2000-2012), newer graphs are in .png format" (Engelbeen, Interview, 2025). He notes that some graphs were not updated between the earlier discussions around 2012 and the current project "due to lack of time or lack of new data" (Engelbeen, Interview, 2025).

4.4.6 Publishing and sharing research data

No data are deposited in a formal repository. As Engelbeen notes: "Everybody is free to ask for use of all our data and graphs, but until now we haven't had any formal requests" (Engelbeen, Interview, 2025). The data exist on personal storage and cloud backups, accessible through direct contact with the authors.

4.4.7 Attributing and citing data

Engelbeen and his co-authors attribute data sources through a combination of formal citations in the reference list and informal inline citations within figure captions and body text. Formal citations follow a standard academic format. For example, the authors cite a data description paper for the Global Carbon Project 2023 data with a full bibliographic entry and DOI in the reference list (Friedlingstein et al., 2023). However, within figure captions and body text, the same dataset, but perhaps not the same version, is often cited informally as "(Global Carbon Project, ongoing)" or "(GCP ongoing)," a shorthand that foregrounds the dataset's continuously updated nature rather than a fixed publication. This kind of dual citation practice reflects a tension between the conventions of academic publishing, which assume stable, versioned sources, and the reality of many datasets, which are scholarly resources updated on a regular basis.

Several data sources are cited by URL only, without formal bibliographic information, particularly for institutional monitoring programs such as NOAA's continuous CO_2 measurements and the HLNUG Linden/Giessen station. The "(ongoing)" designation appears throughout, applied to sources like NOAA CO_2, Wood for Trees, and HadSST3, signaling that the data extend beyond any fixed publication date.

Data sources are also cited within figure captions rather than in the body text, meaning attribution is sometimes tied to a specific visualization rather than to the claim the visualization supports. This places the citation burden on the reader to connect figure captions to the reference list, and in some cases the caption attribution is the only formal acknowledgment of a data source in the document.

4.4.8 Publishing and sharing research

The report was published in December 2024 by the CO_2 Coalition as a PDF report available on their website. The CO_2 Coalition, a 501(c)(3) is a nonprofit advocacy organization

in the United States that promotes narratives "about the important contribution made by carbon dioxide to our lives and the economy" ("About Us," n.d.-b). The Coalition's organizational position shapes how the report frames its findings. The "restoring vital atmospheric CO2" framing in the title reflects the Coalition's position that higher CO2 is beneficial (Engelbeen et al., 2024, p. 1). The publication venue ensures the work circulates with audiences in the ACC skeptical community.

Unlike peer-reviewed journal articles, this report represents grey literature publication through an advocacy organization. Grey literature refers to research produced and distributed outside traditional academic publishing channels, such as reports, white papers, and technical documents released by organizations rather than through peer-reviewed journals. The article acknowledges Gregory Wrightstone "for contributing to these articles and providing valuable feedback" and Angela Wheeler, Joseph Coyle, Robert Burger, and Gordon Tomb for "formatting and grammar edits" (Engelbeen et al., 2024, p. 66). This acknowledgement offers a glimpse into the social infrastructure supporting the work, with community members contributing review, editing, and formatting assistance.

In addition to the CO_2 Coalition publication, the article was deposited in the Open Science Framework preprint repository (Engelbeen et al., 2024). OSF notes on the deposit: "OSF preprints are not peer reviewed and acceptance into OSF preprints is not an indicator of scholarly merit" (Engelbeen et al., 2024). The OSF deposit extends the report's reach into spaces where researchers search for academic work, while the CO2 Coalition publication channels it toward audiences already aligned with the organization's position. The two venues serve different functions: one signals openness and invites scrutiny, the other ensures circulation within their own community.

4.4.9 Analysis

This data journey documents an instance of quality assessment reuse directed inward at the ACC skeptic community. Engelbeen applies his community's own epistemic standards, preference for direct measurements, cross-dataset consistency checking, evaluation of measurement site conditions, to audit and ultimately refute a fellow skeptic's claim. This is not a case where skeptic methods produce a contrarian result. It is a case where applying the community's own standards produces a conclusion that aligns with mainstream science on the

CO_2 attribution question; not because Engelbeen defers to mainstream science, but because his methods lead him there independently.

The report's alignment with consensus on the attribution question does not reflect broader endorsement of mainstream climate science. The report deliberately addresses only the attribution question. The "beneficial CO_2" framing in the title is the CO2 Coalition's organizational position, not a claim the report analyzes. Engelbeen's rigorous treatment of Beck's data reflects the report's deliberately narrow scope rather than any inconsistency in his epistemic commitments. This journey also reveals something about how the ACC skeptic community handles internal disagreement. Rather than treating Beck's claim as outside the community's purview or simply dismissing it, Engelbeen engages with it through the community's own analytical methods and publishes the refutation through a community-affiliated venue. The community produces its own quality assessment mechanisms.

Beck's data fail to fulfill Engelbeen's trustworthiness criteria, as documented in Section 4.4.2.1. What is analytically significant here is not the failure itself but how Engelbeen demonstrates it. Engelbeen uses modern HLNUG monitoring data from approximately the same location as Beck's most significant historical measurements. By demonstrating that the Linden/Giessen site produces severely contaminated readings under modern conditions, Engelbeen argues that Beck's historical measurements at the same location were subject to the same contamination, undermining the site Beck relied upon.

Across all this analysis, Engelbeen works directly with primary data, downloading HLNUG monitoring data, comparing it against NOAA background stations, and manually extracting measurement ranges from historical tables. This reflects a commitment to knowing the data firsthand, a value that runs throughout the ACC skeptic community more broadly.

The publication venue is addressed in Section 4.4.8. What is worth noting analytically is the combination of the two deposit locations, CO2 Coalition and OSF preprint, which serves a dual function: the Coalition publication ensures circulation within the sympathetic community, while the OSF deposit signals transparency and invites scrutiny from outside it. OSF is primarily associated with open academic science, and depositing there lends the report a perception of legitimacy by placing it alongside peer-reviewed and pre-peer-reviewed scientific work, even though OSF itself does not evaluate content for scholarly merit. The platform's association with mainstream science lends the report a credibility the CO_2 Coalition venue alone cannot.

The co-authorship of this report is itself analytically significant. Dave Burton, whose sealevel.info journey is documented in Section 4.2, is a co-author alongside Engelbeen and another author. This collaboration illustrates how the ACC skeptic community functions as a knowledge network: individual members often collaborate on shared data reuse projects, reinforcing each other's work across different domains of climate skepticism, and combining different expertise.

What makes this journey analytically distinctive within the broader corpus is the direction of the audit. Across the full ethnography, ACC skeptic data reuse is directed overwhelmingly outward: at mainstream climate science, government datasets, media claims, and policy decisions. This journey is the only sustained instance in the corpus of a community member conducting an intensive, data-driven internal audit of a fellow skeptic's empirical claims. Community members do critique each other's work, comment threads occasionally surface disagreements and corrections. That this critique was produced collaboratively, published through a community-affiliated venue, and deposited in a public preprint repository rather than suppressed or ignored suggests the community has capacity for internal quality assessment, but, as of now, that capacity appears to operate only at the margins. Rather than producing consensus, the report's core attribution claim drew sustained pushback from the ACC skeptic community members committed to a natural origin for rising CO_2, exposing a genuine internal fault line rather than resolving it. This internal audit looks less like a community correcting itself than like one faction contesting another.

4.5 Replication and Quality Assessment Reuse: Academic Challenges to Policy (McKitrick - Nature Scientific Reports)

4.5.1 Background and Research Design:

Ross McKitrick, a Professor of Economics at the University of Guelph specializing in Environmental and Applied Microeconomics, began this data reuse journey with a puzzle in the climate-agriculture literature and a policy debate on the Social Cost of Carbon (SCC). Two prior studies, Challinor et al. (2014) (C14) and Moore et al. (2017) (M17), had analyzed the same underlying meta-analysis dataset of global crop model outputs but reported different outcomes. In C14's original analysis, the results implied a "moderate global net benefit from the warming associated with doubling the atmospheric CO_2 level", meaning slight yield gains under climate

warming (McKitrick, 2025, p. 1). In contrast, M17 re-analyzed that dataset and reached “much more pessimistic conclusions, projecting declining global crop yields due to the warming” (McKitrick, 2025, p. 1).

These divergent findings had direct policy consequences. The SCC is a dollar value used in federal regulatory cost-benefit analysis to represent the economic harm of emitting one additional ton of CO_2. A higher SCC justifies more stringent climate regulations by increasing the estimated benefits of emissions reductions. The Biden Administration raised its SCC estimate approximately fivefold, with a new 2030 SCC value of $220 under 2% discounting, of which $103, nearly half, is attributed to projected agricultural damages (Interagency Working Group on Social Cost of Carbon, United States Government (IWG), 2013; McKitrick, 2025; US Environmental Protection Agency, 2023). The EPA’s GIVE agricultural damage module, which drives much of this increase, is based on M17’s yield projections (McKitrick, 2025; US Environmental Protection Agency, 2023). This raised the stakes for understanding why C14 and M17 differed so much.

McKitrick’s interest in this issue started through his own prior research and a subsequent academic exchange. In 2020, he co-authored a study on climate sensitivity and agricultural productivity in SCC calculations (K. D. Dayaratna et al., 2020). A comment published in 2022 (Meyer, 2023) drew McKitrick’s attention to the M17 yield-damage findings. In a 2023 reply (K. Dayaratna & McKitrick, 2023) to Meyer’s comment, McKitrick and colleagues noted that C14 and M17 had used “the same data set” yet arrived at different conclusions, but they couldn’t fully investigate the discrepancy at the time because M17 “didn’t report enough detail to make that possible” (McKitrick, Interview, 2025). After concluding that reply, McKitrick decided to directly revisit the original C14 dataset and the M17 analysis to trace the source of the difference. This decision set the stage for a data reuse effort: to obtain the C14 dataset, determine any issues, correct them, if possible, replicate both studies’ analyses, and see how the results might change.

This account reconstructs McKitrick's data reuse from his published article, its supplementary materials and Mendeley deposit, and a semi-structured interview conducted during this study (September 2021–December 2025). The article was published in 2025; the account reflects the data and materials as deposited at the time of my review.

4.5.2 Seeking and selecting data

One immediate issue was that neither of the earlier teams (C14 or M17) had made the meta-analysis dataset publicly available. The C14 paper had indicated a website for the data, but by the time McKitrick looked, "the website where it was supposedly available didn't exist" (McKitrick, Interview, 2025). Obtaining the original data required personal communication with the original data author. McKitrick emailed the lead author of C14 and they "sent the data (an Excel spreadsheet) upon request." This Excel file, named MasterData_021013_IsFinal.xlsx, was the master database underlying C14's study, representing the "final" version of the compiled meta-analysis data from dozens of crop modeling studies. Additionally, McKitrick acquired a secondary data file, baselinetemps.csv, from Thomas Hertel via personal communication. This file supplied baseline temperature values for various locations or climate zones, likely to help parameterize or check climate scenario information in the analysis. McKitrick also obtained M17 regression results and simulation data from Moore via personal communication, a third informal data transfer that proved necessary for the replication analysis.

Table 4.1: Datasets obtained by McKitrick for his article.

Dataset	**How Obtained**	**Format**	**Purpose**
MasterData_021013_IsFinal.xlsx	Personal communication with A. Challinor	Excel (.xlsx)	C14 master meta-analysis database; baseline for C14 and M17 analyses
baselinetemps.csv	Personal communication with T. Hertel	CSV	Baseline ambient temperatures
M17 regression results and simulation data	Personal communication with F. Moore	Not specified	Replication of M17 analysis

Opening the C14 master spreadsheet revealed the central problem. McKitrick recalled, "As soon as I opened it I could see that they didn't have 1,722 complete records because so many changes were missing. I confirmed with [the author] that this was the correct data set" (McKitrick, Interview, 2025). C14 reported N=882 complete records but the dataset as supplied contained only 862, and the remaining 860 of the 1,722 entries incomplete (McKitrick, 2025, p. 5). In other words, while the dataset contained 1,722 entries (as reported in C14), nearly half (N=860) of those entries were incomplete, typically missing at least one key variable needed for analysis. The most frequently missing field was the change in atmospheric CO_2 concentration

associated with each crop yield projection. Other entries lacked changes in temperature or precipitation or lacked all three climate change variables.

McKitrick argued in the interview that neither C14 nor M17 adequately disclosed the extent of the incompleteness. McKitrick noted: “I was surprised neither [C14] nor M17 told the reader how many records were incomplete. They refer to N=1722 but that is misleading for understanding the regression model” (McKitrick, Interview, 2025).

Having the raw spreadsheet meant McKitrick had the same baseline data C14 and M17 used. The data journey thus moved into a recovery phase: tracking down the missing pieces for as many of those 860 incomplete records as possible to see if the missing variables “really were unavailable” (McKitrick, Interview, 2025). A table summarizing the datasets used across this journey and how the sample sizes changed is provided in Table 4.2.

4.5.3 Processing and transforming data

4.5.3.1 Recovering and completing the dataset

McKitrick began recovering the missing values from each underlying study cited in C14’s meta-analysis. Over roughly two months from December 2023 to the end of February 2024, he systematically worked through the references and supplementary materials of the original studies to find the omitted variables. In many cases, the missing value was available in the text, tables, or figures of the underlying paper but had not been transcribed by the C14 authors. “In a few cases the change was available in the underlying paper even though it hadn’t been copied over to the [C14] spreadsheet,” he explained, “in others it was a matter of looking up the IPCC tables and matching the scenario name and time span” (McKitrick, Interview, 2025). Several studies had described scenarios by name without explicitly stating the CO_2 increase; to fill these, McKitrick consulted the corresponding IPCC assessment reports’ appendices to find the CO_2 ppm change over the period in question.

McKitrick chose to do all the recovery work himself rather than delegating: “I didn’t have a research assistant but I wanted to do the work myself anyway so I could look at the underlying studies and get to know the data for myself” (McKitrick, Interview, 2025). Working study by study, he kept detailed notes of every value added or corrected, later compiled into the article’s online supplement: “I kept notes and published them in the online supplement to the article” (McKitrick, Interview, 2025). By February 2024, McKitrick had retrieved data for

roughly 360 previously incomplete records, growing the usable sample from 862 to 1,222, a roughly 40% increase.

McKitrick saved this expanded dataset as a new version of the master spreadsheet, MasterData_021013_IsFinal.ISO3.xlsx, adding ISO3 country codes to each entry, standardized location identifiers that C14 did not have.

Table 4.2: Sample size progression across McKitrick's dataset recovery process.

Dataset	N	Notes
C14 database (total received)	1,722	As received from Challinor
Complete records	862	
Recovered by McKitrick	+ 360	Retrieved from underlying studies and IPCC tables
McKitrick's expanded dataset	1,222	Deposited on Mendeley Data

4.5.3.2 Building the analytical codebase

With the corrected dataset in hand, McKitrick structured his analytical workflow in R, organized around a main script named code.C14.main.R, which "calls and runs all functions and subroutines" for the analysis (README). This main script allows setting key parameters, climate sensitivity and baseline CO_2 concentration, before a series of function files and subroutine files execute specific tasks. The function files (prefixed code.f.) each "create functions to estimate models or tabulate data," handling tasks like regression estimation and assembling result tables (README). The subroutine files (prefixed code.s.) contain reusable code segments sourced at appropriate points in the main script. McKitrick ensured that all code and data files were in one folder with an output directory to catch all result tables and figures. "I try to provide turn-key code with a local folder hierarchy," he explained, "It should be possible to download the archive and run it in R Studio with 1-click" (McKitrick, Interview, 2025).

4.5.4 Analyzing and Interpreting Data

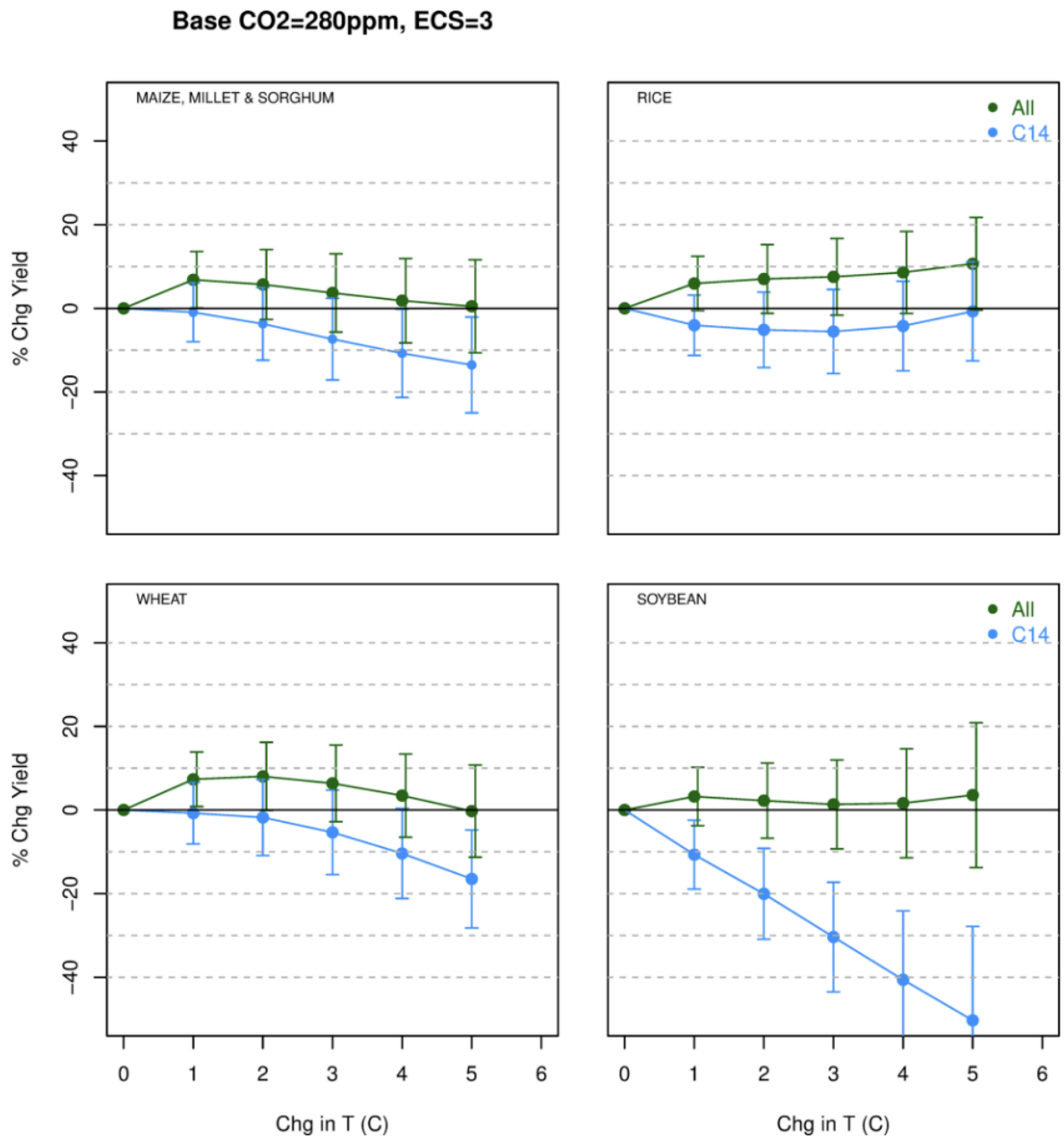
McKitrick designed his analytical workflow to replicate the analyses of both C14 and M17 as faithfully as possible, while also exploring the effect of the expanded data. This method meant estimating two main regression models: one resembling C14's approach and one following M17's approach. M17 had a more complex model than C14, incorporating interaction

terms between variables and a non-linear term for CO_2 fertilization imposing diminishing returns. C14 had used primarily linear terms and treated CO_2 fertilization as a linear additive benefit. Because neither study reported all regression details, McKitrick ran both styles of analysis using his own code and compared results.

He first replicated the original studies on the incomplete original dataset of 862 records to verify he could reproduce the published findings. This confirmed the earlier patterns: the C14-style model suggested mild benefits of warming, while the M17 model produced predominantly negative yield impacts. The correlation between M17's projections and McKitrick's replication on the C14 dataset was 0.96, which McKitrick treats as establishing the validity of the replication (McKitrick, 2025, p. 5).

McKitrick then re-ran both regression approaches on the expanded 1,222-record dataset. The yield decline narrative of M17 disappeared. Under both modeling approaches, global average yield changes were neutral to positive even at substantial warming increments. He reports that "global average yield gains of all crop types under CO_2-induced warming are positive even out to 5 °C warming" (McKitrick, 2025, p. 1). In other words, once the data gaps were filled, the meta-analysis suggested zero or modest gains of crop yields on average, thanks largely to CO_2 fertilization effects balancing or outweighing temperature-driven losses. McKitrick concluded that "the climate change-related agricultural damage estimates in M17 are too pessimistic and the large implied revisions to the SCC are unsupported" (McKitrick, 2025, p. 1). Image 4.13 below shows the yield change simulations.

Image 4.13: Figure 1 from McKitrick (2025) showing yield change simulations (%) for four crop types, maize/millet/sorghum, rice, wheat, and soybean, at warming levels from 1°C to 6°C. Blue lines show results based on the incomplete C14 dataset (N=862); green lines show results based on the expanded dataset (N=1,222). The shift from predominantly negative to neutral or positive yield projections across crop types illustrates the effect of the dataset expansion. Source: McKitrick (2025, Figure 1).



4.5.5 Managing, documenting, preserving, and sharing data

Every decision made during the dataset recovery process was recorded and published in the article's online supplement, providing a trail of the dataset expansion. In the supplement's

section “Notes on expansion of C14 data,” provided as a Word document, McKitrick itemized the adjustments for each source, although without full citations to each source (McKitrick, 2025). He instead directed readers to the C14 paper, which has the “studies listed on pp. 18-29” (McKitrick, 2025, Supplementary Information). Some corrections were straightforward, for Brassard & Singh (2007), two yield change entries appeared mis-transcribed and he corrected them. In other cases, entire batches of missing values were filled in from the original papers, like in Reyenga et al. (1999), and Rosenzweig & Parry (1994), CO_2 increase values “were available but not included” in the C14 data file (McKitrick, 2025, Supplementary Information). Some cases required deriving values from reported end-point CO_2 levels.

Once the article was published, McKitrick deposited data and code, including Excel files, CSV, R scripts, and README, with Mendeley Data. The dataset and code package was cited in the paper as “Data and code to accompany ‘Extended Crop Yield Meta-analysis Data do not Support Upward SCC Revision’, Mendeley Data, V1,” with a DOI. The decision to share publicly was driven by both practical and principled considerations: “I always publish my data and code,” he stated. “It’s partly so I don’t have to spend time afterwards fielding inquiries but it’s also related to research integrity” (McKitrick, Interview, 2025). The R codebase is structured for reproducibility, with a turn-key main script, clearly named function and subroutine files, and a README explaining the folder hierarchy and parameter settings. All code and data files are organized in a single folder with a dedicated output directory.

4.5.6 Attributing and citing data

McKitrick attributes the C14 dataset to Challinor et al. (Challinor et al., 2014) throughout the paper and supplement, noting that the data were obtained via personal communication with Dr. Challinor. The baseline temperature file is attributed to Dr. Thomas Hertel, and M17 regression results and simulation data are attributed to F. Moore, both via personal communication in-text. IPCC assessment reports consulted during the recovery process are cited in the supplement’s notes on dataset expansion. The expanded dataset itself is cited in the paper as a Mendeley Data deposit with a DOI.

4.5.7 Publishing and sharing research

The findings were published as a peer-reviewed article in *Scientific Reports* in 2025 (McKitrick, 2025). *Scientific Reports* is an open-access journal published by Nature Portfolio

covering research across the natural and social sciences (*Scientific Reports*, 2026). Its open-access model means the article is freely available without a paywall, consistent with McKitrick's stated commitment to making his data and methods publicly accessible. The paper's competing interests declaration notes that McKitrick's Fraser Institute and Global Warming Policy Foundation affiliations are unpaid and that neither organization had any knowledge of, involvement with, or input into the research (McKitrick, 2025, p. 8).

4.5.8 Analysis

This data journey documents data reuse directed at a policy-relevant empirical claim. McKitrick recovered missing data from an incomplete dataset, replicated two prior analyses, and produced a more complete and accessible version of the evidence base, directed at challenging a part of the empirical foundation of a federal climate policy decision.

McKitrick is a tenured professor of economics at the University of Guelph and Senio, he operates within mainstream academic structures. He was recently appointed to the U.S. Department of Energy's 2025 Climate Working Group alongside other figures associated with ACC skepticism (Climate Working Group, 2025). His academic position gives him easier access to peer-reviewed venues and the professional networks through which informal data transfers happen. The C14 dataset, M17 simulation data, and baseline temperature file were all obtained through personal communication with researchers McKitrick could reach through academic channels. Someone outside these networks would have found both acquiring the data and publishing in a peer-reviewed venue considerably more difficult, if not impossible.

The journey reveals how policy debates prompt much data reuse. McKitrick's investigation was prompted not only by scientific curiosity but by the Biden Administration's SCC revisions, which relied on M17's agricultural damage projections. By demonstrating that those findings rested on incomplete data and whose conclusions reversed when gaps were filled, his analysis directly challenged the empirical basis for the SCC increase. Data reuse here is simultaneously a scientific and a policy intervention, and those are not mutually exclusive nor unique to ACC skeptics. Policy-oriented research is a legitimate mode of scholarly engagement.

McKitrick's transparency commitments were developed through prior experience. His "thoughts on this were influenced in part from [his] coauthor's…and [his] long battles with Michael Mann over the data and code used to produce his 1998 hockey stick graph" (McKitrick, Interview, 2025). These experiences led him to the conviction that "studies for which the data

and methods are unpublished and/or unavailable should be disqualified for use in public policy formation" (McKitrick, Interview, 2025). This claim is both an epistemological position and a political one. It asserts that scientific claims must be reproducible to be credible, and it positions data withholding as a form of gatekeeping that enables questionable findings to influence policy. He operationalizes this conviction through extensive documentation, traceable data versions, and turn-key reproducibility code.

The data recovery process itself reveals infrastructural friction points. The C14 dataset was theoretically public, the paper mentioned a website, but that website no longer existed. Access required personal communication with the original author, who provided data "upon request" (McKitrick, Interview, 2025). This informal researcher-to-researcher transfer highlights how much scientific data remains outside formal repositories, accessible only through professional networks and goodwill. McKitrick's ability to obtain the data reflects his position within academic networks; someone without those connections might have been unable to acquire the data at all. Once obtained, the dataset's incompleteness was immediately apparent, compounded by a further discrepancy: C14 had reported 882 complete records, but the dataset as supplied contained only 862. Neither C14 nor M17 had disclosed this clearly according to McKitrick. This opacity about data limitations is precisely what McKitrick's transparency opposes. The process of correcting the dataset represents detective work that many researchers do not undertake. McKitrick's decision to do this work himself rather than delegating to a research assistant reflects both resource constraints and a belief that hands-on analysis produces better understanding (McKitrick, Interview, 2025).

> McKitrick's frustration with data dissemination practices is documented directly:
>
> Yes many public data sets are very difficult to use, seemingly on purpose. Climate data sets are often disseminated in netCDF format even if it's just a small matrix of values that could have been posted as .csv or text format. And many studies nowadays disseminate archives using GitHub where the program is broken up into many small files with little documentation so it's hard to parse what is done with the data. Government websites typically make it hard to get the data you want because they have lousy search engines and you get stuck in endless pages of metadata without clear navigation to the actual numbers (McKitrick, Interview, 2025).

These frustrations motivated him to build YourEnvironment.ca as a demonstration of how he thinks environmental data should be disseminated.

The peer-reviewed publication in Scientific *Reports* combined with open data and code deposit on Mendeley Data represents a hybrid legitimation strategy. Publishing in a Nature Portfolio journal confers mainstream scientific credibility. Depositing complete replication materials signals commitment to transparency and invites scrutiny. Both moves are consistent with McKitrick's stated epistemological position: that scientific claims used in policy formation must be reproducible and publicly accessible.

What McKitrick produced is both a technical accomplishment and a political intervention. He produced a more complete and accessible dataset than its predecessor, that future researchers will likely use. In that sense his intervention is positive for the broader evidence base, not only for skeptical arguments. It also presents evidence that a core input to Biden Administration climate policy rested on an incomplete dataset whose apparent implications changed once McKitrick filled the gaps.

4.6 Conclusion

The four data journeys documented in this chapter reveal the range of data practices within the ACC skeptic community, from rapid reactive reuse to sustained infrastructure building. Table 4.3 summarizes each journey across the parallel dimensions documented in Sections 4.2 through 4.5, enabling comparison across cases.

Table 4.3: Comparative summary of four ACC skeptic data reuse journeys.

	Burton (sealevel.info)	**Burnett (Climate Realism)**	**Engelbeen (CO_2 Coalition)**	**McKitrick (Scientific Reports)**
Reuse type	Infrastructural: database and visualization tool	Rapid reactive: media response	Intra-community quality assessment	Replication and quality assessment
Prompt	Hansen's 3mm/year claim; NC-20 coastal policy debate	*Hill* article claiming climate harms apple production	Community debate about CO_2 origins	C14/M17 discrepancy in crop yield literature; Biden SCC revisions
Data sources	PSMSL, NOAA (API), IMGW (direct contact)	FAO FAOSTAT, EPA Heat Wave Index	HLNUG (Linden/ Giessen), NOAA background stations,	C14 dataset, baselinetemps.csv, M17 data all

			Law Dome ice cores	through personal communication
Processing	Perl scripts; geographic weighting; autocorrelation handling; JSON integration across sources	FAOSTAT Compare interface; manual cursor calculations; no underlying data downloaded	Manual cross-dataset comparison; Excel graphs	Study-by-study data recovery; IPCC table consultation; R codebase
Analysis	Interactive regression tool; user-configurable weighted averages; 15 station subsets	Manual percentage calculations from hover tooltips; visual interpretation	Four comparisons across datasets; assessment of measurement site conditions	Two regression models on incomplete (N=862) then expanded (N=1,222) dataset
Documentation	Automated scripts; TLIB version control; JSON with extensive metadata; public changelog	None: data exists only in FAOSTAT interface during analysis	Personal hard drives and cloud; Excel/Paint graphs; no formal repository deposit	Detailed supplement notes; turn-key R codebase with README; Mendeley Data deposit with DOI
Attribution	Consistent; multiple citation formats; IMGW cited directly for Polish data	Links to FAOSTAT; EPA caption with URL	Formal citations alongside "(ongoing)" designations; URL-only for some sources	Full citations throughout; personal communication attribution; IPCC reports cited in supplement
Publication venue	Self-published web tool (sealevel.info); continuous updates	Heartland Institute blog (Climate Realism); same-day publication	CO_2 Coalition grey literature report and OSF preprint	*Scientific Reports* (peer-reviewed) and Mendeley Data
Policy stakes	NC-20 coastal development regulations; sea level rise projections	Media narratives about climate harms to agriculture	Anthropogenic CO_2 attribution debate within ACC skeptic community	Biden Administration SCC estimate; federal climate regulations

Taken together, these journeys illustrate that ACC skeptic data reuse is not a single mode of practice, but a range of activities shaped by the reuser's technical capabilities, institutional positioning, publication context, and the specific claim being contested. Burton builds infrastructure; Burnett responds to media; Engelbeen audits a fellow skeptic; McKitrick replicates and expands a dataset to contest a policy input. Each journey engages with open data infrastructure differently, acquiring, processing, analyzing, and publishing in ways that reflect both individual commitments and community-level epistemic values. The patterns and diversity

of these practices are examined across the full data corpus in Chapter 5. The knowledge infrastructures these reusers rely on, and their entangled relationships with the broader climate science infrastructure, are developed in Chapter 6.

Chapter 5 Data Practices of ACC Skeptics

This chapter addresses the first research question guiding this dissertation: What are the data practices of ACC skeptics? And the subquestion, how does this community develop trust or distrust in the data they reuse? Where chapter 4 traced four individual data journeys in depth, this chapter moves across the full data corpus. Across my monitoring environment, I observed 99 data reusers through semi-structured interviews, content analysis, and participant observation.

The chapter follows the data reuse journey framework developed in Chapter 3 and used in Chapter 4, moving through data reuse prompts and design; seeking and selecting data; assessing data and repository trustworthiness; processing and transforming data; analyzing, interpreting, and abstracting data; managing, documenting, and preserving data; attributing and citing data; publishing and sharing data; and publishing and sharing research. The order of these phases is not prescriptive. In practice, data reuse in this community was iterative, recursive, and variable. Not all phases occurred in every journey, or at the very least, were not visible to me. Some phases overlapped, and boundaries were often blurry. The framework is a writing tool for organizing observations, not a model for how reuse unfolded in practice. Within each phase, findings from the four case studies are woven alongside evidence from the broader data corpus to identify patterns that emerged across the community.[4]

The goal here is to identify patterns that hold beyond the four cases and characterize ACC skeptic data practices more broadly. However, the evidentiary basis varies across the data practice phases. Several left visible traces in blog posts, comment threads, and published outputs, and so are well represented in the content analysis corpus: what prompted and framed a reuse episode, how reusers assessed data quality and trustworthiness, how they analyzed and interpreted data, how they attributed and cited it, and how they published and shared research. Two phases are mixed. Which datasets were selected and what data were shared is visible in the

[4] Portions of this chapter draw on previously published work, including (Pasquetto et al., 2026; Wofford & Thomer, 2023)

public record, but how datasets were first found and why particular sharing decisions were made come primarily from interviews. The least publicly visible data practice phases are processing and transforming data and managing, documenting, and preserving data; the evidence for these draws more heavily on the eight semi-structured interviews and the four data journeys. Where a section rests primarily on interviews rather than the broader corpus, I note it explicitly.

Across the corpus, several overarching patterns recur. ACC skeptics tended to prefer "raw" and observational data over processed data or modeled outputs and directed their distrust primarily at institutions' interpretive and processing decisions rather than underlying measurements. Data reuse was predominantly reactive, prompted by circulating media claims, policy documents, or community debates, and was oriented toward independent verification of those claims. An audit orientation ran throughout: reusers approach data as something to be checked against official narratives. These practices were sustained by a hybrid social and technical infrastructure examined in Chapter 6.

5.1 Data reuse prompt and project design

Data reuse prompt and project design refers to the factors that contributed to the data reuser beginning the project as well as the initial framing of the project.

5.1.1 Research topics

These topics organized their broader research agenda. Each topic was associated with specific arguments, particular empirical claims the community made within that topic area. For instance, a topic like sea level was a recurring area of investigation, while an argument like sea level rise is not accelerating is the specific claim a given reuse episode might address. Four research topics dominated the corpus; each associated with arguments that challenged the empirical foundations of ACC.

1. The first research topic concerns observational climate variables, including global temperature, sea ice extent, sea level, natural disaster rates, snow cover, heatwaves, and wildfires. Arguments within this topic challenged the observational evidence used to support anthropogenic climate change (ACC). These arguments could directly dispute reported trends, question whether the selected variables are appropriate indicators of ACC, argue that observed changes are not statistically significant, contend that the variables

cannot be measured accurately, or dispute claims that the observed changes will produce adverse effects.

2. The second research topic concerned climate modeling. Arguments within this topic held that climate models cannot accurately predict future, or even current, climate conditions because of perceived limitations and uncertainties in model construction and climate dynamics. These arguments commonly cited difficulties in representing cloud processes, the inherent complexity of climate systems, and the role of chaos in limiting the reliability of long-range predictions.
3. The third research topic concerns the causes of observed climate variability. Arguments within this topic held that anthropogenic CO_2 emissions are not the primary driver of observed climate change and that natural processes provide a better explanation for the observed record. These arguments commonly cited factors such as El Niño and other ocean oscillations, solar activity and sunspots, cloud and ice albedo dynamics, and the claim that increases in atmospheric CO_2 follow rather than drive temperature changes.
4. The fourth research topic concerns fossil fuels and energy policy. Arguments within this topic held that fossil fuels are superior to renewable energy alternatives, have contributed substantially to human welfare, and therefore make climate policies aimed at reducing fossil fuel use costly or counterproductive. These arguments commonly cited rising income and life expectancy alongside increased fossil fuel use, declining rates of climate-related mortality and disease, and improvements in food production and access to clean water.

These four research topics were not independent of one another. They were often operationalized and deployed together. And even when not directly connected, there were underlying ways one argument impacted the other. For instance, arguing that observational trends do not support ACC strengthens the claim that climate models are unreliable.

These research topics constituted the community's standing research agenda during my digital ethnography, sustained through years of published arguments, organizational missions, community websites, and ongoing debate within the ACC skeptic knowledge infrastructure. Regardless of whether the interest in the topic was longstanding or newly encountered, most episodes of data reuse began with a specific occasion, event, or prompt, internal or external, that

initiated the research process. The occasion that set a specific data reuse episode in motion is where we now turn.

5.1.2 What prompts data reuse

Every data reuse episode began somewhere: perhaps with an encounter, a gap, a dispute, a question. These are not unique to ACC skeptics; all data reuse begins with something. The most common pattern across the corpus was reactive: a reuse episode began when something in a reuser's information environment caught their attention. This could be a newspaper article, a social media post, a policy debate, a gap in existing data infrastructure, a discrepancy, or a dispute within the ACC community itself. And in response, the data reuser was inspired to go to the data and judge whatever caught their attention against the data. In other words, data reusers encountered a claim or a condition, went to the data, and saw whether it holds. Reactive here still means deliberate, for instance, not just passively reacting to things: for some it was an explicit mission to check public claims about climate as they appear. What reusers then did with the data is taken up in the sections that follow; the concern here is only what sets the work in motion.

This prevalence of reactivity may be partly an artifact of my monitoring environment: blog posts, news articles, and social media make the prompt-to-reuse chain visible in ways that more private or internally motivated reuse does not. However, this reactive engagement pattern was also seen in the interviews I conducted. Whether reactive reuse is more common than internally motivated reuse across the community is difficult to establish from the data available. Regardless, the reason media articles matter, gaps matter, and discrepancies matter is because this community believes that climate science has real stakes for energy policy, regulation, and public life.

Most often, the occasion was a claim someone else had made. It might have surfaced in a news article, a social media post, a government report, a journal, or a podcast, and the reuser converted it into a checkable question, do the data show that? before going to the data to find out. They might go to the same data the claim relied on, to a different dataset, or, when the claim cited no data at all, to data of their own choosing. Burnett's data reuse journey illustrated this pattern clearly where he encountered a news article "Climate change is hampering US apple quality and output: Study" (Udasin, 2025) and framed his entire data reuse as a response to this news article claim. Daily, he conducted a Google News Search for "climate change" to identify

stories to respond to: “I just pick what I think is the most interesting story for the day” (Burnett, Interview, 2025).

This is an organizational routine as much as a personal habit. James Taylor, the president of The Heartland Institute reiterated the organization’s data reuse workflow for their *Climate Realism* project:

> On a daily basis, what we do at Heartland is we do a Google News search every day for climate change and we see what are the top stories, the narratives that are being put out. And then let’s say, okay, climate change is threatening cranberry production. Then we’ll go and we’ll find…you go look at the actual data and you can see cranberry production sets records almost every year (Taylor, Interview, 2025).

He framed this practice as a commitment to verification: “I’m not going to leave it just because I heard it on CNN” (Taylor, Interview, 2025). For instance, in a January 2025 post, ACC012 wrote a Heartland associated blog post as a fact check on an Associated Press report, based on the EU’s Copernicus service, that January 2025 was the hottest January on record (ACC012, Blog, 2025). Against that claim the author reused NASA satellite data to show a 0.34°C drop from the previous January. In this case, a media claim was found, and satellite data supplied the rebuttal.

A similar story unfolded across other community members unassociated with The Heartland Institute. Tom Nelson, who runs *The Tom Nelson Podcast* (Nelson, n.d.) and producer of *Climate: The Movie* (Durkin & O’Toole, 2023), described the same reactive orientation: “I see the claim and then I just very consistently, I search Twitter or Google to see what the counterclaims are, or to see if it holds up” (Nelson, Interview, 2025). He walked through one hypothetical example:

> They will say that, “This storm in India is the worst storm ever in India.” Constantly, there’s all these mainstream articles. So I spend a lot of time just checking out. Is it really the worst ever? And some of it is going to be the worst ever recorded because it’s only been recorded for 150 years, and some of it is. But so often, if they make a claim like that, in two minutes of checking you can find out, no, they forgot that it really was worse back then. Constantly, they’re saying that this current storm was a 500-year storm or a 1,000-year storm. I’m constantly checking in on that because I almost never believe it,

and they almost never have done their homework to prove that it really was (Nelson, Interview, 2025).

He characterized part of his role in the community as "debunking climate change claims online almost daily" ("Tom Nelson," n.d.).

The source need not be the media; a government report or journal article set the same process going. Taylor described responding to "a United Nations' report, some U.S. government climate report or just in the media" (Taylor, Interview, 2025) and monitored scientific journals directly: "there are science journals that I will, on a regular basis, 'All right, let me see Geophysical Research Letters. Let me see what they've published this past month'" (Taylor, Interview, 2025). When the prompt was a policy-relevant document the stakes were especially explicit from the outset. In one example, writing on a blog, ACC013 recounted how his own line of research began with what he calls "an exceptionally influential study" that he read: Sherwood et al.'s 2020 assessment of equilibrium climate sensitivity, whose narrowed range was largely carried into the 2021 IPCC report. That study sent him back to examine its statistical methodology, where he identified what he argued were several errors, and the dispute has since run through successive journal exchanges (ACC013, Blog, 2025). The journal article supplied both the data reuse prompt and the target, and the policy stakes were explicit from the outset, since the figure he disputed had already been adopted into an IPCC assessment.

Not every occasion was a claim someone else has made. Sometimes it was something the reuser notices in the data themselves. Dave Burton's journey began with a gap he noticed while browsing NOAA's tide gauge pages: "I scrolled to the bottom looking for the average just out of curiosity, there's just no average" (Burton, Interview, 2025). The absence, and its implications for sea level rise policy debates, motivated a sustained infrastructural commitment, the construction of sealevel.info. Burton also describes a general habit of noticing inconsistency:

> I have a nasty, suspicious mind. I have a habit of cross-checking things. If I hear two different statistics, which are related to one another, I like to check them for consistency and see if they are compatible. And it turns out when you do that, when you get in the habit of doing that, you often find that people are believing things that are contradictory (Burton, Interview, 2025).

An inconsistency between related figures was its own starting point; McKitrick's case began the same way, with the question of why one dataset yielded different results in two papers. As he explained:

> I had coauthored a paper in 2020 on the role of agricultural productivity in social cost of carbon calculations. Meyer published a comment on it in 2022 in which he mentioned the M17 findings. Those were also very influential on the Biden Administration SCC revisions. In our reply to his comment we pointed out that both studies used the same data set but we didn't explore the origins of the differences in results, partly because M17 didn't report enough detail to make that possible. After our reply was done I decided to go back to the C14 and M17 papers to try to figure out why they drew different conclusions (McKitrick, Interview, 2025).

The discrepancy a reuser notices was not always between two numbers; sometimes it was between an institution's data now and the same data in the past. Writing on his blog, ACC014 noticed that a NOAA page on long-term tornado trends had, in his word, been "disappeared" (ACC014, Blog, 2021). The NOAA page had shown no rise in strong tornadoes since the 1970s, a decline in the most violent categories, and had explained that the growing number of total reports reflected better detection of weak tornadoes, not more storms. ACC014 recovered the archived version of the NOAA page and set it against NOAA's current presentation, which he read as implying a rising trend stripped of those caveats (ACC014, Blog, 2021). What set the episode in motion was the disappearance itself, and the inconsistency he saw in it: an institution treating the tornado record as too unreliable to chart while still publishing confident historical comparisons for temperature, hurricanes, and ice.

In other cases, the occasion was a discrepancy between research and on the ground experiences. For Jim Steele, a published study on butterfly populations that did not match his own field observations became the starting point only after its authors refused his request for the data:

> I ask, 'Can I get your data? So I want to retry your study.' And she said, 'I'm not giving you my data.' She refused to do it. And then her husband, who was also a butterfly expert, said, 'Oh, you wouldn't get anything if we gave you the data or not.' And they did everything they could to prevent me from repeating what she had done. So that just lit a fire under my butt. She was getting a lot of fanfare across all the media about, 'Look,

> everything is moving further north due to climate change.' And that wasn't what the study really showed. So those kinds of things are what motivated me, and I feel like if you have a bad analysis, you're going to make bad remedies, and you really have to understand the climate (Steele, Interview, 2025).

As the quote shows, the refusal motivated him to start conducting his own research.

And sometimes the occasion originated inside the community. Engelbeen's investigation responded to claims made by other ACC skeptics: "repeated publications of claims within the ACC skeptical community that the current increase of CO_2 in the atmosphere was mainly natural, based on wrong data, wrong interpretation of real world processes" (Engelbeen, Interview, 2025). Not all this reuse was directed outward; some applied the community's evidentiary standards to its own members.

Across these occasions the orientation was practically constant: the reuser goes to the data to test a claim or the data themselves, and, as Taylor noted, rarely concluded the claim was right: "I can't recall looking at something as the daily climate scare of the day and saying, oh wow, I guess they're right on that" (Taylor, interview). Whether that reflected a genuine pattern or confirmation bias in how they select and read data falls outside the scope of this study.

5.1.3 Research design and planning

The responsiveness traced in the previous section was, in part, deliberate. For the community's organized actors, it was a designed workflow: the daily news scans and the standing *Climate Realism* project treat answering public claims as the work itself. For individuals it could be more of a habit than strategy, a reflex of long-standing skepticism. What matters here is that the same deliberateness shapes not just which claims a reuser examined, but how the investigation that follows was designed. As a reminder, comparative reuse assesses similarities and differences between datasets, for example by checking observations against a model or one dataset against another. It requires relatively little knowledge of how the data were made, and it is the routine, everyday practice. Integrative reuse is more demanding. It repurposes data to answer a new question, identifying trends, correlations, or patterns, often by combining sources or running new analyses. Because the reuser must understand how the data were produced, integrative reuse is rarer and calls for deeper expertise. Replication belongs with the integrative end: reproducing someone else's analysis takes the same depth of knowledge, but the aim is to test whether a finding holds rather than to ask something new. Two further purposes sit outside

this continuum, because they are not, strictly, ways of analyzing data. Quality assessment reuses a dataset to evaluate its reliability; it is a matter of judging the data. Infrastructural reuse builds the tools and databases through which others reuse data in turn.

In this community the comparative end was by far the more common. Most reuse checked observational data against a claim: does the record actually show the warming, the rise in disasters, or the sea-level acceleration that has been claimed? Integrative reuse, the novel reanalysis that answers a new question, was the work of a smaller number of reusers.

Across of these goals, an audit orientation was present to varying degrees. Quality assessment was its most direct expression: the explicit aim was to check whether official data and methods hold up under independent scrutiny. Comparative reuse functioned less as neutral comparison than as an audit of institutional claims, structured around a prior question of whether the official account was accurate. Even integrative reuse was typically framed as checking whether the data support what institutions and media said they show. The audit orientation shaped how each type of data reuse was motivated and carried out.

Once a reuser identified something worth investigating or researching, the design of their research often reflected a shared orientation: go directly to the data, do not trust institutional processing, and verify the claim yourself. The orientation was often articulated through the language of Richard Feynman, whose insistence on the primacy of observation and experiment recurred across the community as a touchstone for their research philosophy, and was operationalized in design. Feynman is invoked directly 55 times across the full corpus, reflecting his recurrence as a reference point in the material I analyzed. In one podcast, a community member quoted Feynman directly before describing his own approach:

> Now, I'm going to put in here a quote from Nobel Prize physicist Richard Feynman. He said, [paraphrased] "give all of the information to help others to judge the value of your contribution, not just the information that leads to judgment in one particular direction or another. It doesn't matter how beautiful your theory is, it doesn't matter how smart you are, if it doesn't agree with experiment or data, it's wrong." And so, I try to train graduate students to find as much data as possible to see how these claims pan out (ACC015, Nelson Podcast Interview, 2024).

The closing line of the Feynman quote, "that a theory at odds with the data is wrong" was the most repeated Feynman invocation in my data, recurring nearly verbatim across blogs,

Reddit threads, and podcasts. The second common use of Feynman reveals a distrust of experts: "Science is the belief in the ignorance of expertise." Others used it in their work, treating Feynman's quote as a verdict on consensus itself: Feynman, COMM001 writes, "was referring to consensus science, the modern type of political science…It is not empirical science" (COMM001, Blog, 2021).

Together these quotes represented a specific stance and dictated specific behavior by ACC skeptics in how they conducted data reuse. It positioned their work as an audit or correction of a field that they believed puts theory before observation and interpretation before data, and expert consensus before scrutiny. When reusers described their work as an audit, they were drawing on this logic. The job was not to theorize but to check whether the data support the claims being made.

Multiple reusers describe their research in explicitly audit-like terms. On the Nelson Podcast, ACC016 described his project in these terms directly: "I decided that I would do an audit of the temperatures of the Climate Change Institute and get some real data from real stations on the ground" (ACC016, Nelson Podcast, 2025). He wanted to audit the Institute "the same way [he] got audited in the circulation magazine business" (ACC016, Nelson Podcast, 2025). What he imported is not just the vocabulary of auditing but its institutional logic: a regime of external verification carried over wholesale from one industry and applied to another.

> In the publishing industry, it is common to be subjected to rigorous audits for advertising purposes (conducted by the ABC or BPA audit bureaus) in order to justify ad rates, as well as postal audits to qualify for substantial discounts like second class postage or what is now called magazine rates. None of this auditing is present in the climate data claims realm. To provide accountability to their audiences, the data presented by the Climate Change Institute of the University of Maine, the NOAA Models (CFS and others such as ERA-5), the UN IPCC, and the Copernicus Institute, ought to be subjected to a similar rigorous auditing procedure by an independent body. [ACC016] aims to make a contribution by helping to establish some accountability in this area, which is currently entirely lacking (ACC016, Slides, 2025).

This move recasted a set of federal and institutional scientific bodies as unaudited institutions that need to be held accountable. The publishing industry auditors refused to take a claimed figure on trust and instead inspected the underlying reality directly:

> The auditors, they had a very interesting approach. They would say, "okay, you say you print this much… how much paper do you have in your warehouse? Let's go look." They would actually want to walk into the warehouse and see the rolls of paper. So, in the same way, I want to see what is the temperature at individual stations around the world (ACC016, Nelson Podcast, 2025).

The audit was frequently policy motivated from the start: "you get things like the Inflation Reduction Act, which is really a climate bill…so I decided that I would do an audit" (ACC016, Nelson Podcast, 2025). Even though, he insisted his audit "has nothing to do with human induced climate change, climate science in general, anthropogenic global warming" (ACC016, Nelson Podcast, 2025). The audit posture was doing exactly that work: it lets data-checking be presented as a neutral exercise, held apart from the politics that occasioned it.

ACC031 audited through the lens of his own discipline, econometrics, and in doing so turns the distance between fields into a credential. He described climate science as relying heavily on statistical methods that originated outside the field: "Climate science is a lot like economics in terms of the tools that people use. It is to a large extent applied statistical analysis…the techniques are applied statistics, and a lot of those techniques came out of econometrics" (ACC031, Nelson Podcast, 2022). The problem, he argued, is that climate scientists use these tools without fully understanding them: "I guess it just means that they're using tools that they don't understand" (ACC031, Nelson Podcast, 2022). His critique of the optimal fingerprinting method turned on exactly this:

> You guys [climate scientists] invoked what's called the Gauss-Markov theorem which is elementary in econometrics…I taught it last week to a third year class…you guys invoked it but you got it wrong, you left out an important condition, your method doesn't satisfy the Gauss-Markov theorem (ACC031, Nelson Podcast, 2022).

He framed the broader design question as one of disciplinary accountability: "The question isn't why is an economist weighing in on this, but why didn't people in the field notice these problems" (ACC031, Nelson Podcast, 2022). The domain distance that separates econometrics from climate science, ordinarily a barrier to reusing a field's data and methods, is recast as the thing that authorizes the audit. ACC031 claimed to see what the field that adopted those techniques could not. The same stance shaped his teaching, where he had students plot a

century of daily high temperatures from local stations against model predictions, passing on the audit logic as well as practicing it (ACC031, Nelson Podcast, 2022).

Several design principles recur. Skeptics often framed their work as verification or audit rather than discovery. And they tended to treat outsider expertise, like engineering, process chemistry, and data analytics, as an asset rather than a limitation. They portrayed themselves as external scrutiny on a field they believe has been corrupted by flawed incentives.

5.2 Finding and selecting data

The next category of data reuse practices involves finding and selecting data for reuse, which encompasses the activities through which reusers locate, evaluate, and retrieve datasets for their purposes. How reusers find data is one of the harder things to observe directly. Publicly published reuse, blog posts, comment threads, videos, rarely records how a dataset was found, so the discovery process is largely invisible in the published record, and the account here rests primarily on my semi-structured interviews together with interviews from *The Tom Nelson Podcast*. Within that material, three modes dominate: Google searches, community networks, and prior familiarity.

Community networks played an observable role in dataset circulation. ACC018 described encountering a historical CO_2 analysis paper by Ernst-Georg Beck only after it was referenced by another ACC skeptic in a different Nelson podcast episode: “I visited this paper ’cause it was prompted, it was referenced by ACC020 in his talk on [your] podcast” (ACC018, Nelson Podcast, 2025). This chain, from podcast to paper to data, illustrates how datasets could circulate through the community. Interestingly, as a reminder to readers the Ernst-Georg Beck dataset, is the very dataset Engelbeen audits in his data reuse journey.

Some community members accessed climate data through third-party aggregator tools rather than institutional repositories directly. Sites such as Wood for Trees, sealevel.info, and Climate4You pull together datasets from multiple official sources into single interfaces, often with built-in visualization and analysis tools. Members routinely share links to these sites rather than the underlying repositories. These tools are not skeptic projects in any uniform sense, Wood for Trees was built by a software engineer who accepts the scientific consensus around anthropogenic climate change, and sealevel.info by Burton, and as knowledge infrastructures they are taken up more fully in Chapter 6. For instance, in response to a mainstream-aligned YouTube video posted on Reddit, an ACC skeptic commenter responded:

His graph from 2004 is not based on any of the major data sets. It looks like he just drew the graph himself. Here's the real data: https://woodfortrees.org/plot/hadcrut4gl/from:2004/to:2020/plot/gistemp/from:2004/to:2020 (COMM002, Reddit, 2021).

In another instance, COMM003 suggested sealevel.info as a "treasure trove of data from NOAA's 375 world-wide tidal gage records the averages" (COMM003, Blog, 2021). Similarly, another commenter (COMM004) presented the Climate4You site as "an excellent source for a stupendous quantity of climate data and analysis" (COMM004, Blog, 2021). What matters here is that they, not the institutional repositories, were often the community's point of access to data.

Many reusers did turn to the original data repositories, especially those with complex data reuse journeys. Three out of the four data reuse journey subjects rely heavily on established institutional repositories run by NOAA, FAO, the Global Carbon Project, and PSMSL. Engelbeen described this pragmatic reliance directly: "As far as possible from reliable sources in many cases there is no choice at all, because there is only one source: e.g., NOAA for ocean pH and pCO_2" (Engelbeen, Interview, 2025). Across the corpus, the most reused data were observational temperature records, since the central ACC claim is that global temperatures are rising and skeptics return to these records to contest it. Four recur most often:

- HadCRUT from the Met Office Hadley Centre and the Climatic Research Unit (CRU) (Brohan et al., 2006; Morice et al., 2012, 2021). HadCRUT combines a sea surface temperature (SST) dataset and a land surface air temperature dataset. Often, skeptics state that these data are less trustworthy because of CRU's involvement with Climategate.
- Global Temperature Report from The University of Alabama in Huntsville (UAH) Earth System Science Center (Christy & Spencer, 2026). Two ACC skeptical climate scientists, Dr. John R. Christy, and Dr. Roy W. Spencer, created this dataset. This global temperature dataset is derived from NOAA's TIROS-N satellite microwave data and shows fewer extreme increases in temperature than similar datasets.
- Goddard Institute for Space Studies Surface Temperature (GISTEMP) Analysis Dataset from NASA (GISTEMP Team, 2026; Lenssen et al., 2024). GISTEMP is comparable to the HadCRUT dataset in combining a land-surface air temperature dataset and an SST dataset to create a global temperature dataset, and the two are often compared.

- Global Historical Climatology Network monthly (GHCNm) Dataset from NOAA's National Center for Environmental Information (NCEI) (found in eight documents) (Menne et al., 2018). This dataset provides monthly climate summaries from stations around the world. Climate skeptics critique this dataset saying data are skewed due to urbanization and homogenization.

Beyond temperature, reuse clustered around particular arguments: CO_2 data for trends and correlations; sea-ice extent and sea-level records to argue that change is exaggerated or not accelerating; solar, ocean-oscillation, and water-vapor data to advance alternative causes; and energy and development indicators to tie quality-of-life gains to fossil fuels. For a list of all datasets found in the data corpus see Appendix C.

Formal openness, however, did not guarantee easy access, and the four journeys showed a spectrum of effort. Some reuse was a straightforward download (Burnett, Interview, 2025; Engelbeen, Interview, 2025); other cases demanded much more. McKitrick found the Challinor dataset listed in a published paper as available at a URL that no longer existed and had to email the author directly; he likened working through government data portals to "an Easter egg hunt" (McKitrick, Interview, 2025). Burton's practice was the most labor-intensive, stitching records across PSMSL and NOAA, contacting the Polish IMGW directly, and integrating fifteen tide-gauge subsets compiled by different organizations (Burton, Interview, 2025). Direct outreach recurred across the broader corpus. ACC021 and ACC022 wrote to meteorological societies throughout Europe for station-history metadata:

> In 2021 to 2022 we went, we reached out to groups in Europe, across Europe, looking at the different meteorological societies or climate groups in the different countries in Austria, in Slovakia, all over Europe and said "can you provide us with these station history metadata for as many stations as you can?" (ACC021/ACC022, Nelson Podcast, 2023).

ACC027 described a more informal version of the same process, finding a reference to historical data on a government website and then calling directly to obtain it:

> I poke through their website and in one of the websites I found a reference that said there was data going back to like 1916 or something but I couldn't find it anywhere right so I called, I literally called a guy in the government…and I said where's this data can you get it to me (ACC027, Nelson Podcast, 2025).

Whether through institutional repositories, community networks, or unofficial aggregators, data selection in this community reflected a prior interpretive stance. Reusers broadly knew what they were looking for before they began searching. The datasets they selected, and the tools they used to access them, were shaped by a set of trust criteria that determined which sources counted as reliable and which did not. How this community evaluated and assigned trustworthiness to data and repositories is examined in the following section.

5.2.1 Assessing data quality and trustworthiness

In this data community, not all data were created equal; unadjusted or "raw" data were considered superior to all other forms of knowledge because they were the least affected by human subjectivity. By "raw," skeptics referred to data directly collected from instruments with no additional processing, adjustments, homogenization, or manipulation. Most of the interviewees express skepticism towards the corrections, adjustments, models, or interpretive frameworks that institutions apply to measurements. In other words, distrust was directed primarily at the interpretive, processing, and analytical decisions made by institutions and researchers rather than at the underlying measurements themselves. The data were more likely to be trusted, while the interpretive infrastructure built on top of it were not.

The community was not uniform, however, in what data they considered trustworthy. Some community members questioned whether the initial readings themselves were reliable. One ACC skeptic researcher quoted in a blog post argues that "global mean temperatures before 1980 are based on untrustworthy data," noting that before full satellite observation began, only a small portion of the Earth had been observed with sufficient accuracy and frequency (ACC025, Blog, 2019). Others rejected the conceptual framework of global average temperature entirely. Cohler characterized "the global temperature" as "a fiction" and distinguished it from local temperature measurements, which he treated as physically meaningful measurements used in weather forecasting (Cohler, Interview, 2025). He grounded this claim in his own mathematical work on temperature aggregation published in a preprint (Cohler, 2026), as well as in prior critiques of global temperature averaging (Essex et al., 2007). He extended this critique to what counts as data at all: "A lot of things that people call data and call measurements today are not measurements. A lot of organizations get away with saying, 'Here's the data for this,' when that data is actually purely model calculations" (Cohler, Interview, 2025). In fact, over the course of

my digital ethnography, I noticed an increasing distrust in the data themselves, and their capability to measure their intended phenomenon**.** As ACC025 shows in this statement:

> Both of these reports themselves are "Data Disasters" – in that the data upon which they are based is so flawed that it is not fit for purpose. The flaw is, or should be, obvious to anyone even glancing at the data (ACC025, Blog, 2021).

Still, the distrust of institutional processing was the most consistent and widespread position across the corpus, but it was not the only one.

Overwhelmingly, ACC skeptics favored reusing observational data over other types of data. Taylor articulated the quality hierarchy directly:

> Is this objective data that is not subject to manipulation, to cherry-picking, to whatever else? If I can get the actual data, that's the best. And so for example, when I give presentations on climate change, I might cite tornadoes. And if I can go to the National Oceanic and Atmospheric Administration's webpage and get, not the summary, not what the bureaucrats are saying about climate change and tornadoes, but what are the actual data (Taylor, Interview, 2025).

He framed this as a simple principle: "The more raw or closer to the collection instrument, the better" (Taylor, Interview, 2025). Across the corpus this orientation recurs consistently. One Reddit commenter stated it flatly: "I'll believe it when we have raw data" (Reddit, 2021). Another framed it as a matter of objectivity: "The only truly objective method to evaluate climate model projections is by comparing them with real-world data" (CO_2Science, n.d.).

> ACC017, a prolific data reuser and tool-builder, described his preference for raw data: I use the raw data, the raw daily temperature data and it seems to be pretty consistent, it seems to give believable answers but they don't release graphs of that data in public. They do massive adjustments to the data before they create their graphs (ACC017, Nelson Podcast, 2022).

He also sought out sources that cannot be revised after the fact: "I think the people fudging the data hate those old newspaper reports because they can't fudge them" (ACC017, Nelson Podcast, 2022), using digitized historical newspaper archives to document extreme weather events that predate the modern instrumental record. That move points to a second sense of "raw" at work here: not only data close to the instrument, but data that can no longer be

altered. An old newspaper was trusted less for its precision than for its fixity, even as present-day reporting was deemed unreliable; the value lied in being beyond the reach of later adjustment. The same contrast organized what data were selected and trusted, as when one reuser set a "relatively unmanipulated" NASA satellite dataset against the Copernicus "black box mélange of surface temperature data" (ACC012, Blog, 2025).

Climate science, as a discipline, relies on Global Climate Models (GCMs) rather than observational data, which ACC skeptics preferred. Instead, ACC skeptics often rejected these models outright, considering them inferior to observational data. Climate models were portrayed as reflecting the assumptions of their makers rather than the behavior of the climate system. "Stop with misleading models, let the data speak" (ACC023, Blog, 2020). Observational data were thus often used to critique proxy data and climate model projections. ACC024 in a conference presentation cites the Bulletin of the American Meteorological Society noting that models may be optimized for metrics tied to "goals, expertise or cultural identity" of a modeling center, interpreting this as proof that "it is the scientist, not the science that determines what is 'anticipated' and 'acceptable'" (ACC024, YouTube, 2021). Burton articulated the underlying principle directly: "Measurements are better than models. Models can be very good, but only if the thing being modeled is well understood" (Burton, Interview, 2025).

5.2.1.1 How Reusers Develop Trust and Distrust in Data

Community members developed trust and distrust in data through several criteria - documentation transparency, consistency across independent datasets, stability across versions over time, the directness of the underlying measurement, the processing a dataset has undergone, and the perceived neutrality of the institution behind it - that recur across the corpus, though they applied these criteria with different levels of rigor and formality. Within the semi-structured interviews, Engelbeen articulated his criteria most explicitly:

> If there were different datasets, looking at the differences and why there were differences and how the differences were explained (if at all). Further the reputation of the data provider for integrity of data and eventual corrections of the data in the past and external critiques on these corrections from third parties (Engelbeen, Interview, 2025).

McKitrick described his criteria more functionally:

> I don't assess data sets for fraud or fabrication as I don't have any training in that. I look for availability, traceability to original sources, and relevance. For instance in climate

> studies I would look for whether the data is being used for its intended application (McKitrick, Interview, 2025).

Burnett's trust was grounded in prior experience rather than formal evaluation. The spectrum here, from Engelbeen's explicit multi-criteria evaluation to Burnett's familiarity, maps broadly onto their institutional positioning and the demands of their respective publication formats. Engelbeen and McKitrick published detailed analyses built to withstand scrutiny, so their evaluation was explicit and documented; Burnett, whose output was a blog post, leaned instead on sources he had already found reliable, since the format rewarded turnaround over fresh appraisal.

Examining data documentation was a common first step in evaluation. Community members read READMEs, methodology papers, and FAQ pages to understand how datasets have been constructed and processed. This can itself become a friction point. Cohler described encountering two different URLs both claiming to be USHCN version 2.5, each with its own README, and being unable to determine from the documentation whether they contained the same data or different data. The directories had different structures and different file dates, with one stamped 8-25 and the other 8-28, and one appeared to contain what he described as possibly raw or pre-processed data. He could not determine from the documentation alone whether these were the same dataset, different versions, or something else entirely:

> I started going through the READMEs…and so finally, I just threw my hands up in there and I said, Look, I'll just ask ChatGPT. ChatGPT, what's the story on this? They both claim that they're USHCN 2.5, why are they different if they're the same? (Cohler, Interview, 2025).

The resort to AI to interpret documentation is itself an observation about the state of data documentation and proliferating datasets online.

One community member examining the GHCN dataset methodology described reading through the FAQ documentation and finding the explanation for why the century-scale trend was higher in a newer version: the algorithm adjusts earlier temperatures to conform to newer measurement conditions, detecting approximately twice as many change points in the newer version than in the older one. Rather than reading this as a methodological improvement, he framed it as evidence that "the data you get is different from the raw data version to version" (ACC028, blog, 2021).

Consistency across independent datasets was a recurring criterion for evaluating whether data can be trusted.

> Having multiple sources confirming that is a good thing. So when there are ice core data from Greenland and from Antarctica that tell us the same thing about the earth's climate history, that's good because they're two objective sources that agree (Taylor, Interview, 2025).

Engelbeen noted that the Scripps and NOAA CO_2 datasets "don't differ with more than 0.2 ppmv, which proves that the CO_2 measurements worldwide are very accurate" (Engelbeen, Interview, 2025). Burton described a general habit of cross-checking: "I have a nasty, suspicious mind. I have a habit of cross-checking things. If I hear two different statistics, which are related to one another, I like to check them for consistency and see if they are compatible" (Burton, Interview, 2025). He also flagged suspicious outliers directly in the data: during an interview he pointed to several months of tide gauge data at Seward, Alaska and noted "there are several months there that cannot be right," identifying the 1964 earthquake as a likely cause of the anomalous readings (Burton, Interview, 2025).

Comparing different versions of the same dataset over time was another trust assessment practice. Some community members did this systematically. Tom Nelson described another community member's practice of downloading the full raw dataset on consecutive days to check for unexplained changes:

> He would download it one day, download the whole dataset the next day, and data from the 1930s would be changing and there's no explanation as to why. So yeah, that was a red flag there, that there's data being manipulated after the fact, and we don't know why and there's not an explanation of why (Nelson, Interview, 2025).

Engelbeen connected this practice directly to questions of data legitimacy:

> If for some reason the past data are not reliable, that should be clearly explained and motivated. There is a clear lack of such explanations in near all temperature datasets, which make the past temperature data from direct measurements highly questionable (Engelbeen, Interview, 2025).

Climate4You made the same test a standing feature. One of its analyses, which the site calls a test of temporal stability, plotted the net change in each of five global temperature datasets between their May 2008 and October 2021 versions, and concludes that "a temperature

record which keeps on changing the past hardly can qualify as being correct" (Climate4You, n.d.). As elsewhere in this section, a record that was revised over time was treated as unreliable on that basis.

The type of measurement mattered as well as its source. Engelbeen questioned data trustworthiness "when there are huge differences between similar datasets" (Engelbeen, Interview, 2025). His skepticism extended to proxy measurements:

> Especially several temperature proxy measurements, like tree rings, are completely unreliable: for the same place, one series can show 'colder', another 'no change' or 'warmer' temperatures… Depending of the choice of the researcher (with his/hers own biases), you can have warmer of colder temperatures or no temperature change at all (Engelbeen, Interview, 2025).

His preference for measurements that physically captured the phenomenon of interest over those that reconstruct it through an intermediary was illustrated by his comparison of ice cores and stomata data [ice cores preserve bubbles of ancient air whose CO_2 can be measured directly, whereas stomata data infer past CO_2 from the density of pores on fossilized leaves]:

> The stomata data are calibrated against ice core CO_2 over the past century, but nobody knows how the local bias evolved over the centuries in the past, due to changes in local vegetation or vegetation in the main wind direction (Engelbeen, Interview, 2025).

For temperature datasets he preferred:

> UAH satellite measurements, as these are certainly least biased to any direction and are frequently calibrated against balloon data at several places. Other satellite datasets are often calibrated, based on 'models', which are more prone to biases of the model makers (Engelbeen, Interview, 2025).

As a reminder, UAH's dataset was created by ACC skeptics and was highly trusted by this community for satellite temperature data. Still that trust is not unanimous. ACC016 distrusted satellite records across the board, UAH included, preferring temperatures measured at ground stations "where people live" over readings inferred from orbit (ACC016, Nelson Podcast, 2025).

Trust assessment extended to the conditions under which data were collected. The volunteer project surfacestations.org, started by Anthony Watts, organized volunteers to survey

temperature stations to document "biases and errors through faulty siting, encroachments, or maintenance issues" (*Surface Stations*, 2021).

Burton's engagement with satellite altimetry data illustrated how these criteria play out in practice. He described discovering that the AVISO dataset includes a 0.3 millimeters per year adjustment added to account for ongoing ocean floor sinking, an estimate derived from models rather than direct measurement: "There's no way to measure that, but they've got untestable models where they make guesses at it" (Burton, Interview, 2025). He contacted AVISO to request that the option to view data without the adjustment be restored and received what he describes as "a very polite refusal" (Burton, Interview, 2025). The adjustment, in his view, meant the dataset no longer reported sea level but rather "what they think the sea level trend would be if it weren't for this factor" (Burton, Interview, 2025).

Consequently, when acquiring data, some ACC skeptics closely examined data repositories, documentation, curatorial decisions, and scholarly products to find processing that has, in their view, contaminated the raw data. This processing could include adjustments applied to raw measurements, such as homogenization, corrections for station changes, instrument and temporal changes, and the gridding and infilling of gaps. In a Reddit post, one community member critiqued the GHCNm dataset by directly quoting its documentation:

> And since in the introduction to the data (CDRP-ATBD-0859 Rev 1 GHCN-M Mean Temperature-v4.pdf) it states:
>
> > "Data are collected from NOAA in situ networks as well as other national and international providers. They are subjected to a series of processes that combine data from various sources, perform quality control, homogeneity corrections, and output the data for customer access and permanent archive."
>
> And their readme.txt from their ftp site says in the Introduction:
>
> > ". . . The greatest difference from previous version is a greatly expanded set of stations based on the large data holdings in GHCN-Daily as well as data collected as part of the International Surface Temperature Initiative databank (ISTI; Rennie et al. 2014)."
>
> So the data you get is different from the raw data version to version. Considering these data have been adjusted in the combined sources with obviously large changes up to Version #4 (See Figure 3 in the above mentioned pdf) how would I know they haven't

been “adjusting” their data to better align with everyone else? (COMM006, Reddit, 2020).

Some made this comparison systematically. ACC017 described his method plainly: “I just take the raw dataset and the final adjusted dataset and for each year do an average and calculate the difference to see what the amount of the adjustments are” (ACC017, Nelson Podcast, 2025). To do it at scale he built his own software tool, which ingests records from many stations at once. The point was to measure how far the published record sits from the raw measurements, and to read the size of that gap as the amount of alteration the data have undergone.

ACC021 and ACC022 took this a step further, gathering their own evidence to check an institution’s processing. They obtained station history metadata from over 800 weather stations across more than 24 European countries by reaching out directly to national meteorological societies:

> Each of these groups had looked in the documentation for what are called station history metadata. So that’s where the station observers would say, we had to move our weather station. We’ve replaced it with a new one. So they have recorded a lot of these major changes that took place at the different times of the record” (ACC021/ACC022, Nelson Podcast, 2024).

This metadata was then used to evaluate NOAA’s homogenization adjustments against physical records of actual station changes.

ACC skeptics also noted when different versions of datasets would be removed or no longer supported actively by repositories and accuse repository staff of changing or removing data to support the theory of ACC. If available, they compared different versions of datasets to note changes using the Wayback Machine.

This pattern is significant for two reasons. First, it explains why the community continued to use mainstream climate science data even while contesting mainstream science conclusions. The raw measurements were more likely to be treated as neutral while the interpretive layer was treated as politically or epistemologically compromised. Second, it shapes what kinds of reuse were possible and valued in this community. Reuse that went back to the original measurements and applied analysis was treated as more credible than reuse that relied on processed or modeled outputs.

ACC skeptics often distrusted repositories because they associated them with political institutions or perceived ideological agendas. Groupthink was seen as a major threat, with the scientific consensus on anthropogenic climate change portrayed as politically driven. As COMM010 stated the following:

> Science is antithetical to science collectivism, they don't go together. Today's climate science collectivism is a political order bridging a religious order. These zealots embrace group think, they extol the virtues of consensus; collectivism, it's their power to truth mechanism. Science, however, is truth to power, not power to truth (COMM010, Blog, 2025).

Skeptics claimed dissenting views are excluded from journals, pushing them to publish preprints that they described as receiving "defensive and antagonistic" reviews (COMM007, Reddit, 2021). This perceived institutional capture extended to data repositories. The concern was not just that models were wrong but that the institutions maintaining the underlying data had been captured by the same forces driving the models. One Reddit commenter stated this directly: "I don't trust anything released by NOAA either. And for good reason" (COMM008, Reddit, 2021). Another asked: "Do you think they have the satellite and consistently report the measurements, or do you think like me that they select the data that supports their narrative, from wherever they can find it, regardless of its credibility?" (COMM009, Reddit, 2021). One principle holds the trust framework together: data were trusted in inverse proportion to the amount of institutional processing they carried. Rawness operated as a proxy for neutrality, and the processing layer, adjustment, homogenization, modeling, was where the community located most of the risk of ideological contamination. Evaluating data here was, in large part, a matter of tracing how far a given figure sat from the instrument that produced it.

5.3 Processing and transforming data

Processing and transforming data includes activities through which reusers prepare data for analysis. This encompasses preprocessing, cleaning, normalizing, aggregating, and converting data into formats suitable for further work. A spectrum of processing appeared across the four journeys, from no processing at all to interoperable technical automated systems. This range reflected not only differences in technical capacity but differences in publication format, research goals, and epistemological commitments about what it means to engage directly with data.

In the content analysis and participant observation, I found limited descriptions of data processing by ACC skeptics for what I believe are several compounding reasons. First, many data reusers did not publicly publish or write about their data processing activities, meaning the evidence in this section relies primarily on the eight semi-structured interviews and Nelson podcasts rather than content analysis or participant observation. Second, as shown in the previous section, skeptics often disparaged data processing that researchers or repositories carried out on the data, valorizing "raw" data. When ACC skeptics' data processing activities were made public, they often were not extensive, perhaps because of these beliefs. A third factor is that many community members accessed data through interactive tools like Wood for Trees or FAOSTAT where processing is built into the interface and no external processing is required.

Burnett illustrated both patterns: all processing occurred within FAOSTAT's interactive interface, no data were downloaded at the processing stage, and percentage calculations were performed manually from cursor-hover values within the visualization tool. This was not a failure of technical capacity but a reflection of his publication format and goals, which were rapid reactive blog posts that did not require the overhead of full data extraction and reprocessing.

Engelbeen conducted relatively extensive processing within Excel. For his analysis of the seasonal CO_2 cycle, another section of his CO_2 Coalition report which was not discussed in Chapter 4, he built a custom Excel macro: "I had to make an Excel macro program to convert the long series of monthly CO_2 measurements of different stations since 1958 into the year by year evolution of seasonal CO_2 amplitude over the years" (Engelbeen, Interview, 2025). He also applied the Takahashi (2002) formula to calculate temperature's short-term influence on ocean surface pCO_2, and scaled HadSST3 temperature data by a factor of 3.5 to equalize amplitudes with CO_2 rate of change for visualization purposes. These processing decisions required domain expertise and methodological judgment even within a familiar tool.

McKitrick's two-month data recovery process, filling 360 missing records by cross-referencing underlying papers and IPCC scenario tables, was itself a form of processing. The incompleteness he discovered upon opening the spreadsheet was not resolved by going back to the repository but by going back to the original literature the meta-analysis dataset was based on.

Burton's processing was the most technically elaborate of the four, running from weighted averages in Excel to automated Perl scripts that pull data across repositories through

their APIs. One choice is worth noting: he deliberately represented missing months as null values rather than dropping them, so Google Charts would not create a misleading graph.

These four journeys establish a spectrum that extended across the broader corpus. One common processing activity was file format conversion to make data more accessible for spreadsheet analysis. Data repositories often provide various file formats, including CSV and TXT files, making the need to convert file formats unnecessary. For example, in a CO_2Science blog post, the author used the HadCRUT CSV version of the dataset over the netCDF version, along with the TXT file of CO_2 data from NASA. Engelbeen described CSV as his preferred format precisely because it imports directly into Excel and converts easily to other formats (Engelbeen, Interview, 2025). When sharing data with others, some reusers went further: one community member, when examining albedo data, used the NetCDF file provided by NASA Earth Data for her own work but converted it to a plain text file when sharing with others in comments, perhaps for ease of access by less technical users (ACC030, Blog, 2021).

In one of the few larger examples of data processing described in the broader data corpus, ACC036 took CERES radiation data in its native NetCDF form and converted each of the 32 datasets into a three-dimensional array of latitude, longitude, and time, storing them in a form he could reload without redoing the conversion:

> It's in NetCDF format, which has the extension .nc, so you'll need software able to read NetCDF files… I went through and converted each of the NCDF datasets into a 3D array (latitude, longitude, time). Then I saved them in the R native format. This lets me access them without having to extract and convert them every time I need them. This gives me a folder with about 1.45 gigabytes of data (ACC036, Blog, 2021).
>
> The processing made the rest of his reuse more efficient.

ACC030's posts documented processing in more detail than most in the corpus, because she published the scripts that produced her results. For instance, in one blog post, she pulled two monthly global series out of a CERES top-of-atmosphere file — outgoing shortwave radiation (the variable gtoa_sw_all_mon, the sunlight Earth reflects back to space) and incoming solar radiation (gsolar_mon) — and derived albedo by dividing the first by the second, the step her code wrote as $1/$2, before passing the result to a linear regression with gmt gmtregress (ACC030, Blog, 2021). Albedo is not a field in the file; it is the quantity that division produces.

Her global fire index took more processing. NASA's Active Fires data arrive as a grid of cells, each holding a fire-intensity value, and she reduced the whole grid to one global number per period to plot how it moved over time. Two of her choices determined what that number meant. First, she weighted each cell by its surface area, on the reasoning that a one-degree cell covers less ground near the poles than at the equator and that averaging the cells equally would, in her account, let the small polar cells count for as much as the larger cells at the equator. Second, she averaged across the whole globe rather than over land alone, including ocean cells she treated as carrying no fire; this, she said, was why the index could not exceed roughly 0.29, since she counted land as about 29 percent of the surface. She described the values in those terms: "Fire pixels have intensity. 0.05 could mean 5% of Earth has strong fire, or 10% has medium fire. It also can not exceed 0.29" (ACC030, Blog, 2021).

A pattern ran across these processing practices that follows from the epistemological commitments in 5.2: distrusting the corrections institutions apply, many skeptic reusers preferred exclusion to correction. Rather than adjusting data they considered unreliable, they removed it. ACC017 stripped away all adjusted data and worked only with raw records. In another example, when ACC030 was working with the NASA global fire data index that had occasional missing values, she chose not to interpolate them: "Even one missing day means I have to fill in data. I didn't want to do that" (ACC030, Blog, 2021). She instead fit a linear trend to the record as it stood rather than impute missing values. ACC021 and ACC022, the team behind the European station-history reconstruction described in 5.2, were a partial exception. They excluded urban stations and built rural-only subsets, but where they did adjust, they justified each change against the physical station-history record rather than by algorithm: "We show you exactly how much we adjust and we try to justify according to the history record that we have. We have to do it on individual basis, we cannot use a computer program" (ACC021/ACC022, Nelson Podcast, 2024). They took this hand-checked, case-by-case method to be what distinguished their approach from NOAA's.

5.4 Analyzing, interpreting, and abstracting data

Analyzing, interpreting, and abstracting is the stage at which reusers turn data into results and claims, through activities like comparing, searching for patterns, reproducing prior analyses, and visualizing. Across my data corpus, analysis was organized less commonly around generating new knowledge than around checking whether official data and claims held up.

Reusers compared datasets against one another, compared observations against models and proxy reconstructions, searched observational records for trends and relationships, and reproduced prior analyses to determine whether published findings could be replicated. Although the techniques ranged from simple visual inspection of datasets to statistical modeling, they were united under this common orientation: analysis was conducted as a form of audit. Data were used to evaluate whether institutions, researchers, and public communicators had accurately represented the phenomena they describe.

The analytical activities observed across the corpus fell into three recurring categories: comparing datasets and claims, identifying trends and relationships, and reproducing prior analyses. These activities often overlapped within a single reuse episode, but distinguishing them analytically helps clarify how ACC skeptics transformed data into evidence.

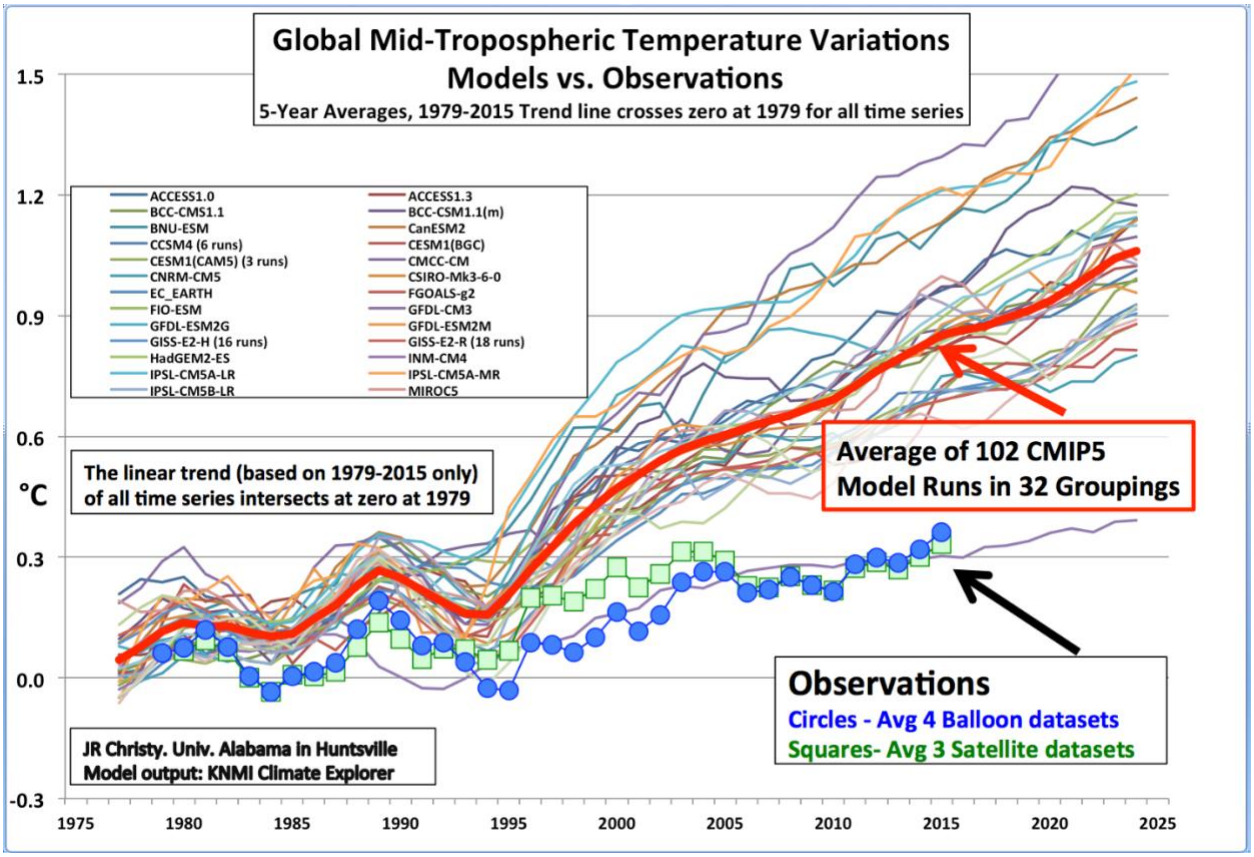
5.4.1 Comparing datasets, models, and claims

Comparing was the community's most common analytic activity, and the one in which the audit orientation was most visible. Reusers frequently placed datasets alongside one another, compared observational records against model outputs or proxy reconstructions, or evaluated empirical records against public claims. The comparison was typically structured around a prior question: does the evidence support the official account?

One common form of comparison placed observational data against model projections or proxy reconstructions. Observational measurements were treated as a benchmark against which inferred or modelled values were evaluated. One community member described observational records as "ground truth data," distinguishing direct observation from inference-based forms of evidence (ACC041, Website, 2021).

Below is an example produced by Dr. John R. Christy, who used the KNMI Climate Explorer to plot climate-model output against observational temperature records. He charted mid-tropospheric temperature from 32 models, representing 102 individual simulations, as five-year averages, and overlaid two observational series, one from weather balloons and one from satellites (Testimony of John R. Christy, 2017, p. 5). He reported that the models warm faster than the observations, by his account about two and a half times faster on average and roughly three times faster in the tropics, and that a single model, the Russian INM-CM4, was the only one that tracked the observations closely. The observational record was treated as the benchmark and the model ensemble as what was tested against it.

Image 5.1: Dr. John Christy's comparison of modeled and observed mid-tropospheric temperature. Lines show five-year-average variations from 32 climate models (102 simulations); circles show weather-balloon observations and squares show satellite observations (*Testimony of John R. Christy*, 2017, p. 5).



A second form of comparison examined datasets that measure the same phenomenon but are produced by different organizations. These comparisons often surfaced disagreement among the datasets rather than convergence. An example of this comes from ACC041, where he compared five major global temperature datasets, made up of two satellite series (UAH and RSS) and three surface-based series (HadCRUT, NCDC, and GISS). He classified them into his own "quality classes," placing the two satellite records in the highest class, HadCRUT in the second, and NCDC and GISS in the third, on the premise that the records were not of equal scientific quality (ACC041, Website, 2021). He showed the disagreement by superimposing the series on

common axes, normalized to a shared 1979–2008 average, since their baselines differ, and by plotting the difference between the surface and satellite averages, which he reported drifting upward since about 2003, with the surface records showing roughly 0.1°C more warming than the satellite records since 1979 (ACC041, Website, 2021). Similar comparisons appeared throughout the corpus as reusers evaluated competing temperature, sea-level, and atmospheric datasets.

Comparison was also directed toward public claims. Burnett's data reuse journey illustrates this pattern particularly clearly. News reports, policy documents, scientific papers, and social media posts were translated into empirical questions and then assessed against observational data. In these cases, the comparison was not between datasets alone but between a public claim and the evidence used to support or challenge it.

A clear example from the broader corpus is ACC017's response to an Associated Press article reporting that climate change had made wildfires worse, which cited the National Interagency Fire Center showing average annual US burn acreage rising from 2.7 million acres in 1983–1992 to 7.5 million in 2011–2020. ACC017 reframed the claim as a question about the longer record and set the same kind of data against it, citing US Forest Service and historical newspaper figures that put burn acreage far higher earlier in the century, about 22 million acres in 1937 and over 50 million in 1930–1931, alongside a federal estimate of roughly 145 million acres burned annually in the pre-industrial period (ACC017, YouTube, 2021). He said the article "cherry-picked the year 1983 to start his comparison," a year near the bottom of the record, so growth measured from there did not appear in the longer series. The comparison was between a public claim and the evidence organized to contest it.

5.4.2 Identifying trends and relationships

A second analytical activity was the search for trends and patterns in observational records. Reusers examined temperature, CO_2, sea ice, sea level, energy, and natural-disaster datasets to determine whether the patterns visible in the data supported prevailing climate claims. Unlike comparison, which evaluates one dataset, model, or claim against another, trend analysis focuses on the behavior of a dataset through time and the relationships between variables.

The specific conclusions reusers reached varied widely. Some reusers reported cooling, expanding sea ice, declining disaster counts, or the absence of statistically significant trends. Others accepted that trends exist while disputing their magnitude, significance, or attribution to

anthropogenic CO_2 emissions. Still others argued that changes in measurement practices, station coverage, or data processing make long-term trends difficult or impossible to determine. What united these analyses was not a shared conclusion but a shared purpose: determining whether the observational record supported what institutions, researchers, and media organizations claimed it showed.

Trend analysis often took relatively simple forms. Burton's browser-based regression is one example, offering users the capability to fit trend lines to tide-gauge records to calculate rates of sea-level change and evaluate claims about acceleration. Similar analyses recurred throughout the corpus, where linear regressions and visual trend estimates were applied to long-term environmental records. The emphasis was frequently on identifying what the data appeared to show directly rather than relying on published interpretations. In another example, ACC040 plotted October temperature data for Tokyo from 1987 onward and concluded that there had been no warming in Tokyo in October since 1987 (ACC040, Blog, 2021).

Relationships between datasets were also examined through correlation and temporal alignment. Reusers frequently compared temperature records against solar activity, ocean oscillations, atmospheric CO_2 concentrations, cloud cover, or other variables they regarded as plausible drivers of climate variability. The objective was often to identify alternative explanations for observed changes or to challenge the strength of relationships emphasized within mainstream climate science. Such analyses varied considerably in sophistication, ranging from visual inspection of overlapping graphs to formal statistical procedures.

An example from the broader data corpus is shown when ACC030 downloaded CERES radiation data, derived albedo as the ratio of reflected to incoming solar radiation, and fit a regression across 2000–2021. The post reported that albedo had fallen over the period and converted that decline into a radiative forcing, a measure of how much something pushes the Earth's energy balance out of equilibrium, concluding that it had warmed the planet about twice as much as CO_2 over the same years (ACC030, Blog, 2021). Here the analytical work moved beyond simple trend identification to constructing relationships between multiple variables and interpreting their broader implications. Whether produced through statistical modeling or visual inspection, these analyses were generally presented as direct engagements with observational evidence. The underlying premise was that patterns contained within the data should be visible and understandable without extensive theoretical interpretation.

5.4.3 Reproducing and reanalyzing prior studies

A third analytical activity involved reproducing existing analyses and reexamining published results. While less common than comparison or trend analysis, replication occupied a prominent place within the community's self-understanding because it directly embodies the audit orientation. Rather than asking a new question of a dataset, the reuser asked whether a previously published result could be reproduced by a nominally independent third party.

McKitrick's data reuse journey provides the clearest example. His investigation began with the observation that two studies using the same underlying dataset produced different conclusions. The resulting project involved reconstructing the analyses, recovering missing data, tracing methodological decisions, and systematically reproducing the original calculations. The objective was to determine whether the published findings followed from the data and methods employed.

Similar practices appeared elsewhere in the corpus. In ACC034's carbon-cycle analysis, the author reconstructed the carbon-cycle framework used by the IPCC and then altered a key assumption concerning natural CO_2 contributions after 1750. The resulting model generated substantially different conclusions regarding the relative contribution of anthropogenic emissions to rising atmospheric CO_2 concentrations (ACC034, Preprint, 2021). Although the analysis introduced a modified interpretation, the project began with reproduction: rebuilding the original framework in sufficient detail to evaluate how its assumptions shaped the outcome.

Replication was particularly valued because it aligned closely with the community's broader understanding of scientific accountability. Published findings were treated not as settled conclusions but as claims that remained open to independent verification. Reproducing an analysis allowed reusers to inspect methodological choices, identify assumptions, and determine whether alternative decisions produced different results. In this sense, replication served as a mechanism of methodological scrutiny.

5.5 Managing, documenting, and preserving data

Managing, documenting, and preserving data covers how reusers organize, label, generate metadata, and store data. Because community members rarely write publicly about how they store and document their own data, this section draws primarily on the interviews and Nelson podcast transcripts rather than on content analysis or participant observation. The exception is organizational and website-based practice where data are stored publicly.

The four journeys revealed a spectrum of preservation practice that ran from no durable preservation at all to automated version control, and as with the processing spectrum in 5.3 it tracked publication format and goals rather than technical skill. At one end, Burnett preserved nothing beyond the downloaded image that appeared in a post: rapid, reactive blogging needs no durable data trail, so none was kept, and the visualization was the only artifact that traveled forward. Engelbeen's practice was more deliberate but still informal and personal, with the data and graphs stored and backed up on computers, hard drives, and memory sticks.

McKitrick occupied a more structured position. His versioned file naming convention served as a preservation strategy, moving from MasterData_021013_IsFinal.xlsx through successive versions to a dated final file, with an organized local folder hierarchy, a README documentation file, a published supplement recording his data-recovery decisions, and a formal long-term deposit in Mendeley Data.

Burton's preservation infrastructure was the most developed of the four journeys, using TLIB, his own version control software, to automatically track modifications and update last modified dates, automated scripts to handle ongoing PSMSL and NOAA updates, and a multi-format strategy serving JSON, CSV, Excel, and PNG outputs that preserve different levels of contextual information. The JSON outputs are self-documenting, embedding provenance metadata directly. What separated them was not how capable each reuser was but how much ongoing life their data needed: a one-off blog chart needs no archive, while a continuously updated public tool needs version control to function.

This spectrum was not unique to the four case data reuse journeys. Many skeptic organizations utilized data to populate their websites and databases, providing convenient access to data in one location. When community members and organizations stored or republished data, they tended to favor application-neutral formats such as CSV files. For example, Wood for Trees provides "raw" data in tabular plain-text files accompanied by metadata. The metadata includes information on who processed the data, the original source, a link to the original data, the file name, and a brief description. Other websites and groups, such as Climate4You and Going Green Canada, relied on data repositories only, providing links to publicly available datasets without creating their own databases. They often chose to provide links to the documentation of these datasets to promote further understanding of the data. However, relying on these data repositories for data preservation means that any changes made by the official organization can

affect the availability of data. The EPA Climate Change Indicators page, removed in December 2025, is a vivid example: Burnett had reused a graph from that page across dozens of articles, and it is now unavailable at its original URL, though copies persist across blog posts and social media, stripped of their original context.

Beyond website-based preservation, some community members maintained their own personal archives and discussed them publicly in interviews. ACC032, a retired engineer collaborating with the ACC021 and ACC022 team, began downloading the full GHCN dataset from NOAA's website around 2010 or 2011 after noticing that adjustments to station records were changing without explanation. ACC021 discussed what ACC032 found:

> He went back to check the adjustments that had been done to a weather station and he realized that the adjustments had changed…he realized that every day NOAA recompiles the adjustments for their station records and every day those adjustments are different" (ACC021/ACC022, Nelson Podcast, 2023).

ACC032 escalated from monthly to weekly to daily downloads, and had by then preserved nearly 2,000 to 3,000 versions of the GHCN dataset, including versions no longer available on government servers. This practice sat at an extreme end of the preservation spectrum, motivated by a specific concern about institutional data integrity. Keeping data on personal infrastructure rather than relying on institutional repositories was consistent with the community's broader skepticism toward institutional stewardship.

ACC017 described a similar if less intensive personal archiving practice: he found Twitter useful because "they had archives you could download so I could put information on Twitter and then digitally search later and find it — it's a good way for me to archive information" (ACC017, Nelson Podcast, 2022). For ACC017, social media functioned as a searchable personal archive for the historical newspaper records and data he encountered in his research.

5.6 Publishing and sharing research

Publishing and sharing of research refers to the dissemination of findings derived from data reuse. This community has a hybrid publication network that includes mainstream peer-reviewed journals, ACC skeptic-run peer-reviewed journals, grey literature, blogs, and social media. The data reuse journeys themselves contain a wide range of publication venues: a peer-reviewed research journal (McKitrick), a think-tank affiliated blog (Burnett), a grey literature

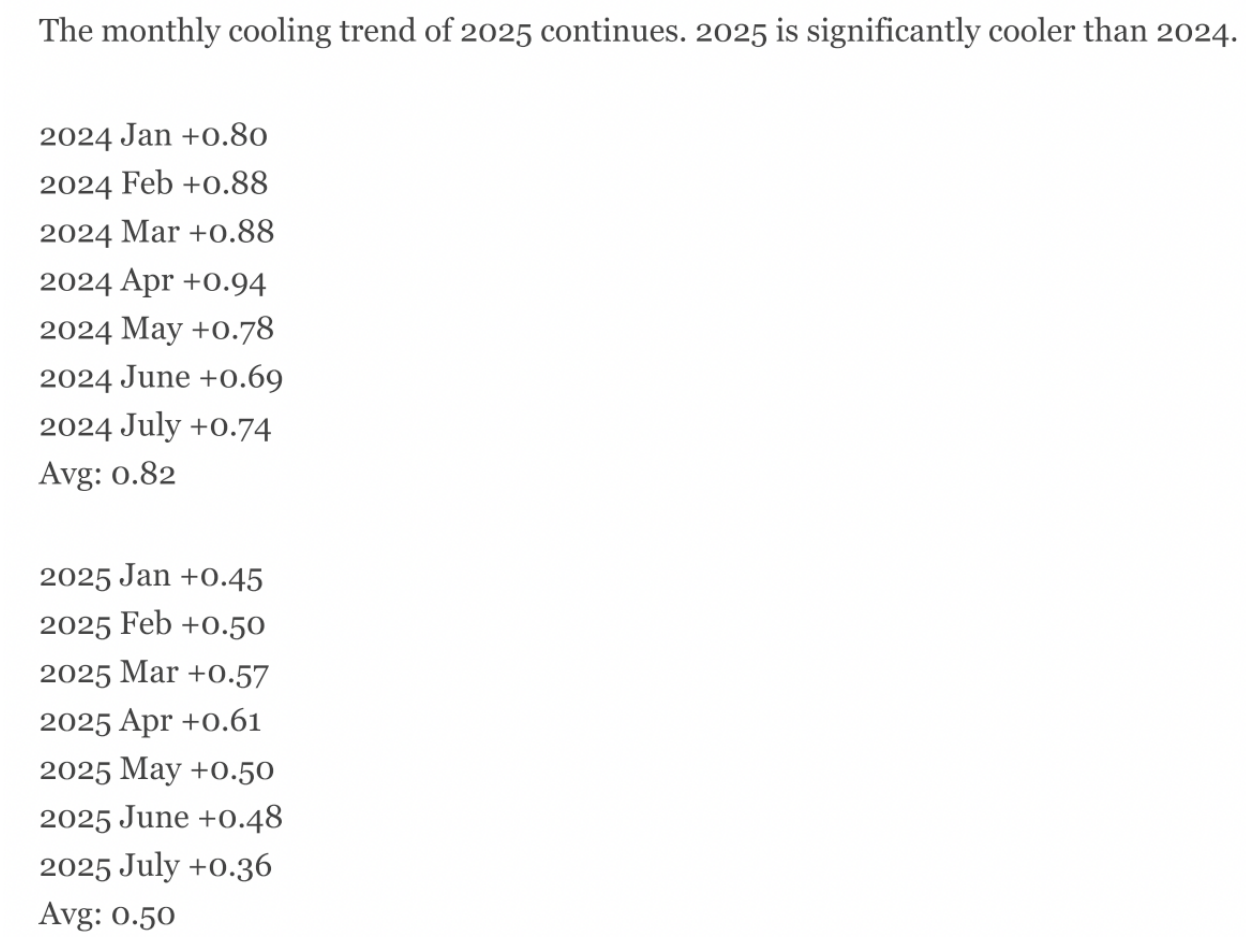
report for an ACC skeptic nonprofit organization with an OSF preprint deposit (Engelbeen), and a self-maintained tool (Burton). In only one case (McKitrick) did the work pass through formal peer review. The others relied on alternative legitimation strategies: institutional affiliation, government-style formatting, OSF deposit, or the longevity and transparency of the tool itself.

This spectrum maps closely onto the broader community's publication practices. Some community members published in peer-reviewed journals, ranging from mainstream outlets to community-run journals and venues with more permeable gatekeeping, which is examined in Chapter 6. ACC skeptics predominantly shared through blogs, forums, and social media networks, informal infrastructures that functioned as validation systems and interpretive communities.

Frequently, blog posts with data reuse cases had extremely long comment sections, reaching up to hundreds of pages. These comment sections functioned as a form of informal peer review. When errors were pointed out that the data reuser agreed with, the work was frequently changed. For instance, one skeptic responded to an error in his figure by stating, "Thank you very much for finding my labeling error in Figure 10 that inadvertently switched the labels for land and deep ocean, ending up disagreeing with my own calculations. I updated the PDF above to correct this error" (ACC034, Blog, 2021). Commenters provided constructive critique, and reusers revised their work accordingly. ACC035 described an even more intentional version of this process through his use of Substack: "I can put ideas out there, I can put data out there…a lot of the stuff that winds up in the peer-reviewed literature that I contribute to…you can chase them back to some appearance on my blog or Substack" (ACC035, Nelson Podcast, 2023). He framed this explicitly as "extended peer review" (ACC035, Nelson Podcast, 2023), treating informal publication as the first stage in a research process that may eventually lead to formal peer-reviewed output.

A claim and a data reuse response sometimes occupied the same digital space: on blog comment threads, Reddit, and social media, a claim circulated and a reuser answered it with data in the same thread. In one such thread responding to a 2025 blog post about the DOE Climate Assessment report, COMM001 posted University of Alabama, Huntsville (UAH) satellite temperature data directly into the comments to contest warming claims under discussion (Image 5.2).

Image 5.2. COMM001 posted UAH satellite temperature anomaly data directly into a blog comment thread in response to claims about warming trends (COMM001, Blog, 2025).



5.7 Attributing and citing data

Attributing and citing data refers to how reusers acknowledge the source of the data they use. Citation practice in this community tracked institutional positioning. Reusers working within academic publishing norms used formal citation, with full references, DOIs, and persistent identifiers, because their venues and audiences expected it. Those working through advocacy and blog channels attributed more informally, using inline links and mentions that pointed readers to a source rather than supplying the apparatus of formal citation. The informal forms were also less durable: without persistent identifiers, a linked source can move or disappear, as several dead links across the corpus show.

All four data reuse journeys acknowledged their data sources, but the form ranged from formal APA-style citations with DOIs, to informal inline links, to the "(ongoing)" citation convention that foregrounds continuous updates over fixed publication dates. None of the four used persistent identifiers consistently. The tension between static citation conventions and continuously updated datasets was visible across all four journeys.

McKitrick, operating within academic publishing norms, treated data citation as a formal obligation: full citations in the Nature paper, a Mendeley Data deposit cited with a DOI, and a supplement that documents every data recovery decision with explicit source references, functioning as an extended data citation record. Personal communications with Challinor, Hertel, and Moore are cited in the Methods section. Burton's attribution was consistent if not always uniform in format: two full citations for PSMSL, links to NOAA CO-OPS, station-specific citations for individual tide gauges, and attribution metadata embedded at the top of archived text files. The Świnoujście station page cites IMGW, PSMSL, and NOAA separately as three distinct sources, reflecting the complexity of the data provenance. Engelbeen's citing practices were less consistent: some sources were fully cited, others linked only, and no formal data citation accompanied the graphs made available on request. Wood for Trees was used for plotting, but the underlying datasets were cited separately rather than through the platform. Burnett's attribution was the most informal of the four: he linked to the FAOSTAT Compare interface rather than to the underlying dataset or its metadata, the EPA Heat Wave Index graph includes attribution text when embedded but that link is now dead, and the journal article that prompted the reuse episode was mentioned but not formally cited in the blog post.

This spectrum, from McKitrick's formal DOI citations to Burnett's dead links, maps onto the broader pattern in the community. Reusers publishing in academic venues were more likely to have full citations for the data. For instance, a journal article published in the Journal of Applied Meteorology and Climatology challenging the urban heat island adjustments in the Global Historical Climatology Network cites the relevant data papers in the reference list, complete with DOIs (Spencer et al., 2025). The formality of the citation aligns with the setting and the purpose of the article. The venue requires it, the authors write as academics, and a critique of GHCN adjustments must name precisely the datasets and methods it disputes.

At the more informal end, bloggers named the producing institution datasets but pointed to them more often through inline links rather than full citations. ACC040 placed a source line

beneath a self-made graph "Data: JMA," which links to the agency's monthly mean temperature spreadsheet, and credited the source in the text: "according to data from the Japan Meteorological Agency (JMA)" (ACC040, Blog, 2021). Another blogger, ACC030, named CERES as the source of albedo data and routed readers to it, "You can download here(link), after you register here(link)" (ACC030, Blog, 2021). These attributions identified the source and sent a reader to it, but they carry none of the durability of a formal citation. They do not have persistent identifiers, and their continued functionality depends on web pages that may move or disappear.

5.8 Publishing and sharing data

Publishing and sharing data describes how reusers make datasets available to others, whether by depositing them in a repository, distributing them on request, sharing them through collaborative channels, or building public-facing tools that display them. Reusers in this corpus frequently expected, and sometimes demanded, data sharing from the mainstream climate science KI. However, their own data sharing practices varied. Ultimately, data sharing was grounded in a commitment to reproducibility and transparency, and it was pedagogical as much as scientific: the goal was not only to publish findings but to recruit others into the practice of independent verification.

Reusers articulated what they expected of the mainstream climate science knowledge infrastructure regarding data sharing: that underlying data be released on request, that original records be preserved alongside any revisions, and that changes be documented and open to inspection. These were framed as conditions of legitimacy. Cohler traced the demand to disputes of the early 2000s, when, in his account, researchers were "hiding data, changing data, lying about data, purposely avoiding FOIA requests" (Cohler, Interview, 2025). His remedy came from legal practice:

> Old data should always be kept. What's the term? Forensic data analysis, you need to see how things changed all along the way. If you're lawyers who deal with documents and online things, they need to have a rock solid chain of custody of what happened, where, when, by whom, et cetera. And so that's what you need in data, because otherwise, you can lose track of stuff very quickly because there's so much stuff there (Cohler, Interview, 2025).

The same expectation appeared when the infrastructure was seen to obstruct it. A blog post recounted a Freedom of Information exchange in which the Met Office, asked for a station's classification, first asked "'Whilst we look up what the CIMO rating the site is, we will need to know why you are enquiring and what you are doing with the information?'" (ACC037, Blog, 2025). The author translated that response as follows: "'The climate is breaking down, Net Zero is the only solution. We are just not going to tell you how we know,' seems to sum up the Met Office's attitude" (ACC037, Blog, 2025). A commenter treated the point as a matter of entitlement: "FOI rules do not allow government agencies to delay or deny information nor ask why you want the data" (COMM011, Blog, 2025). Reusers in this community wanted data that could be traced with a documented history, institutions that answered questions, and requests met without justification, friction, or delay.

Their own data sharing practices were not always as transparent as what they demanded from the mainstream climate science KI, though many reusers held themselves to the same standard. Data sharing ranged widely even across the four data reuse journeys from Chapter 4, from public republication and formal reproducibility deposits at one end to no sharing of underlying data at the other. McKitrick framed his sharing commitment explicitly: "I always publish my data and code. It's partly so I don't have to spend time afterwards fielding inquiries, but it's also related to research integrity." (McKitrick, Interview, 2025). He deposited both the data and the code in Mendeley Data after publication. His "thoughts about this were influenced in part" from "long battles over the data and code used to produce his 1998 hockey stick graph" where eventually he and his collaborator "got most of the data and methods released" that were used to produce the hockey stick graph (McKitrick, Interview, 2025).

Burton's sealevel.info provides continuous, publicly accessible, multi-format data publication, with JSON files, CSV files, Excel spreadsheets, and PNG images updated frequently for NOAA and PSMSL holdings and manually for edge cases. Engelbeen shared informally, on request, with no formal deposit: "Everybody is free to ask for use of all our data and graphs, but until now we [haven't had] any formal requests" (Engelbeen, Interview, 2025). Burnett did not share the underlying data: the PNG image pasted into the blog post was the only artifact that traveled forward, and the link to the FAOSTAT Compare tool was provided but not to the underlying dataset itself.

Beyond the four journeys, other reusers described their data sharing philosophies and guidelines. For instance, Taylor, speaking collectively for the Heartland Institute, stated the following:

> I don't care what your credentials are and what you've accomplished. So I would hope and think, and I believe that for anything that we're reporting … If someone asked me, "Okay, well where did you get that graph from? Tornadoes or floods or whatever else," I'll say, "Here's where I got it. Here's how I got it." I would hope that everyone would be as transparent as possible. We certainly aim to be, and I believe we have been. We'll continue to be (Taylor, Interview, 2025).

In a collaborative peer-reviewed paper, ACC019, ACC021, and ACC022 made all their processed data available as supplementary: "we have included all of our data as supplementary information, we're not hiding anything, all of our data is available for anybody to go through" (ACC019/ACC021/ACC022, Nelson Podcast, 2023). They also made their 27 TSI reconstructions publicly downloadable: "you can download the series as well, we provide all of our data in supplementary material, anybody that wants to use this data" (ACC019/ACC021/ACC022, Nelson Podcast, 2023). The paper's own data-availability statement bears this out, directing readers to downloadable supplementary files that include its figures and tables, the Excel calculations, underlying data, the analysis code, and output (ACC019/ACC021/ACC022/ACC034, Journal Article, 2023).

Others agreed: sharing is what makes transparency and reproducibility possible. Cohler articulated the underlying logic directly:

> Science has to be reproducible. If it's not reproducible, it's useless. If you can't say, 'Here's the data and here's the calculations I did.' And if someone else can't take the same data and get the same results, then your results are no good (Cohler, Interview, 2025).

This sharing orientation was pedagogical as much as it was scientific: the goal was not only to publish findings but to recruit others into the practice of independent data verification. Many reusers took a participatory approach, encouraging their followers to create their own visualizations and reproduce their findings. This impulse drew in part on a conviction that a single person can unsettle a consensus, a belief one Reddit commenter stated as follows:

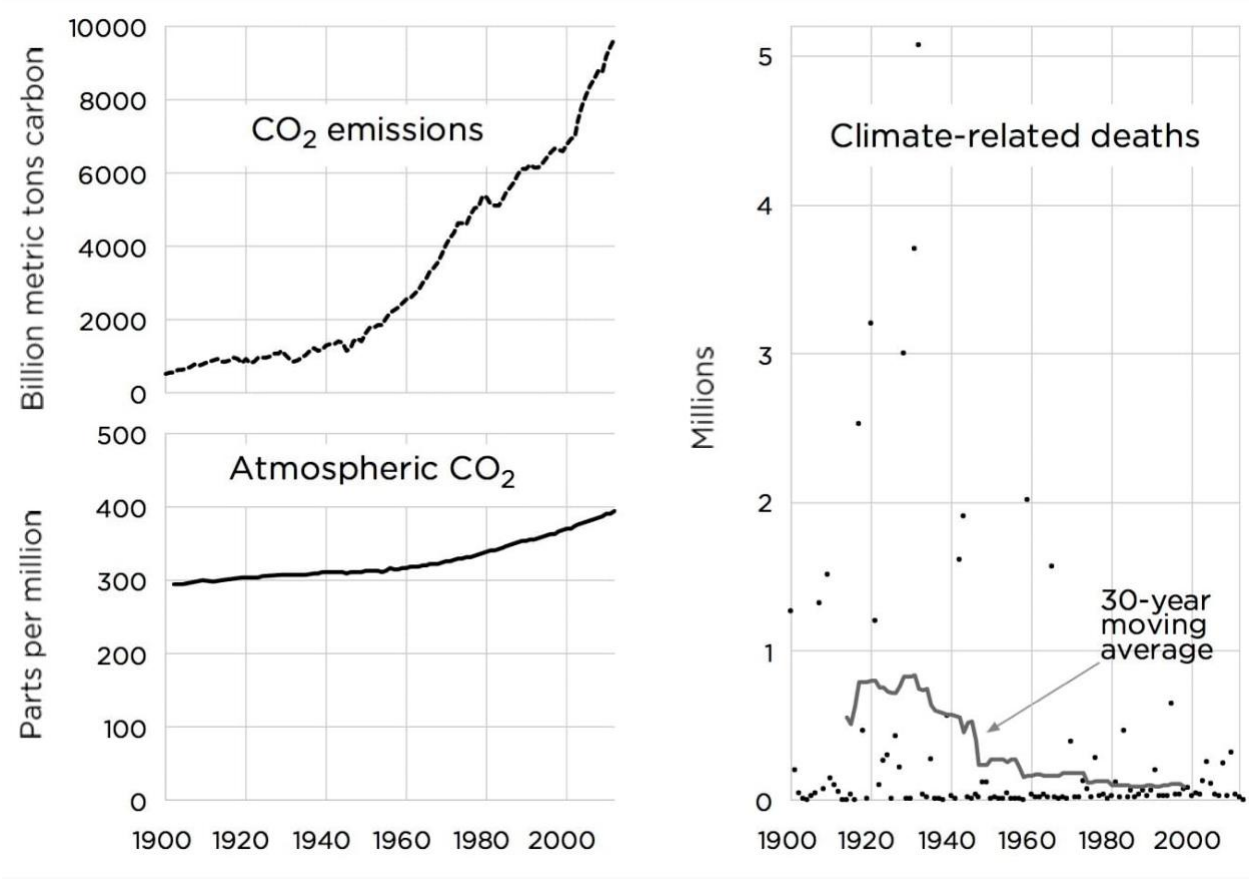
> Science isn't a popularity contest. Science is a standardised process to find the true facts about a theory. Even if just one person provided information which throws a 'scientific

> consensus' into question - it breaks that 'consensus'. As Einstein stated when 100 scientists wrote a letter saying he was wrong: "were I wrong, one professor would be enough" (COMM007, Reddit, 2021).

One example is the companion website to a book published by ACC038, which discusses the ethical benefits of fossil fuels. The website's purpose is to let readers rebuild his charts: the site supplies the figures and, "where possible," the "numerical data which makes it possible for the interested individual to recreate the figures" (ACC038, Website, 2021). Each figure is paired with its sources and some of the steps needed to reproduce it. The note for the graphs on fossil fuel use and climate-related death is shown below, along with a screenshot of the graphs themselves in Image 5.3.

> The CO2 emission data can be used as is from the CDIAC data sets. The same is true for the ice core and Mauna Loa CO2 data from the Scripps data set. The climate-related deaths can be obtained by querying the EM-DAT database for all climatological, hydrological, and meteorological disasters, sorted by year for the entire globe. Since publication of the book we decided that the disaster category of "complex disasters" included in the book version should not be considered as it covers disasters like famines that can, but do not necessarily, include climate-related causes. That results in some changes to annual values, including that of the previous record year of 1932 with 5 million deaths. The new record year is 1931 with over 3.7 million deaths.

Image 5.3. “Figure 1.9/5.1: More Fossil Fuels, Fewer Climate-Related Deaths” (ACC038, Website, 2021).



Not all sharing requests were completely fulfilled. ACC036, when asked by a commenter to share all of his code and data, described the obstacles candidly: the data folder was approximately "a gigabyte and a half," the code was "FAR from user-friendly," dependent on a cascade of functions he had built over years, and he would "MUCH rather do new and interesting research than spend time beating my code into usable shape" (ACC036, Blog, 2021). He did, however, provide the commenter with a folder of basic functions and detailed instructions:

> With that said, the basic functions that I use are in a folder here. If they drive you nuts, don't blame me, they did that to me long ago. If they depend on other functions, let me know, I'll see what I can dig up (ACC036, Blog, 2021).

From there he walked the commenter through his entire setup step by step. This included where the CERES data come from, what format they arrive in, and how he converts and stores

them (the processing described in 5.3). He closed with an open offer of further help: "I hope this helps. Let me know if you need more info" (ACC036, Blog, 2021). The reluctance was understandable and real, but so was the instruction. Even a request he resisted fully satisfying became a step-by-step tutorial.

A second refusal brought up what counts as relevant data. Asked to release the exchanges with the AI model used to prepare a review paper, Cohler declined. He stated the following about the request, and the reasons why he declined:

> We [got a request] from one person, but it's ridiculous. It's a ridiculous request. … we had Grok write a response to the person. And we said, "Even if you think about this for one moment, that is exactly the same as saying, tell us every conversation you had with each other." … Who's going to do that? Nobody's going to do that. You can't do that. And the whole point of science is that the paper speaks for itself. … if there's anything untrue, tell us what it is. It's all in the paper … This gets to the data issue too, right? If you write an original research paper where you're actually doing a new experiment or something, then you're doing calculations. You got to verify all those calculations. You got to make sure the data was accurate … but we didn't even do that level because this is a review paper. So every one of those citations we cite, they did that in that paper. So when you review this paper, you have to review it as a review paper. In other words, you have to make sure, did it cite the sources accurately and did it use their conclusions correctly … and then all the numbers that we cited, which all came out of those papers, did they do the numbers right? (Cohler, Interview, 2025).

Both refusals are instructive. Where ACC036's turned on effort and the state of his code, Cohler's turned on principle, a claim about what verification requires. For Cohler, the object of review is the published review paper and the sources it cites, not the process that produced it, so the model exchanges fell outside what he owed a reader. The review paper generates no new data of its own, so for Cohler verification meant checking that the cited sources were used accurately, rather than providing the underlying "data" or the model conversations. Together the two cases show that even a community that treats transparency as a condition of legitimacy sets limits on its own sharing.

Chapter 6 Knowledge Infrastructures They Rely Upon

The data practices documented in Chapters 4 and 5 did not occur in isolation. They depended on infrastructure: the repositories that store and maintain data, the tools that make it accessible and analyzable, the social networks that circulate and validate datasets, the epistemic values that guide judgments about evidence and trustworthiness, and the publication platforms through which findings travel. Paul Edwards defines knowledge infrastructures as "robust networks of people, artifacts, and institutions that generate, share, and maintain specific knowledge about the human and natural worlds" (Edwards, 2010, p. 17). This chapter takes up the second research question of the dissertation: the knowledge infrastructures that ACC skeptics rely upon and build.

The chapter is organized in three parts. The first examines the more technical components of the infrastructure: the data repositories and archives the community relies on, the computational tools it uses and builds, and the publication platforms through which findings circulate and are legitimated. The second turns to the more social components of the infrastructure: the networks and communities that organize the work, and the epistemic values that hold the whole together. The third part examines infrastructural inversion as both a community practice and a feature of the alternative KI itself. First, it examines the politics of data access, archiving, and deletion, where the community makes the infrastructure's normally invisible technical and administrative decisions publicly legible. Second, it examines accusations of data misuse as sites of infrastructural friction, where competing knowledge infrastructures intersect and their underlying assumptions, values, and dependencies become visible.

The division between technical and social components is analytical, not ontological. Repositories, tools, and publication venues are treated as technical infrastructure. Networks, mentoring, community review, and collaborative knowledge production are treated as social infrastructure. In practice the two are intertwined: a repository is maintained by people and embeds their decisions, and a community is held together in part by the artifacts it builds and

shares. The separation organizes the analysis. It does not claim that any component is purely one or the other.

Reference to mainstream science recurs throughout this chapter, but only to orient the reader to the infrastructure ACC skeptics respond to, not as a comparison case: this dissertation studied one community and claims about how its practices differ from those of mainstream scientists lie beyond what the evidence here can support. The theoretical implications are developed in Chapter 7.

6.1 Technical infrastructure

I cover three components of the technical KI this community relied upon and built: the data repositories and archives that store and provide climate data, the computational tools that make those data accessible and analyzable, and the publication platforms through which findings circulate and acquire legitimacy. Each draws on the mainstream climate science KI, and each is also a site where the community builds its own alternatives. This section takes them in turn.

6.1.1 Data repositories and archives

> And the beauty is… even John Q. Citizen can find data showing that temperatures have been warmer than today for most of human history. And pretty much all the other topics we address, most of that information is freely available on the internet. And when that information is available as data, as opposed to a study, a summary, or something that's subject to manipulation. I mean, it's wonderful… we have the actual data. (James Taylor, Interview, 2025).

All four data reuse journey subjects drew primarily from the same institutional repositories that mainstream climate science depends on: NOAA, FAO, PSMSL, the Global Carbon Project, and the Met Office Hadley Centre. This reliance is not incidental. It reflects the entanglement between the skeptic knowledge infrastructure and the mainstream climate science knowledge infrastructure, which it often critiques. The skeptic knowledge infrastructure does not stand apart from open mainstream science data. It is built on top of that data and depends on its continued maintenance and accessibility.

The community's relationship with these repositories was not simply about reusing the data provided. The community is not uniform in how far up the measurement chain its skepticism reached, but the distrust of institutional processing was the most consistent. For most,

this distrust did not lead them to abandon institutional repositories, partially because there are no other options for such global data. Instead, it led them to comparing versions of datasets, returning to less-processed data, and scrutinizing for evidence of contamination.

Friction points in the data repositories were not neutral events for this community. McKitrick's reading of the removal of older wildfire data is one example: "They don't explain what changed or why the numbers can't be compared…Selectively deleting records creates the impression of an agency wanting to hide the numbers" (McKitrick, Interview, 2025). ACC042 described a similar interpretive move when reflecting on a local NOAA weather station that has been decommissioned since 1985 but continues to appear as an active source through interpolation, with each daily reading marked "E" for estimated:

> That station has not been there since 1985, but the station interpolation algorithm requires you to have data all the time. So what happens at the station, even though it stops and we don't have observations, they go to stations around it and then attempt to estimate what the temperature would have been if that station were still there (ACC042, Nelson Podcast, 2024).

ACC042 treated this as evidence of a deeper pattern. In his account, individual friction points are not isolated errors but evidence of systemic bad faith. Whether intentional or not, these infrastructural events undermined trust among communities already inclined toward institutional skepticism.

This scrutiny extended beyond the data itself to the computational infrastructure used to produce them. Where processing code was unavailable for independent verification, community members treated that inaccessibility, or really any inaccessibility, as an infrastructural failure. ACC031 explained that while the National Centers for Environmental Information published a paper describing how the GHCNm dataset is processed, the code was held "specifically by the US government that we cannot test externally [or] even replicate it" (ACC031, YouTube, 2021). He concluded that the code must be made public to understand how the dataset is calculated. The processing pipeline, not just the data it produces, was treated as a component of the knowledge infrastructure that should be open and auditable.

6.1.1.1 Openness as a Spectrum

All four journeys revealed a gap between data being open in name and actually accessible in practice. For McKitrick, the C14 dataset was listed in the published paper as available on a

website that no longer existed; access required professional standing, knowing how to email the lead author directly, and having that author respond. For Burton, PSMSL had not updated Polish tide gauge data in roughly 25 years. For Engelbeen, the CDIAC link he relied on no longer resolved after CDIAC data were transitioned into ESS-DIVE, illustrating how repository migrations can break established reuse pathways. For Burnett, the EPA Climate Change Indicators page was removed in December 2025, taking with it a graph he had reused across dozens of articles.

These were not only obstacles that individual reusers encountered and overcame. They are findings about the infrastructure itself. Openness is not binary; data are not simply open or closed. Here, data exist on a spectrum of accessibility shaped by things like documentation practices, website maintenance, institutional decisions about archiving, political and administrative decisions about what to host and retain, and the social networks that provide workarounds when formal access fails. Nominal openness and actual accessibility are not the same thing, and the distance between them is uneven across reusers.

Movement and removal within the official infrastructure widened this distance. Data and documentation were sometimes relocated, as when many NOAA datasets moved from NCDC to NCEI. Datasets are sometimes removed because of errors, newer versions, or administrative changes. Links to datasets and documentation were not always updated to reflect these changes. For reusers who built their practices around specific datasets and URLs, reorganization and removal disrupted established workflows.

When reusers saw what they treated as gaps or failures in the official infrastructure, they often responded not by working around them but by building something. McKitrick built a more complete version of the C14 dataset and deposited it on Mendeley and built a website to demonstrate how he argues climate data should be presented to the public (yourenvironment.ca) (McKitrick, n.d.). Burton built a tool that combines sea-level tide-gauge datasets usually siloed across repositories. ACC021 and ACC022 compiled station history metadata from meteorological societies across more than 24 European countries, on their argument that NOAA's homogenization adjustments did not account for physical station changes documented in national records and made the compilation available as supplementary material to their published papers. These built alternatives appearws across every dimension of the infrastructure: datasets, tools, repositories, and publication venues. These community-built components often

sought legitimacy by adopting the conventions of scholarly infrastructure, a pattern developed in the discussion of publication platforms in 6.1.3.

Building alternative datasets and repositories did not resolve the underlying dependency. Datasets posted to blogs or private archives could circulate without clear provenance, stripped of the contextual information needed to evaluate them. Once a dataset was downloaded and recirculated, later corrections to the source did not always reach the copies already in circulation.

6.1.2 Computational tools and platforms

The tools ACC skeptics used to access, analyze, and visualize climate data were a central component of the community's knowledge infrastructure. They did more than enable particular forms of data reuse. They mediated access to datasets, shaped what kinds of analyses can be performed, embedded assumptions about transparency and accountability, and connected community data practices to the institutional repositories on which they depend. The tools form a layered infrastructure. The community-built layer sits on top of official data and depends on the stability and continued accessibility of institutional sources. It enables analyses that the original repository interfaces do not support, and it creates new dependencies and new points of failure.

The most basic computational layer was the analyst's own software. Many community members analyzed data and produced visualizations in Excel, R, or Python. ACC043 who uses both states the following in a blog post after completing the data analysis in R:

> R makes it easy to do the calculation, but it is unsatisfying since we don't get much understanding from running it or from the output. So, let's do the same calculation with Excel and go through all the gory details," a preference for hands-on engagement that reflects the broader DIY analysis value (ACC043, Blog, 2021).

He treated the duplication as a way to expose each step of the calculation. This duplication of data analysis aligned with the skeptic values of teaching others how to reuse data and questioning assumptions behind analysis.

Beyond personal software, three kinds of shared interactive tools appeared across the corpus, distinguished by who builds and maintains them: tools provided by official repositories, tools built by third parties unaffiliated with the skeptic community, and tools built by the ACC skeptic community.

Official repository tools carry institutional authority but are frequently limited to a single repository's holdings and are designed for the repository's primary users rather than for the

analytical needs of skeptic reusers. The most frequently used in the corpus is NSIDC's Charctic Interactive Sea Ice Graph, which visualizes the Arctic and Antarctic seasonal sea ice cycles from 1979 onward. Others include NOAA's tide and currents pages, FAOSTAT's Compare Data interface, NASA's interactive data viewers, and KNMI's Climate Explorer. These tend to be more constrained than the third-party and skeptic-run tools, limited to their own repository's holdings. Skeptics engaged with official repository tools, especially when those tools offer affordances that community-built alternatives do not. KNMI Climate Explorer, for instance, uniquely for an official tool allows users to investigate climate data, including time series, model scenario runs, and more; it appeared in the corpus specifically for its model scenario visualization affordance, which skeptic and third-party tools do not yet replicate. When suggesting tools for time series analysis, by contrast, community members directed one another to community-built alternatives rather than KNMI.

The second category is third-party tools built by developers unaffiliated with the skeptic community: Wood for Trees and Rimfrost.no. Wood for Trees is the most popular tool in the corpus, hosting "C++ software tools for analysis and graphing of historical climate time series data, and an interactive graph generator where users can play with different ways of analyzing data" (*Wood for Trees: Home*, n.d.). It aggregates official datasets including HadCRUT4, SIDC sunspot numbers, and Mauna Loa CO_2, and provides built-in processing including running means, trend lines, Fourier analysis, and derivative calculations. What is infrastructurally significant about Wood for Trees is that it was built and is maintained by a single software engineer who accepts the scientific consensus on anthropogenic climate change. The tool's creator described it as value-neutral in its stated stance. Its homepage declines to tell users what to conclude:

> It's not the place of this Web site (or anyone else) to tell you the answers, even if I could! This is just a tool to help you dig into the data to help you form your own opinions. Whatever you decide the most important thing is that you learned what the issues in analysis are and how to test your ideas against real data (*Wood for Trees: Home*, n.d.).

The tool depends on one person's continued willingness to maintain it and on continued access to the official datasets it aggregates. Its creator notes that "with sharp tools comes great responsibility…beware of short, cherry-picked trends" (*Wood for Trees: Home*, n.d.). Wood for Trees added the ability to perform linear least-squares regression. Still, the creator notes "that it

can be fairly dangerous…depending on your preconceptions, by picking your start and end times carefully, you can now 'prove' whatever you want with temperature trends" (*Wood for Trees: Home*, n.d.). This affordance, perhaps, contributes to its popularity with ACC skeptics, and the creator's statements reflect his own awareness of how the tool can be misused.

Rimfrost.o, a Norwegian site, offered an aggregated visualization platform covering thousands of datasets, including temperature, ice, CO_2 measurements, river discharge, precipitation, glacier mass balances, ocean energy, ocean water levels, sunspot numbers, and skiing conditions. It relied on legacy Java browser applets, became incompatible with current browsers, and over the course of my digital ethnography it became inoperable. The site's maintainers described the platform as the product of "a small group of enthusiasts" working with a large and aging codebase built at a time when Java support in web browsers was standard (Rimfrost Experience, n.d.). According to the creators, maintaining the platform became progressively more difficult. The maintainers attempted to preserve the system through Webswing, a compatibility layer that allowed the legacy application to run in modern browsers without requiring users to install Java locally. However, as of 2026 this solution also appeared to have ceased functioning. Rimfrost illustrates a recurring tension. The platform was created to provide access to data that would otherwise require navigating numerous institutional repositories, but its continued operation depended on technical expertise, ongoing maintenance, and resources that were in short supply. Like institutional infrastructures, which face their own vulnerabilities, tools such as Rimfrost are subject to forms of infrastructural fragility. In this case, the primary risks were technological obsolescence and the withdrawal of volunteer labor.

The community also built its own tools like sealevel.info. These were built by community members to fill perceived gaps in official infrastructure and to enable analyses that official tools do not support. Sealevel.info, built by Burton and documented in Chapter 4, provides interactive regression analysis and visualization for over 1,200 tide gauge stations. Its spreadsheets "consolidate data from NOAA, PSMSL, and other sources to simplify examination of tide-gauge data for long term sea-level trend analysis" (sealevel.info).

Another tool developed by ACC017 took a different approach. Rather than aggregating specific climate datasets, it was built around AI commands that allow users to query, filter, trend, and compare any large dataset without writing code. Though general-purpose in design, ACC017 promoted it to the ACC skeptic community through YouTube tutorials. In climate applications,

demonstrated through tutorials on ACC's YouTube channel, the tool handled data at scale. For instance, in one tutorial, ACC017 drew on 1.3 million daily maximum temperatures for Ohio to identify a cooling trend in January temperatures since 1900, then cross-referenced a 1906 heat anomaly in the station data against a digitized New York Tribune article reporting heat waves across the state that same day (ACC017, YouTube, 2025); in another, he used the tool to compare Antarctic sea ice extent on the same date across different years, concluding that 2025 sea ice levels were close to the 1978–2025 average (ACC017, YouTube, 2025).

For the community members who built these tools, the aim was to make verification ongoing and public. Rather than auditing a single dataset or method and publishing the result, they built a parallel infrastructural component so that auditing can continue and others can take part. The origin of a tool is independent of its community use: Wood for Trees was built by someone who accepts the consensus on anthropogenic climate change, and both it and sealevel.info now function as infrastructure the community relies on, shares, and deploys as evidence.

Community members routinely shared links to unofficial tools rather than to institutional repositories. One Reddit commenter directed a newcomer to Climate4You as a primary data resource, describing it as "a wide collection of available data" (COMM005, Reddit, 2021), while another calls Climate4You simply "the best, also useful: https://www.woodfortrees.org/" (COMM012, Blog, 2021). Sealevel.info was used similarly: one commenter posted multiple sealevel.info links in rapid succession to contest sea level acceleration claims: "I don't see acceleration http://sealevel.info/MSL_graph.php?id=130-001…" (COMM013, Blog, 2021). What is analytically notable is that these tools are not official.

6.1.3 Publication Platforms

The publication venues ACC skeptics use to share findings were described in Chapter 5 in terms of what they afford. Blogs and social media offer rapid turnaround, broad reach, and informal legitimation. The blogosphere, specifically, is also where the community's audit outputs, reanalyzed datasets, alternative reconstructions, and critical readings of peer-reviewed papers, are most frequently published and shared, which makes it a central node in the skeptic knowledge infrastructure. In fewer cases, mainstream journals offer peer review credibility. Here I treat these venues as components of the knowledge infrastructure and ask of them the same

questions asked of repositories and tools: who controls access, what values they embody, and how the community builds alternatives when existing venues exclude its work.

Within the sample, ACC skeptics, with some exceptions, posted their findings on blogs, presented at skeptical conferences, and messaged on social media. The ACC skeptics had a robust culture of informal, scholarly-like communication through organizational blogs, personal websites, and social media. They also ran their own peer-reviewed journals, which have developed in response to what the community perceives as gatekeeping within mainstream peer review. This perception is captured by ACC041 who describes peer review as a "device to maintain groupthink" (ACC041, Nelson Podcast, 2022).

The ACC skeptics' publication venues visually and rhetorically mimic the publication venues of mainstream science. *Climate at a Glance* posts adopt visual styles similar to those used in EPA's Climate Change Indicators, situating the content within a familiar government-style reporting aesthetic. This visual mimicry is not incidental: it positions *Climate Realism*'s outputs as occupying the same evidentiary register as federal agency publications, lending them an air of official authority. The CO_2 Coalition report uses scientific formatting, with citations, figures, and a structured argument, but it has not undergone peer review. It is published by an advocacy organization; yet its formatting signals scientific legitimacy to readers unfamiliar with the publication's context. These alternative infrastructural components often mimic parts of the mainstream climate science to maximize their perceived legitimacy. For example, one skeptic organization calls its lists of world temperature datasets a data repository, and its weekly newsletter a journal. Many sites also provide resource lists of mainstream climate science data, tools, and organizations alongside their skeptic counterparts, presented as equally legitimate. These components adopt the trappings of scholarly discourse to establish legitimacy within and beyond the community.

One example of this mimicry is *Science of Climate Change* (SCC), an ACC skeptic-run journal founded in 2021 and hosted by the Norwegian Climate Realists. Its homepage states that its objective is "to publish — different to many other journals — also peer-reviewed scientific contributions, which contradict the often very unilateral climate hypotheses of the IPCC and thus, to open the view to alternative interpretations of climate change" (SCC, n.d.). In simpler terms, SCC states that it peer-reviews and publishes work that challenges IPCC conclusions, work it argues other peer-revied journals that are part of the mainstream climate science KI will

not accept. The journal adopts the full conventions of academic publishing, including an editorial board, its own peer review system, and volume and issue numbers, while positioning itself outside mainstream venues. Ernst-Georg Beck's historical CO_2 measurements, the dataset Engelbeen audits in Chapter 4, were published in an extended version in SCC (Beck, 2022). Community members present articles published in SCC as equivalent to other mainstream climate science publications, a legitimation strategy that depends in part on mimicking academic conventions rather than on the processes those conventions are meant to signal. What distinguishes SCC from the journals it resembles is not visible in its conventions, which match theirs, but in its editorial mission to publish work contradicting the IPCC and its position outside the established venues. The evidence here does not support a claim about how its peer review compares to that of those venues.

SCC represents the community building its own publication infrastructure from scratch. But skeptics also navigate within the existing publication ecosystem. Several community members publish in mainstream journals. Along with McKitrick, ACC019, ACC021, and ACC022 publish in *Research in Astronomy and Astrophysics* and *The Astrophysical Journal*, both published by IOP Publishing, as well as in *Earth-Science Reviews*, an Elsevier journal. These are legitimate peer-reviewed venues by conventional measures. Some of the skeptics' work also appears in MDPI journals, including *Climate and Energies*. MDPI occupies a more contested position: it is indexed in Scopus and Web of Science and claims peer review, but has faced sustained criticism over self-citations rates and biased special issues (Oviedo-García, 2021). For instance, in one of the MDPI venues, ACC019, ACC021, and ACC022, publish a journal article where the listed academic editor is another ACC skeptic. This is consistent with the MDPI special issue model, which allows guest editors to manage peer review for specific issues, creating conditions in which community members may exercise editorial influence over work by their colleagues. Some community members have the credentials and institutional standing to access mainstream peer review and do so while also publishing in venues with lower barriers to publication. The community's publication infrastructure is not uniform but stratified, with different venues serving different purposes and audiences.

6.2 Social infrastructure

The social components of the infrastructure were the networks that organize the community's work and the epistemic values that hold it together. The networks recruited and

trained new participants, circulated datasets and analyses, and subjected claims to informal review. The values, commitments to individual inquiry over expert authority, to hands-on analysis, and to transparency, shaped what the community treats as credible knowledge and how it sustained its practices over time. As set out at the start of the chapter, the separation from the technical infrastructure is analytical: the networks are held together in part by the tools and venues described above, and those artifacts are maintained by the people described here.

6.2.1 Social networks and communities

One of the most persistent symbols of ACC skeptic data reuse was the lone data auditor, the independent-minded individual, uncorrupted by groupthink, sitting at a computer and letting the data speak for themselves. Each of the four data reuse journey subjects in Chapter 4 presented themselves in this way. McKitrick did the recovery work himself rather than delegating to a research assistant because he wanted to know the data personally. Burton built and maintained the data analysis tool by himself and described himself as "a dog with a bone" who just does not "let things go" (Burton, Interview, 2025). Engelbeen spent decades studying the carbon cycle himself. Burnett answered questions himself rather than relying on the newspaper article's analyses.

But none of them operated alone. McKitrick's transparency commitments developed through years of community engagement, specifically during Climategate. Burton's initial Excel calculation did not account for the clustering of tide gauge stations. It was the skeptic email list that surfaced this problem and pushed him toward a more rigorous approach. And it was the organization opposing North Carolina's coastal development regulations that gave Burton's work political stakes and a reason to keep going for fifteen years. Without that community intervention, sealevel.info might never have existed. Engelbeen's webpage on CO_2 origins, developed through community debates in the late 1990s, was the direct basis for the CO_2 Coalition article examined here. Data reuse in this community was socially organized, shaped by collective norms, collective knowledge, and collective purpose. The lone auditor was an identity that the community promoted, but the data reuse journeys told a different story.

The social infrastructure of this community functioned in several ways. First, comment sections on blog posts served as an informal peer review. Blog posts presenting data reuse cases often had very long comment sections, reaching hundreds of pages. When errors were pointed out that the data reuser accepted, the work was changed. These threads operated as a parallel peer

review network, where claims circulated, were challenged, and were refined within the community before potentially being amplified into broader public discourse. Its gatekeepers were accountable primarily to their audiences rather than to disciplinary bodies. The comments sections were also often filled with both ACC skeptics and non-skeptics.

Second, experienced community members actively recruited and trained newcomers. They invested time in training novices in their scholarly practices and offering explanations to those they see as genuinely interested in understanding their positions. Because of this direct engagement, audience members could begin producing and sharing evidence in support of oppositional science themselves. In this way, data reusers raised new data reusers. When discussing data sharing or publishing efforts, ACC skeptics emphasized the importance of involving more people in reusing climate data. If a commenter wanted to reproduce work, the reuser provided data, instructions, guidance, and in some cases, the code used. This sharing orientation wase pedagogical as much as it is scientific. These community members differed from most science communicators in at least two key ways: first, they actively collected and re-analyzed data; second, they did so in collaboration with their audiences, systematically transforming them from passive followers into active epistemic contributors. The community's social infrastructure was not incidental to its data practices. It is the mechanism through which those practices were transmitted, refined, and sustained.

ACC021 and ACC022's experience illustrates how this social infrastructure operates in practice. After spending four years developing findings they could not publish through conventional journals, they launched an open peer review site in 2014, posting eight working papers and inviting the entire scientific community to identify errors. They had "a lot of engagement at the time," ACC021 recalled, with critical and supportive comments arriving within hours of launch (ACC021/ACC022, Nelson Podcast, 2022). Rather than ignoring critical responses, the team directly contacted scientists with relevant expertise, visiting research groups and asking "off the record, on the record, to have discussions, have we got anything wrong" (ACC021/ACC022, Nelson Podcast, 2022). This engagement produced a network of collaborators recruited not through institutional channels but through the community infrastructure of podcasts, blogs, and direct outreach. These publics, in Finn's (2018) framing, "do things together" and thus "believe things together." They shared a commitment to

participatory, transparent scholarly practices, using data sharing and reuse to signal epistemic credibility.

The community's shared purpose was itself part of its social infrastructure, the collective purpose named above. Policy stakes were explicit in three of the four journeys and implicit in the fourth. McKitrick's investigation was prompted by the Biden Administration's revised Social Cost of Carbon estimates. Burton's analytical curiosity acquired policy stakes when North Carolina debated coastal development regulations premised on accelerated sea level rise projections. Burnett's entire publication model was oriented toward contesting media claims that carry policy implications. Engelbeen's journey was the partial exception; the policy stakes are indirect, operating through the community effort to settle the CO_2 attribution question, and perhaps to turn to the policy debates the community believes it can win.

The community also overlapped with other skeptic communities, and its infrastructure traveled with it. I observed overlap between ACC skeptics and anti-vaxxers, anti-maskers, and skeptics of the 2020 U.S. presidential election results. The same members applied their data literacy to other public datasets, including 2020 U.S. presidential election data, as well as COVID-19 data. In one comment thread on an ACC blog, COMM014 applied the same recalculation practice used on climate datasets to Canadian government COVID-19 data, downloading the official figures and recomputing death percentages by vaccination category to contest the official narrative:

> The data are presented in a way that makes unvaccinated status look much more dangerous than vaccinated. But if you do the math of the outcomes "hospitalizations" and "deaths" as a percentage of "total cases" in each vaccination category, as opposed to percentages of the total cases tested positive, things look quite different. (COMM014, Blog, 2021).

Wood for Trees itself added COVID-19 case data in March 2020, its creator noting that the existing climate analysis tools could be applied directly to coronavirus data, an infrastructural expansion that reflected the same logic. The same tool and the same audit practices, extended to a new domain.

ACC021 extended his work explicitly into COVID-19, co-authoring a 38-scientist review paper on pandemic management published in the *International Journal of Public Health* describing the motivation in terms continuous with their climate work: "We realized that there

was a group of scientists who disagreed with the narrative, but we really weren't hearing anything about them. So we decided to write a paper on the disagreement in science" (ACC021, Nelson Podcast, 2025). This suggests that the data practices documented here are not specific to climate skepticism. They reflect a broader orientation toward auditing public data and institutional claims that has developed within the ACC skeptic community and been applied elsewhere.

6.2.2 Epistemic Values as Infrastructure

The data practices documented in Chapters 4 and 5 were held together by a set of shared epistemic values. Values did infrastructural work in this community. They coordinated dispersed individuals to do work together, determined which artifacts are treated as trustworthy, and supplied the criteria by which the community trained novices and evaluated work. These values were what sustain their knowledge practices across time. Where the social networks in 6.2.1 were where data practices travel, the values were what made those practices legible, the shared standard that let ACC skeptics recognize one another. Three values performed a substantial amount of this coordinating work: the primacy of individual inquiry over expert authority, objectivity understood as the absence of human mediation, and transparency. What happens when these values are pushed to their limit is taken up in Chapter 7. Here the concern is descriptive: what the values are, and how they hold the community's knowledge practices together.

6.2.2.1 Individual inquiry and the limits of expert authority

The community treated first-hand engagement with data, rather than deference to credentialed authority, as the path to reliable knowledge. Taylor states the principle plainly:

> People will say, 'Oh well, NOAA says this or this professor says that.' And that doesn't mean anything. I mean, the data is what matters because no matter how smart you are, you may not know everything. And even if you do, people sometimes will manipulate things to serve their purpose (Taylor, Interview, 2025).

The same stance recurred across the corpus. "In Science, the message is important, the messenger is not" (ACC034, Preprint, 2021). As previously stated, the Feynman line, "science is the belief in the ignorance of experts," circulated widely. From their perspective, non-skeptics defered to experts rather than think critically: "experts on everything, [so] no need to think for

[them]selves then" (COMM015, Blog, 2021). In a Reddit post's comments, commenters mocked the rhetorical weight of credentials. The first, COMM016, stated "'Scientists say' is all they need" and COMM008 responded, "That, and a mention of 'experts'…I guess that fools most people, but it doesn't fool you and me!" (COMM016/COMM008, Reddit, 2020). Individual reasoning was set against any type of collective authority or consensus as a matter of principle: "One can only be a critical thinker as an individualist, one can't be a critical thinker as a collectivist" (COMM010, Blog, 2025).

At first glance, this stance appeared to reject expertise altogether. Yet that was not the case. Rather, it distinguished between expertise that is perceived as independent and expertise that is perceived as constrained by institutional consensus. The issue was therefore not whether expertise matters, but which experts were considered trustworthy and why. Experts who are seen as uncorrupted by groupthink, independent of institutional consensus, are valued and trusted. The community foregrounded these experts' credentials: "distinguished academic and professional climate scientist" (ACC045, Blog, 2021), "past president of the American Association of State Climatologists" (The Heartland Institute, 2021), or "a phd [sic] himself" (COMM007, Reddit, 2021). In a self-published work, ACC046 thanks "climatologists, chemical engineers, reservoir engineers & simulators, heat transfer experts, solar physicists, geologists, and geophysicists" (ACC046, Preprint, 2021). Experts who were skeptical of ACC are portrayed as untainted by groupthink, their credentials lending credibility while still encouraging independent verification. Authors who publish findings that call into question ACC are granted a "special credence to the sincerity of their belief in the validity of their data" for "'sticking their necks out' when reporting something considered by much of the scientific community to be incorrect" (ACC047, Website, 2021).

6.2.2.2 Objectivity

The community treated objectivity as something achieved by minimizing human judgment and letting data stand on their own. The ACC skeptics portrayed themselves as #climaterealists against #climatealarmists, framing their own practices as free from the social biases that corrupt institutional science. As two commenters put it, non-skeptics live in a "self-induced haze [and] are [incapable] of being objective enough to recognize what is self-evident for normal people" (COMM013, Blog, 2021) and "always go for the heartstrings, bypassing any real objectivity" (COMM017, Blog, 2021). The Institute for Energy Research states that "public

policy… should be based on objective science, not emotion or improbable scenarios" (Website, 2021). This ideal was also directed outward: mainstream scientists were criticized for failing to meet it, for allowing sociality to corrupt what should be an individual engagement with data. It generated a preference for primary data over processed outputs as the objective ideal, and for individual verification over expert authority.

Raw and observational data were ranked above processed summaries, and both above models. Taylor made the ranking explicit when describing how he assessed dataset:

> Is this objective data that is not subject to manipulation, to cherry-picking?…if I can get the actual data, that's the best…not the summary, not what the bureaucrats are saying about climate change and tornadoes, but what are the actual data…that's the best because that's objective, it's quantifiable (Taylor, Interview, 2025).

In this formulation, objectivity is treated as a property of proximity to measurement rather than of method or institutional procedure. Raw measurement sits at the top of the hierarchy: closer to the collection instrument is treated as closer to objectivity.

Models sat at the bottom. Taylor dismissed them as the most manipulable and unobjective artifact:

> How much of it is based upon objective data that I can verify versus models? So much out there is garbage because it's models. Models, I could program a model today that would have aliens invading the earth. It's called space invaders…Doesn't mean it's happening (Taylor, Interview, 2025).

In other words, models encoded their maker's choices. As ACC041 put it the computer model "simply mirrors the intellectual choices of the modeler" (ACC041, Website, 2021). ACC024 cited the *Bulletin of the American Meteorological Society*, noting that models may be optimized for metrics tied to "goals, expertise, or cultural identity" of a modeling center. He interpreted this as proof that "it is the scientist, not the science that determines what is 'anticipated' and 'acceptable'" (ACC024, YouTube, 2021), thus violating objectivity. While no artifact was entirely free from subjectivity, more "objective" ones are used to question others. This hierarchy of objectivity maps directly onto the community's data practices documented in Chapter 5: the preference for raw over processed data, for observational over modeled data, for self-conducted analysis over institutional summaries.

6.2.2.3 Transparency

The community's transparency commitments were the most developed expression of these epistemic values and are consistently operationalized. McKitrick and Burton were the clearest examples in this sample of deep commitment to transparency. McKitrick deposited turn-key R code with one-click reproducibility and documented every dataset recovery decision in a published supplement. Burton built self-documenting JSON files with extensive metadata and maintained a public changelog. McKitrick operationalized his argument through his assertion that studies with unavailable data should be excluded from public policy formation entirely (McKitrick, Interview, 2025). Not all community members shared the same level of commitment to transparency. Burnett's efficiency-oriented workflow, answering the question quickly with trusted pre-made visualizations, sat in tension with the deeper transparency commitments of Burton and McKitrick. Engelbeen shared data informally upon request but had made no formal deposit. These variations are worth noting.

The community framed transparency as the ground on which its work should be trusted and reached for the language of open science to do so. A skeptic-led report states that "honest scrutiny and scientific transparency should be at the heart of our policymaking," and casts its open-comment process as "an opportunity for dialogue, learning, and clarification of areas of disagreement" (ACC048, Report, 2025). One blogger invokes Oppenheimer:

> We do not believe any group of men adequate enough or wise enough to operate without scrutiny or without criticism. We know that the only way to avoid error is to detect it, that the only way to detect it is to be free to inquire. We know that in secrecy error undetected will flourish and subvert (Oppenheimer, quoted in ACC014, blog, 2021).

Burnett tied transparency to public accountability:

> When you're going to pass laws…when you're going to restrict freedom or advance certain industries over other industries…the underlying reason should be understandable to your lay Wall Street Journal reader (Burnett, Interview, 2025).

The KI also had edges it did not police. The community's skepticism was applied rigorously to mainstream science institutions and their data practices, but less consistently to its own. Funding source questions directed at community members were sometimes dismissed as ad hominem rather than legitimate transparency concerns, even when they asked the same questions of mainstream climate scientists. The selective application of skepticism, rigorous toward certain

knowledge producers and permissive toward others, was itself an infrastructural feature, shaping which questions got asked and which remained unexamined.

These three values were the connective tissue of the knowledge infrastructure. They were why the community preferred some data and tools over others, how it recognized good work, and what it transmitted to novices.

6.3 Infrastructural Inversion and Friction

Bowker and Star (1999) describe infrastructural inversion as an analytical move: the deliberate reversal of figure and ground, foregrounding the normally invisible background of practice to make its assumptions, values, and exclusions visible and analyzable. The ACC skeptic community's engagement with mainstream climate science KI was characterized by a sustained commitment to infrastructural inversion. For this community, infrastructural inversion was the audit orientation in operation. Auditing a dataset, a processing pipeline, or curatorial decisions meant taking the normally invisible work of the mainstream climate science KI and rendering it visible as an object of scrutiny. They did not take the infrastructure for granted. They scrutinized it, contested it, and exposed it, making visible assumptions about data, access, and authority that remain invisible to most users. First, I examine the politics of data access, archiving, and deletion, where the community makes the infrastructure's normally invisible technical and administrative decisions publicly legible. Second, I examine accusations of data misuse as sites of infrastructural friction, where competing knowledge infrastructures intersect and their underlying assumptions, values, and dependencies become visible.

6.3.1 The politics of access, archiving, and deletion

Data access was where infrastructural inversion took its most explicit form in this community. When official repositories moved, removed, or restricted data, community members noticed the technical mechanisms involved, named them publicly, and treated them as politically legible. ACC029's criticism of NOAA's robots.txt configuration is a clear example. He wrote in 2017::

> I hope that President Trump will issue an executive order, ordering that government and government-supported websites STOP DELETING old versions of ‘their’ (our!) data sets, to hide the extent and nature of their alterations, and STOP using robots.txt and server configuration tricks to BLOCK the archiving of old versions of their data and web

> pages by The Wayback Machine. What possible excuse can NOAA have for such behavior? (ACC029, Blog, 2017).

The robots.txt configuration is a technical convention that most users never notice. ACC029 noticed it, named it, and made it a site of political contestation. Their complaint identifies a tension between open data rhetoric and the decisions that shape what can be accessed.

The same pattern of infrastructural inversion, noticing a technical decision, naming it, and making it a site of political contestation, appeared across the broader corpus in response to other data removal events. The removal of pre-1983 wildfire acreage data from the National Interagency Fire Center website was the most frequently cited example. Multiple community members noticed that historical data showing high burn acreage in the 1930s had been removed and replaced with a disclaimer citing concerns about data quality prior to 1983. H. Sterling Burnett named what he believed the agency's underlying motivations were directly:

> The problem was the data was very inconvenient. It had been used in dozens, if not hundreds of peer-reviewed literature. But now suddenly it was inconvenient because it didn't tell the narrative the government was wanting to tell. So they cut off past data (Burnett, Interview, 2025).

Taylor discussed the deletion as well as the community's counter-response:

> They erased everything before the lowest year of fires…and then they say, 'Oh, well, we don't have confidence in the quality of that data.' Well, why not? Why did you have confidence forever and now you don't? So yeah, fortunately we have the Wayback Machine where we can go back (Taylor, Interview, 2025).

The Wayback Machine, the same archiving tool ACC029 had complained NOAA was blocking in 2017, became the community's primary tool for recovering removed data or noticing changes by 2021. A similar dynamic appeared in the corpus around tornado trend data removed from the NOAA NCEI website: ACC014 retrieved the original page through the Wayback Machine and published a side-by-side comparison with the replacement page, noting that the original had shown a declining trend in strong tornadoes while the replacement showed no such discussion (ACC014, Blog, 2021).

ACC042 extended this interpretation to the behavior of ACC non-skeptics at the start of the first Trump administration: "There were people staying up all night, going online, saving all

the climate data because, you know, he's going to come in and erase it all" (ACC042, Nelson Podcast, 2024). He read this not as legitimate data preservation but as evidence that those doing the downloading had already been altering records themselves and assumed the other side would do the same: "Because according to Saul Alinsky, that's how we would have done it. So you blame the other side for what you would do" (ACC042, Nelson Podcast, 2024).

It is worth noting that the community was not uniform in how it applied this infrastructural critique. James Taylor and H. Sterling Burnett directed their concern at what they characterized as politically motivated data removal and changes under the Biden administration. Burton explicitly extended the concern to the current administration: "There's been a lot of hoo-ha at the moment about Trump might delete a bunch of climate data. And I don't know whether that's going to happen or not. I don't approve of it if it does" (Burton, Interview, 2025). Tom Nelson was similarly clear: "I'm not a fan of deleting data though…more data is better and more open discussion is better" (Nelson, Interview, 2025). Not all community members shared this position. ACC028, writing in January 2025, welcomed the removal of climate content from federal agency websites, describing the Biden-era EPA climate pages as "claptrap" and writing that he "can't wait to see a lot of this stuff disappear" (ACC028, Blog, 2025). For ACC028, the infrastructural critique ran in the opposite direction: the problem was not that data are being removed but that the wrong content had been present for too long. ACC029 has been consistent across administrations: his 2017 complaint about data deletion was directed at NOAA under the Obama administration, and he frames his transparency commitments as principled rather than partisan (ACC029, Blog, 2017). The critique, in other words, was not simply a weapon deployed against one political side. For some community members, it reflected a consistently held commitment to open data; for others, it was applied selectively, directed at institutional arrangements they distrusted while welcoming removal of content they considered politically compromised.

6.3.2 Data Misuse accusations as infrastructural friction

Star and Ruhleder (1996) note that infrastructure becomes visible when it breaks down. Accusations of data misuse work the same way. They are moments of friction where the assumptions, norms, and dependencies that usually sit in the background of data reuse are forced into view and contested. This friction is the point where two communities' incompatible norms come into contact over the same data, making the norms themselves visible. In that sense the

accusations are the audit orientation in operation: the community's habit of treating every claim as something to be checked, turned on specific works and the people who produce them. Documenting them is not a matter of certifying which uses are legitimate. It is a matter of showing where the infrastructure's seams are, the points where different epistemic communities, holding different norms, draw on the same open data and reach incompatible conclusions about what counts as proper use. Following 2.4.2, these accusations are handled two ways. Most are objects of study: claims that surface infrastructural tensions, recorded without being resolved. In a narrower set of cases, where a practice is recognized as misuse both by mainstream climate science and by the community's own members, that recognition is documented.

The clearest case of friction as a clash of norms was McKitrick's reading of the C14 and M17 papers, two published climate studies based on the same data he obtained and reanalyzed (introduced in Chapter 4), which, from his perspective, committed misuse through inadequate disclosure of data incompleteness. Both papers do disclose the scope of their data use: C14 notes that roughly half the entries were omitted from the regression, and M17 reports its observation count. McKitrick read these disclosures as insufficient. Referring to N=1,722 when only 862 records were usable for regression was, in his framing, misleading, particularly given that those findings fed directly into federal climate policy. What this accusation reveals is a friction point where two different sets of norms about transparency come into contact. McKitrick's transparency commitments demand full accounting of what data were used and why. The disclosure conventions of science publishing in 2014 and 2017 treated a brief methodological note as sufficient. The friction arises not because disclosure is missing but because the two communities apply different standards to the same act of methodological disclosure. This friction only surfaces when someone following different norms, and perhaps outside the original community, scrutinizes the work. This is the pattern the rest of the section traces: the same open data, used by communities with different standards, producing incompatible judgments about what counts as proper use.

Another example is from H. Sterling Burnett's perspective, where he thought journalists were committing misuse by citing a single study as definitive without seeking out real-world data. What this accusation reveals is a point of friction between two infrastructural components, again, operating under different norms: scientific publication and media communication. Burnett's entire publication model at Climate Realism was organized around this friction,

attempting to insert a counter-infrastructure into the translation layer between science and public communication.

Data misuse accusations ran in more than one direction: from skeptics toward non-skeptics, from non-skeptic commenters toward skeptics, and, less often, from skeptics toward one another. Misuse accusations clustered around a recognizable set of practices. The most common was the selection of temporal endpoints, naming a start or end date that yields a desired trend. Related is the substitution of local or regional data for global claims, and its reverse, dismissing global averages in favor of local observations. A third concerns correlation presented as cause. A fourth is directed not at data but at the people who steward them: publishing private correspondence and naming curators. The discussion below treats each. What these accusations reveal is not only individual questionable behavior but specific moments of infrastructural friction: where epistemic values collide with disclosure norms, where archiving policies leave no record of what changed between dataset versions, where the translation layer between scientific publication and public communication carries little accountability, and where quality assessment depends on informal community mechanisms rather than institutional ones.

Cherry-picking start and end dates was the most frequently named practice. Non-skeptic commenters called it out in skeptic posts: "The trick that the article is using is cherry picking one month" (COMM018, Reddit, 2021), providing a Wood for Trees trend link as counter-evidence to a post that selected a single UAH data point rather than a trend. Another non-skeptic noted that a skeptic blog "is proven over and over again to cherry pick data, picking a single month in a specific area" (COMM020, Reddit, 2022). Skeptics accused mainstream science of cherry-picking the 1979 start date for satellite records precisely because it coincides with a relative cold period, making subsequent warming appear more dramatic. As one commenter put it: "Climate alarmists usually pick 1979 to start their data reporting because it was a relative minimum cold period and thus makes it easy to show a trend correlated to CO_2 increases" (COMM021, Reddit, 2021). The shared vocabulary is itself notable: cherry-picking is a methodological concern all sides recognize and invoke.

A related accusation concerns the substitution of local or regional data for global claims. One non-skeptic commenter challenged a skeptic analysis as "It's not a finding, it's cherry picked regional data compared to a planet's predicted average" (COMM019, Reddit, 2022), calling out a local-to-global move the original did not acknowledge. This local-to-global substitution appeared

across the corpus, where temperature records from specific stations or regions were used to make claims about global trends, or conversely, where global averages were dismissed in favor of local observations that contradicted them.

A third concerns correlation presented as cause. Some ACC community members read correlations between rising fossil fuel use and improving welfare indicators, life expectancy, access to clean water, falling disaster deaths, as evidence that fossil fuels are beneficial. ACC038's analysis creates figures that reflect this: correlations between rising fossil fuel use and improving health outcomes, declining poverty, and falling disaster deaths are presented as analytically significant without addressing whether the correlation reflects a causal relationship or a confounding variable (ACC038, Website, 2021). A fellow skeptic, ACC023, who elsewhere argues for letting data speak, named the logical problem in a different context:

> Even if the correlation would have been strong, we must bear in mind that correlation and causation are two fundamentally different concepts. For instance, there is a strong correlation between CO_2 increase and global poverty decrease, but more CO_2 does not cause less poverty in the world (ACC023, Report, 2020).

The irony is that the same logical error ACC023 names, treating correlation as causation, is called out in some skeptic analyses.

A fourth kind of accusation is directed not at data but at the people who maintain them. ACC026 published his correspondence with an EM-DAT database curator in a blog post, naming the curator and her director, reproducing their emails, and framing the curator's acknowledgment that pre-1998 disaster increases reflected reporting changes as evidence that she had endorsed a misleading UN report: the two "both knew in 2019 about this issue — admitted openly in the email…then two years later act as reviewers of a report in which data acknowledged to be incomplete is used as if it were valid" (ACC026, Blog, 2021). This is friction directed at the human layer of the infrastructure, the people who steward the data rather than the data themselves. The EM-DAT documentation did not, at the time, flag the reporting-rate limitation. As a practice, it marks the point where data critique extends to the curators behind the data. The friction here is between two sets of norms. Publishing private correspondence and naming the curator and her director enacts the community's commitment to radical transparency: the principle that nothing relevant to evaluating a claim should be withheld from public scrutiny. In this case, that principle collides with professional norms that private correspondence is not

published without consent and that criticism should focus on data and methods rather than the individuals who maintain them. Radical transparency provides no such protections; its logic favors disclosure. The dispute is not simply about the validity of the reporting-rate critique. It reveals a disagreement about where transparency should stop and what obligations are owed to the people who steward open data.

Misuse accusations also ran within the community, although they were relatively rare in my sample. In a comment thread, one community member critiqued another:

> That's [ACC017]'s data mining - not good science! It is the sea ice extent for one day of the year – August 26. And it arbitrarily selects the two lowest points pre-2000, and then compares them with the highest point in the past five years (COMM012, Blog, 2021).

The community's informal review structure through comments sections could audit its own research, as COMM012 showed. But the infrastructure of this community was organized around external audits rather than internal accountability. Scrutiny directed consistently outward at mainstream science turned inward only occasionally, through informal mechanisms that were often less visible than the original data reuse they critiqued.

None of these accusations absolves the others, and that is precisely the point. Misuse is not a fixed category that can be adjudicated from a neutral position. It is a site of infrastructural friction where different epistemologies, values, and political stakes converge and become visible. When accusations of misuse appear, they are rarely just technical disputes. They are moments where the assumptions built into the infrastructure are exposed and contested. Contested reuse is also generative: it offers a moment of infrastructural inversion, making visible the components that allowed it to occur and pushing infrastructures to clarify governance, sharpen documentation, and rethink access controls.

The generative dimension of this dynamic is worth noting. ACC skeptics do find and correct actual errors in the data, contributing to the improvement of climate science. McKitrick's expanded C14 dataset is the clearest example: a skeptic intervention that produced a genuinely superior version of the evidence base that future researchers will likely use regardless of their views on climate policy. Rather than treating contested reuse solely as a problem to be prevented, these cases suggest it is also a signal: a place where infrastructural values and actual practices are misaligned, and where the design of open data systems may need to account for the full range of communities and purposes they serve.

6.4 Pointing Forward

This chapter has examined the knowledge infrastructure that ACC skeptics rely upon and build: the repositories and archives they draw from, the tools and publication venues they use and construct, the social networks that sustain them, the epistemic values that coordinate their work, and the moments of inversion and friction that make infrastructure visible. Two findings run across these components. First, the skeptic KI is built on top of the mainstream climate science KI and depends on its continued maintenance and accessibility, even as it contests it. Second, it is organized around audit: the sustained scrutiny of the systems that produce, store, and circulate climate knowledge.

The relationship between these infrastructures is not one-directional. McKitrick produced a more complete and openly archived version of the C14 dataset than the original authors. Burton recovered Polish tide-gauge data that global repositories had overlooked and incorporated it into his own tool. Engelbeen's analyses reached mainstream-aligned conclusions using skeptical methods. At the same time, the tools built in response to institutional fragility are themselves vulnerable to fragility, often maintained by a single individual and dependent on the official data sources they draw from. These examples suggest a relationship that cannot be reduced to a single description. The points of contact between the two infrastructures are varied, contextual, and sometimes contradictory.

These observations require theoretical development, which Chapter 7 takes up. It turns from what the community does to what the case reveals about open data more broadly, moving from the community's audit orientation to the entanglement of the skeptic and mainstream climate science knowledge infrastructures, and closing with what this means for those who design, govern, and sustain open data systems.

Chapter 7 Discussion

The preceding chapters documented what ACC skeptics do with open data and the knowledge infrastructures they rely on and build to do it. This chapter turns from what the community does to what the case shows about open data more broadly. The argument runs sequentially:

open data infrastructures make audits easier; audit-oriented communities can form around that possibility; and those communities can develop into alternative knowledge infrastructures that both depend on and contest the infrastructures from which they draw. ACC skeptics are not anomalous; they are perhaps a predictable outcome of openness.

Section 7.1 takes up this audit orientation as the community's main organizing logic: a stance of systematic scrutiny directed at the knowledge infrastructure that produces, stores, and circulates climate knowledge. The section shows how open data makes this audit stance possible, and how the stance becomes built into the ACC skeptic community's own data infrastructure. Section 7.2 examines the values that underpin the audit orientation - transparency, objectivity, and interoperability - arguing that the community extends these values to the point they can produce effects that work against the institutions that uphold them. Section 7.3 turns to the relationship the two knowledge infrastructures produce, the mainstream climate science KI and alternative skeptic KI, arguing that their overlap is better understood as entanglement than as one of extraction. Section 7.4 reframes accusations of data misuse as sites of infrastructural friction, moments where competing epistemic norms collide over the same open data and become visible rather than resolved. Section 7.5 concludes with practical implications for the people who design, govern, and sustain open data systems.

7.1 The audit orientation embedded into the KI

The most defining characteristic of the ACC skeptic data community is its audit orientation. The community's characteristic epistemic stance is one of systematic scrutiny directed at the knowledge infrastructure that produces, stores, and disseminates climate

knowledge. As the findings chapters show, the audit orientation is widespread. Data repositories are examined for evidence of manipulation or selective deletion. Processing pipelines are treated as auditable components that should be open to independent verification. Dataset versions are compared across time to detect changes. Institutional decisions about what data to archive, update, or remove are treated as politically legible acts rather than routine curatorial choices. This audit orientation runs through the community's data practices, its tools, its social infrastructure, and its publication venues. It is not peripheral to what the community does. It is the organizing logic through which data reuse becomes meaningful.

The audit framing is the community's own; this is how they describe themselves (rather than how I am choosing to describe them). As introduced in the background, Steve McIntyre's *Climate Audit* website, one of the first and most foundational ACC skeptic websites, took the language and logic of financial auditing directly. Just as a corporation cannot certify its own books, the argument runs, climate science cannot be trusted to check its own work, so an independent party with no stake in the outcome is needed to verify that results follow from the data. McIntyre and McKitrick proposed this audit as a standard higher than peer review, which they characterized as a form of "pal review" (Edwards, 2019). ACC016 framed his project as an audit of a mainstream climate institution, modeled explicitly on the accountability audits of the publishing trade (ACC016, Slides, 2025; see 5.1.3).

Yet the audit orientation has not uniformly produced the results skeptics might have anticipated. Richard Muller, a physicist initially skeptical of the surface temperature record, conducted an independent reanalysis built around extensive data and code transparency. His project, the Berkeley Earth Surface Temperature project, ultimately produced results closely aligned with existing temperature records and contributed to Muller's public acceptance of anthropogenic warming (Rohde et al., 2013; Edwards, 2019). Anthony Watts organized more than 650 volunteers to photograph and assess U.S. weather stations, expecting to demonstrate a substantial warm bias in the temperature record. The resulting analysis confirmed widespread siting problems but found overall temperature trends to be nearly identical to the official record after existing adjustment procedures were applied (Edwards, 2019; Fall et al., 2011). In another case described by Edwards (2010a), secondary analysis by skeptics identified a genuine error in a peer-reviewed study. These examples complicate any straightforward characterization of the community as a uniformly bad actor. Audit practices have sometimes corroborated mainstream

findings, sometimes identified real problems, and occasionally improved the underlying evidence base itself.

The audit orientation is not only an epistemic stance. It is a method of infrastructural inversion, taking the normally invisible work of the mainstream infrastructure and rendering it visible as an object of scrutiny. Carried out at scale, that inversion becomes built into the skeptics' own knowledge infrastructure. The Wayback Machine was used across the community as an institutional surveillance instrument, used to recover removed datasets, compare old and new versions of official pages, and document changes that repositories do not flag. The quality assessment practices documented in the broader corpus reflect the same logic. Surfacestations.org organized hundreds of volunteers to physically survey temperature stations across the United States, documenting siting problems that Watts argued introduced systematic bias into the GHCN dataset. The survey required volunteer coordination infrastructure, data collection protocols, and a publication platform: an audit organized as a citizen science project (Bonney et al., 2009; Irwin, 2002). Comment sections on community blogs functioned as distributed audit infrastructure, generating hundreds of pages of methodological scrutiny directed at both mainstream and community-produced work. Taken together, these are not isolated practices. They constitute infrastructure organized to facilitate audits: a set of components specifically designed to make the official climate knowledge infrastructure interrogable.

The ACC community adopted an audit logic that already saturates contemporary life and how we interact with institutions. In the seminal book, *The Audit Society: Rituals of Verification*, Power (1997) argues that checking is what people do in situations of doubt, conflict, and mistrust. Things only become objects of explicit verification once trust has broken down. The "audit explosion" he documents is the institutionalization of that response with formal verification spreading across the financial, environmental, medical and academic sectors. According to Power, these audits are "rituals of verification" that promise to restore the trust whose very absence prompted them (Power, 1997). Strathern (2000) describes this explosion of audits as an audit culture, where verification protocols designed for financial accountability have overflowed the sector and have become a prerequisite for governance of any kind. The uptake of this logic in open data infrastructures is not coincidental. Open data infrastructure creates the conditions under which audit becomes possible for any motivated community. Edwards (2019, p. 38) makes the connection to the ACC community directly: an "audit culture" arose around

climate science that "sought to open the black boxes of scientific processes" under the banner of transparency and promoting scrutiny from outside the boundaries of the mainstream climate science community. The dissertation documents this audit culture through their data practices and how the audit orientation has been embedded into the infrastructure the community has built.

The term audit, however, is doing different work in different communities, and situating the skeptic audit against two neighboring uses sharpens what is shared, and what is distinctive about it. The first is the data repository community's own auditing practice. Digital repositories submit to audit and certification processes such as ISO 16363 and CoreTrustSeal to earn the trust of their designated communities (Yakel et al., 2013, 2024). These audits are voluntary with the repository initiating the audit, agreeing upon standards, and expecting certification as the outcome. The digital repository trust literature has begun to push back on this arrangement: the certification standards define trustworthiness through policies and objects that can be verified in an audit, prompting Yakel and colleagues to ask, trusted by whom (Yakel et al., 2024). Certification also presumes trust in the auditors themselves, whose selection and judgment shape outcomes yet remain comparatively opaque, leading Frank and Wylie (2023)to call for greater attention to how auditors are chosen and greater transparency in their work. Power (1997) asks whether audit reports "exist to produce certificates of comfort or are they essentially adversarial" (p. 126), and observes that an institutional audit leans toward comfort. Repository certification sits at the comfort pole, which is the basis of Bak's and Moss's critique that certification produces a false sense of security, effective at signaling trustworthiness through formal verification but ineffective at actually moving trustors to trust (Bak, 2016; M. S. Moss, 2011). Repositories should strive for trust through accountability: an inward-focused commitment to being answerable to the communities the repository serves rather to a certification (Yakel et al., 2013). In comparison, the typical skeptic audit inverts many of these features. The skeptic audit is most often imposed rather than invited, conducted by a community the repository never designated, against standards chosen by the auditors. It also sits at Power's adversarial pole; discomfort is the expected product (Power, 1997). Accountability and the audit orientation are not opposites, as they invoke the same ideals such as transparency and responsiveness. The difference is who determines whether these ideals are met and to what end.

The second common audit is the algorithmic audit, which is an investigation of an algorithmic system for behavior that harms the people subject to it. Algorithmic auditing derives

from similar audits in housing and employment discrimination research (Bandy, 2021; Sandvig et al., 2014). Its proponents call for impartial third parties to test opaque systems on the public's behalf (Sandvig et al., 2014), assess algorithms' negative impacts on the rights and interests of stakeholders (Brown et al., 2021), and extend the frame to ecological and community harms (Rakova & Dobbe, 2023). In comparison, the skeptic audit shares the algorithmic audit's posture or vocabulary as independent outsiders scrutinizing a powerful system. However, algorithmic auditors scrutinize closed, proprietary systems, on behalf of people with little power over them. ACC skeptics scrutinize what is likely the most open and most audited scientific KI today, whose data and code are frequently public. This comparison shows that the audit is not tied to fixed political position.

What, then, distinguishes an audit oriented toward improving an infrastructure from one oriented towards discrediting it? Not the vocabulary: all audits espouse transparency, openness, and independent verification, and all claim to correct institutional failures (Power, 1997). The communities that conduct audits hold these values and the audit is the practice through which these values are enacted. How a community deploys its audits, and where it declines to, therefore reveals more about its orientation than the values it professes. Within my data, ACC skeptics rarely applied this audit orientation to their own data practices. The selective application of skepticism documented in Chapter 6 is itself evidence of an audit-oriented community that does not audit itself.

The application of trust or distrust in the ACC community is asymmetrical: who deserves an audit, and who can be trusted? Distrust concentrates on institutions read as ideologically captured by groupthink like NOAA, NASA, the Climatic Research Unit, mainstream journals, and peer review. Trust is extended more readily elsewhere, to skeptic-produced artifacts, such as the UAH satellite temperature series, to sources treated as resistant to revision, including historical newspapers and Wayback-archived webpages, and to intra-community research. What reads as generalized skepticism can be described as a redistribution of epistemic authority; trust tracks perceived political and institutional alignment rather than a uniform evidentiary standard. This is in opposition to the community's self-image as neutral verifiers. The audit posture claims neutrality, but the lines along which trust is granted and withheld is not always neutral.

Not all participants are unaware of this tension. In our interview, Dave Burton reflected on the tendency for people to accept evidence that confirms their existing beliefs, noting that he

had come to recognize the same pattern among those he generally agreed with politically and scientifically:

> People tend to believe what they want to believe. And evidence for a lot of people just isn't really important. And so anyhow, I find I used to think, oh yeah, my side are much more logical than the other side. And then COVID happened and the anti-vaxxers came out of the wall, and a bunch of my friends that I thought were so logical bought into this nonsense, and it's like, aw darn it, it's them too… I really believe that you need to combat that tendency to try that. It's so much more comfortable to not hear things that challenge your preconceptions (Burton, Interview, 2025).

The ACC skeptic audit KI can produce correct findings, real errors found and real gaps filled, but it can persist without them. These correct findings sit alongside a much larger body of audit outputs that mischaracterize ordinary scientific practice, dataset updates, homogenization adjustments, curatorial decisions, as evidence of manipulation, or that treat inevitable small errors as confirmation of systemic bad faith. The audit infrastructure does not require its findings to be correct to persist. It requires them to be legible as findings within the community's own epistemic framework, one constructed to read ambiguity as suspicious.

The community audits through values it shares with the science it scrutinizes: transparency, objectivity, and interoperability. The community holds those values in intensified form, extending each past its ordinary limits. That extension is the subject of the next section.

7.2 Radical scientific values

ACC skeptics do not reject the epistemic values of mainstream science. They embrace a more radical version of them. Transparency, mechanical objectivity (Daston & Galison, 2010), aperspectival objectivity (Daston, 1992), and interoperability are all values that mainstream science itself holds and defends. The ACC skeptic community pushes each of these values to an extreme, where the value begins to turn against the mainstream institutions that also claim to uphold it. This is what makes the community so difficult to argue against from within mainstream science's own framework. As Biagioli and Pottage (2022) argue in a related context, the strategy makes norms absolute and all-encompassing until violations are bound to follow, and until the norm begins to produce effects opposed to its original purpose. These specific values are particularly important because they make the community's audit orientation possible.

7.2.1 Transparency

Transparency is the value most clearly documented in this dissertation as subject to this dynamic. Looking at just two data journeys, both McKitrick and Burton seem genuinely committed to transparency. McKitrick's supplemental documentation is meticulous. And Burton has spent fifteen years building a tool that makes sea level data more accessible, more documented, and more interrogable than the official repositories it draws from. As Burton puts it: "Measurements that are transparently obtained and transparently processed are better than measurements that aren't" (Burton, Interview, 2025). That is a value that is hard to argue against. That genuineness is part of what makes the dynamic difficult to counter. Transparency is not a value unique to ACC skeptics; it appears across knowledge communities with very different political commitments (Cifor & Rawson, 2023; Dillon et al., 2019; Vera et al., 2018). ACC skeptics present their commitment to transparency as a reason their practices should be trusted.

But these journeys also show what happens when transparency is pushed to its extreme. Biagioli and Pottage (2022) describe what they call dark transparency: a deregulatory strategy camouflaged as a defense of science's values of openness and verifiability, radicalizing those norms to the point where they turn against science itself. They developed the concept to analyze the first Trump administration's EPA, where transparency requirements were pushed far enough to exclude legitimate science from policy use. I want to be careful not to map that argument directly onto the ACC skeptic community I observed, since I cannot establish that they are acting strategically in the way the concept implies. But the consequences can be the same. The position that studies with unavailable data should be disqualified from policy formation would, if applied consistently, exclude a significant portion of science from regulatory use, since complete data openness is not always possible because of privacy, proprietary information, and practical constraints. The proliferation of dataset versions across the ecosystem creates a landscape where it becomes genuinely difficult to identify what dataset should be used. That confusion is not intended by the reusers who produce it. But it is not neutral either.

Requests for increased transparency, for access to data, code, and documentation increase curatorial burden regardless of the requester's intent. As Biagioli and Pottage (2022) note in analyzing Climategate, data requests can slow down research and deplete researchers' time and resources whether scientists comply or refuse. Compliance consumes research time, while refusal becomes evidence of subterfuge. The community's widespread use of FOIA requests

operates in the same double bind. Transparency is a value that open science cannot easily argue against. And that is part of what makes its extreme application so consequential, regardless of whether that application is strategic or sincere.

7.2.2 Mechanical objectivity

Mechanical objectivity governs which data can be trusted in this community. Daston and Galison (2010) define mechanical objectivity as a historically specific idea that emerged in the nineteenth century, which holds that reliable scientific knowledge is produced by suppressing human judgment as much as possible, and letting nature speak through data, instruments, and standardized procedure. The ideal is procedural: just as a machine operates without will or preference, the scientist operating under mechanical objectivity functions as a neutral conduit between the natural world and the scientific record, following rules, so that the idiosyncrasies and theoretical commitments of individual researchers are kept out.

ACC skeptics hold this ideal normatively. It shapes what they think science should look like and which data they trust. Observational and raw data are valued precisely because they appear to minimize human mediation; models and processed outputs are suspect because they appear to maximize it. In practice, as documented throughout this dissertation, skeptic data reuse involves extensive interpretive choices. But the ideal still does real work. It frames their critique of mainstream science and what artifacts are perceived as legitimate.

The clearest expression of this ideal is the treatment of minimally processed data as inherently more trustworthy. Processing is read as where bias enters. This is a divergence from the epistemologies of mainstream climate science, which treat preprocessing and calibration as conditions for building global long-term knowledge from local data. Edwards (2019) argues that adjusting for known instrument changes does not distort a dataset but, in his term, "truthifies" it (Edwards, 2019, p. 22). Processing, in mainstream climate science, is what makes the record consistent enough to read at all over time. In other words, processing does not corrupt the record; it is what makes the record legible at all.

In practice, the community's relationship to processing is more complicated than their raw-data ideal suggests. A common strategy was exclusion over correction: ACC017 stripped away adjusted data and worked only with raw records, others dropped urban temperature stations and built rural-only subsets. Removing suspect data is preferable, perhaps, to adding a new

interpretive layer. ACC019, ACC021, and ACC022 were a telling exception. In one project, they did make adjustments, but case by case, with each change justified against physical station-history records. They presented this manual approach as superior to mainstream automated homogenization. If processing is needed, the process should be transparent and auditable. The ideal is therefore not the complete elimination of judgment, but judgment that can be inspected and verified.

This orientation results in a kind of data determinism centered around the belief that data, if collected in what they perceive as legitimate and uncontaminated conditions, will speak for themselves and resolve scientific disputes. Data are treated as the answer. When ACC skeptics encounter traces of social processes in mainstream science artifacts, they read those traces as evidence that their own position is the right one. This framing understates the interpretive work that all data reuse requires. As documented throughout this dissertation, the ACC skeptic community makes extensive choices about which datasets to use, which time periods to select, which processing steps to apply or avoid. The data do not speak for themselves (Gitelman, 2013). The data are made to speak, through choices that are affected by the community's epistemological commitments, including the commitment that unprocessed measurements are more likely to be neutral and that institutional processing is suspect.

7.2.3 Aperspectival objectivity

Within the community, objectivity is valorized in another way, in conjunction with mechanical objectivity. Aperspectival objectivity governs which experts can be trusted. Daston (1992) describes aperspectival objectivity as an aspiration to a view free of any particular standpoint, knowledge that bears no trace of the position from which it was produced. The community applies that aspiration to people. Expertise is valued to the extent that it appears detached from the institutional position being audited. The community valorizes independent, disinterested scrutiny, the outsider's freedom from a stake in the result, and it pushes that value until it redefines which expertise counts. The disciplinary expertise ACC skeptics celebrate is often drawn from outside climate science, and the distance is treated as an advantage rather than a limitation.

This is why the community disparages expert authority while celebrating the credentials of its own members (Pasquetto et al., 2026). ACC031's econometric training was deployed explicitly as audit expertise: he argued that climate scientists lack the statistical training to

recognize the methodological errors he can identify (ACC031, Nelson Podcast, 2022). Physicists, engineers, and geologists recur in the corpus as preferred authorities, valued precisely because their disciplinary distance from climate science is read as freedom from its institutional biases.

This is a redefinition of which expertise counts, one that privileges auditing competence over situated domain knowledge and treats outsider status as a credential rather than a disqualification. Edwards (2019) describes the same inversion in early climate audits, where disciplinary experts were portrayed as government scientists whose institutional position was disqualifying. Climate scientists themselves are not automatically excluded from this trusted circle. In other words, climate science expertise is not rejected completely. It becomes trustworthy when detached from mainstream consensus.

7.2.4 Interoperability

Interoperability is the final value subject to this dynamic. The mainstream climate science knowledge infrastructure maintains separations between datasets for good reasons, such as differences in measurement methodology, spatial and temporal coverage, intended application, and uncertainty characterization. Not all datasets are appropriate to combine. Community-built tools make these datasets interoperable in ways that official infrastructure does not, which appears to serve open access and analytical flexibility. But the interoperability the tools provide can obscure the reasons datasets are kept separate. The tools may be more attractive to users without climate science expertise, who lack the necessary context to recognize why those separations exist or to seek out the institutional and contextualized sources that would supply it. Interoperability pushed beyond what the underlying data can support produces comparisons that look rigorous but are not.

In each case, a scientific value is extended until it begins to work against the function it was meant to serve. Through these values, the ACC skeptic community positions itself as doing science more correctly than the institution it contests. That position is difficult to answer from within a scientific framework that relies on the same values. Additionally, if transparency must be total and objectivity requires data uncontaminated by human judgment, then any acknowledged uncertainty, any revision, any gap in disclosure can be read as a failure of science rather than a normal feature of it. Scholars have named the broader strategy Scientific Certainty Argumentation Methods, or SCAMs: the exploitation of the inherently probabilistic character of

scientific findings to demand absolute proof as a precondition for policy (Freudenburg et al., 2008). Because empirical science cannot deliver certainty, the standard can never be met.

SCAMs cut in two directions at once. Normal expressions of uncertainty, like error bars, confidence intervals, or the spread among models, are reframed as evidence that the underlying figures are unreliable. Uncertainty is leveraged when the institution acknowledges it and weaponized when it does not. The same logic treats revision as disqualifying. Climate4You concludes that "a temperature record which keeps on changing the past hardly can qualify as being correct" (Climate4You, n.d.). But revision is what maintaining a long-term record requires. Edwards (2010, 2019) calls this "shimmering data": as techniques improve, versions are reissued, and no dataset is ever final. What the community reads as instability is the record being kept correct. When the science offers a probabilistic finding, the audit demands a certainty that it cannot provide, the absence of certainty becomes the finding. This is the individual-level counterpart to the organized production of doubt documented at the level of institutions and funding organizations (Oreskes & Conway, 2011).

7.3 Symbiotic knowledge infrastructures

The two knowledge infrastructures this dissertation documents, the mainstream climate science KI and the alternative skeptic KI, overlap, and that overlap produces a relationship better understood as symbiosis than as solely extraction. The symbiosis is by turns parasitic, competitive, and occasionally mutualistic, and mimicry runs through all of it. Infrastructures do not operate in isolation; they overlap. In her book on infrastructures during disasters, Finn (2018) uses the example of sending a picture message to illustrate this point, noting that the act depends not just on telecommunication infrastructure but on a web of supporting systems such as electricity, transportation, and payment processing. Through this digital ethnography, I saw ACC skeptics' data practices supported by multiple infrastructures, including the mainstream climate science KI and an alternative KI they have built themselves. The question I turn to now is not whether KIs overlap but what kind of relationship the overlap produces.

Ecology offers a useful vocabulary. In ecology, symbiosis is the umbrella term for organisms living in close association. Symbiosis names the relationship, not the nature of that relationship. Biological systems interact in ways that are parasitic (one organism benefits at the host's expense), commensalistic (one benefits while the other is unaffected), competitive (both vie for the same resources), and mutualistic (both benefit). I use symbiosis/symbiotic in this

sense throughout the dissertation. The colloquial use, in which symbiosis implies mutual benefit, is not what I mean.

Infrastructures, in Finn's account, have generative power: they do not merely serve existing communities but actively produce new ones (Finn, 2018). The "hypertransparency of open data, open code, commodity software tools, and alternative publication venues" makes this possible (Edwards, 2019, p. 21), allowing skeptics to selectively use infrastructural components where they are feasible and advantageous. The skeptic knowledge infrastructure, understood as entangled rather than simply parasitic, does not only extract from mainstream science. It recruits, trains, and sustains a community that does things together and, as a consequence, believes things together (Pasquetto et al., 2026).

Aspects of this relationship are undoubtedly parasitic: ACC skeptics build an alternative knowledge infrastructure from components of the mainstream climate science knowledge infrastructure they often seek to discredit. They use open data to claim epistemic authority for conclusions that contest the institutions that produced that data. This dependency is not neutral: ACC skeptics rely on the work of mainstream science institutions often without contributing to it, and in some cases actively complicate it by increasing curatorial burden and creating conditions in which data removal and restriction, often mistakenly, become politically legible responses. The skeptic KI exerts pressure back on the mainstream climate science KI, eroding public trust and weakening the conditions under which mainstream science operates (Aklin & Urpelainen, 2014; International Panel on the Information Environment (IPIE) et al., 2025; Lewandowsky et al., 2015; Oreskes & Conway, 2011; Weddig, 2022). The harm is real.

This parasitic framing, however, is too simple to account for everything this research found. Parasitism implies the following relationship: the parasite extracts and benefits, and the host is weakened. The dynamics documented in this dissertation are more complicated. The four data reuse journeys in Chapter 4 alone illustrate at least two other relationship types. Competition is visible in how sealevel.info positions itself relative to NOAA and PSMSL. The tool draws on the same underlying data as official repositories while offering more flexible and interoperable access. Users who find what they need at sealevel.info may never travel to NOAA or PSMSL directly. The two infrastructures are competing for users interested in sea level ice data. Mutualism, rarer but present, appears when skeptic data practices produce outputs that benefit the mainstream KI regardless of intent. McKitrick produced a more complete and openly

archived version of the C14 dataset than the original authors, one that future researchers may use regardless of their views on climate policy. His dataset competes in one sense, but it also extends and completes the record in ways that are genuinely useful beyond the skeptic community. Engelbeen's analyses produce mainstream-aligned conclusions using skeptic methods. This is an instance where the alternative infrastructure's outputs reinforce rather than contest the consensus. Commensalism, where one party benefits and the other is genuinely unaffected, is theoretically possible in these relationships but difficult to document in practice. Almost every apparently neutral use of mainstream climate science KI in this corpus carries potential for harm given the adversarial context in which it occurs. The absence of commensalism suggests that in a highly contested epistemic domain, genuinely neutral infrastructure overlap may be rare.

Mimicry threads through all these relationships. Even the competitive and mutualistic moments depend on the skeptic KI's ability to present itself as doing something recognizably scientific, rigorous, and trustworthy. The skeptic KI sustains itself and claims legitimacy by adopting the conventions of the infrastructure it depends on. This mimicry is visible several levels of the alternative infrastructure documented in this dissertation. The journal, *Science of Climate Change*, adopts the formatting conventions of peer-reviewed publishing, editorial boards, volume numbers, article types, and open access declarations while positioning itself explicitly as an alternative to mainstream gatekeeping. Community-built tools adopt the visual and functional language of official repository interfaces, presenting data in formats that signal institutional credibility. Transparency rhetoric, documentation, and reproducibility standards are deployed in ways that mirror the epistemic values of mainstream science, lending credibility to analyses that contest its conclusions. The mimicry does not need to be perfect to be effective. It needs only to be convincing enough to recruit community members, claim epistemic authority, and sustain the alternative infrastructure's legitimacy in the eyes of those already inclined to question mainstream science.

ACC skeptic data practices can and do harm the mainstream climate science KI. But harm is not the only thing happening, and designing open data systems capable of navigating these dynamics requires a more precise account of what those dynamics are. Overlapping KIs produce symbioses, relationships that are parasitic, competitive, and occasionally mutualistic. As we open more components of knowledge infrastructures to the public, alternative knowledge infrastructures are more likely to arise. Those who study the sustainability and maintenance of

knowledge infrastructures should take the interplay between mainstream knowledge infrastructures and skeptics' knowledge infrastructures seriously, since the latter shapes the sustainability and credibility of the former. Designing open infrastructures that can navigate these relationships wisely, rather than assuming that openness produces only beneficial reuse, is one of the more consequential challenges this landscape presents.

7.4 Data misuse as infrastructural friction

Chapter 2 introduced data misuse as a designation applied from a position rather than an objective category and defined it as a site of infrastructural friction. The findings of this study allow that position to be developed into a contribution. Rather than treating data misuse as a peripheral risk, it is more productive to understand it as a site where power, values, and epistemologies collide and are revealed, and therefore as an important lens for thinking about how infrastructures adapt and change. As documented in Chapter 6, misuse accusations most often arise at the boundaries between different data communities, where things like standards of disclosure, methodologies, and expectations of how data should travel may differ and come into conflict. No universal definition of misuse would resolve these collisions. What counts as misuse also shifts as communities, technologies, and social conditions evolve. Misuse is rarely about a simple technical error. It is always contextual: what counts as improper reuse in one domain might be routine in another, and what counts as harmful at one point in time might not have been recognized as harm decades earlier (Pasquetto et al., 2024).

This dissertation did not take a harm-centered approach to defining or identifying data misuse. The starting point for that decision is a definition my colleagues and I proposed in prior work: misuse as "any kind of data use that does not reflect on and take responsibility for the potential harm caused by such use" (Pasquetto et al., 2024, p. 11). This definition deliberately foregrounds harm rather than intent, allowing misuse to be assessed contextually.

The difficulty is that harm is itself contested. What one community identifies as harm, another may not recognize as such. Consider anti-mask organizing during the COVID-19 pandemic. Lee et al. (2021) found that anti-maskers used "orthodox visualization techniques" on open government data to support their "unorthodox beliefs" that mask mandates were ineffective or harmful. Public health authorities regarded that reuse as causing harm, through increased infections and deaths; its practitioners understood it as averting harm. When the same reuse is

harm-causing from one position and harm-averting from another, harm becomes a difficult foundation for a researcher attempting to define misuse across communities.

A harm-centered approach remains possible, and for some actors it remains the right one. A repository deciding whether to restrict a dataset, respond to a contested reuse, or intervene in a dispute cannot avoid making a judgment, and harm is a defensible basis for making it. What this dissertation adds is a condition: the judgment must be made from a situated and explicit position, that is, from a standpoint the repository names as its own rather than treats as universal. Harm is typically defined from one side of the friction. When a repository adopts the harm definition of the community whose data it stewards, it is taking a side, and that is a legitimate thing for a repository to do. In a hypothetical example, a repository hosts state COVID case and mask-mandate data. Anti-maskers download it, chart cases against mandate timing, and argue the mandates did not work. Each side names a harm. Public health researchers see the reuse itself as harmful, misrepresenting the data in ways that erode support for a public health measure. The anti-maskers see the mandates as the harm, and the reuse as exposing it. The repository has to decide whether to respond, and it cannot do so from a position outside this disagreement. If it acts, it is applying public health's standard for what the data can support, not a neutral standard and not the anti-maskers'. My condition is only that it say so: that it make explicit what it counts as harm and misuse here and why, rather than present the judgment as neutral stewardship. The contribution is not that a harm definition of reuse should be abandoned but that its situatedness should be made explicit as I mention in the practical implications of 7.5.3.

For the analysis in this dissertation, however, adjudication was not the task, and this is where friction is useful. The friction literature, discussed in Chapter 2, largely examines communities attempting to establish common ground. The communities in this study do not always do so. One produces and stewards the data; the other audits them. In this setting, friction is not a stage on the way to interoperation. It is the durable condition of the relationship, and further disclosure does not resolve it but supplies more material to audit. McKitrick’s reading of C14 illustrates the structure. Both papers disclose the scope of their data use; McKitrick applies a standard of disclosure under which those disclosures are insufficient, particularly for findings that informed federal policy. Neither party fails to reflect on harm. They hold incompatible standards for what an adequate account of how data were made requires.

Treating misuse as infrastructural friction, then, means treating accusations as the analytic object. Accusations are observable, they run in both directions, and they mark where the infrastructure's seams are. Skeptics charge mainstream scientists and repositories with manipulating, selectively reporting, and misrepresenting climate data. Mainstream scientists and curators charge skeptic reuse with cherry-picking, misrepresentation, and reputational harm. Documenting both does not require ruling on either. It requires showing what the collision reveals: which norms are in contact, which values are being enacted, and which parts of the infrastructure made the dispute possible.

Two qualifications keep this from collapsing into relativism. First, misuse and harm can still be named. Where a practice is recognized as misuse both by mainstream climate science and by the community's own stated standards, that recognition is documented, cherry-picking temporal endpoints meets this condition, and the ACC community's own commenters say so. Second, treating accusations as objects of analysis does not make their consequences symmetrical. For instance, publishing a named curator's private correspondence to an antagonistic audience carries risks for the curator that the accusation carries for no one else.

What this reframing offers is a shift in where analytic attention goes. A harm-centered account asks whether a given reuse caused harm, which requires a position from which to answer, and repositories that occupy such a position can and should use it, explicitly. An account centered on friction asks what the dispute reveals about the system in which it occurs.

7.5 Practical implications for data intermediaries and repositories

The dynamics documented in this dissertation are not only theoretically significant. They have concrete implications for how open data infrastructures are designed, governed, and sustained. ACC skeptics are not anomalous users. They are one instance of a broader phenomenon: communities engaging with open data infrastructure in ways its designers did not anticipate and may not endorse. This section draws on the dissertation's findings to propose implications for data intermediaries and repositories.

Before turning to these recommendations, a caveat about the conditions under which any of them would be implemented. The recommendations in this chapter are addressed to repositories and data intermediaries that are, at the time of writing, under acute strain. Public environmental data infrastructures in the United States face funding withdrawal, staff reductions, and in some cases the removal of datasets and documentation (J. Lloyd, 2026; OOI, 2026;

Santarsiero, 2025; US EPA, 2025). Repository managers are being asked to "think like a business" and pivot their business models (Gregory & Strecker, 2026). Under these difficult conditions, my recommendations can read as indifferent to what repositories are going through, and what they can realistically do.

I, therefore, want to be especially honest about challenges to these recommendations. Most of what I recommend in 7.5.2, and even some of 7.5.3, are not conceptually difficult or novel. Persistent URLs, documented dataset changes, explicit statements of what data can and cannot support, staff with time to answer questions: these are not novel interventions. They are unimplemented because they are unfunded. The barrier is often resources, and no amount of design recommendations can substitute for staff and stable budgets. And the material conditions of data repositories are not incidental to this dissertation's topic because they are partially produced by the same political dynamics members of the ACC skeptic community participate in. Sustained public skepticism toward climate science contributes to a political environment in which funding for environmental data infrastructure becomes more contestable. Some of the ACC community's complaints, like curatorial deficits, unexplained dataset changes, or broken links, are a result of underfunding that is partly caused by the political project some in the community advance. Therefore, what follows in this chapter is not a prescriptive set of obligations, but rather an account of what a well-resourced repository could do while acknowledging that many are not.

7.5.1 Reframing open data governance

The open data movement rests on assumptions this dissertation complicates. These three assumptions are based on Leonelli's distinction between openness as sharing and openness as judicious connection (Leonelli, 2023). The first is that openness is a sufficient condition for reuse, that making data available is the primary obligation of a repository, and that what happens afterward is the reuser's responsibility. Prior research has continuously shown that openness is necessary but not sufficient for data reuse, and the same can be said about data curation; good data curation is often not sufficient for data reuse (Borgman & Groth, 2025; Edwards et al., 2011; Leonelli & Tempini, 2020; Pasquetto et al., 2019; Thomer et al., 2022; Wallis et al., 2013; Wu et al., 2024). As Leonelli (2023) argues, having access to data is not conducive to knowledge production unless the right skills, infrastructures, governance, and administrative support are available to foster reuse, which requires a degree of understanding and trust among the parties

involved. The ACC skeptic community has the skills and support to reuse data meaningfully. Its members obtain the data, run the analyses, and often produce technically competent work. The missing ingredient is not access or capability. It is the trust and shared interpretive context Leonelli identifies, and no amount of additional access supplies it. Some research has examined whether repositories should bear responsibility for what reusers do with their data (Cragin et al., 2010; Frank et al., 2015; Leonelli, 2016; O'Brien et al., 2024; Pasquetto et al., 2024; Verhulst, 2021). Where repositories do take responsibility, recommendations cluster around several interventions: proactive risk assessment and responsibility frameworks, restricted access and data enclaves for sensitive datasets, participative and reflexive governance of data practices, community engagement around the cultural and political contexts of data use, and greater transparency in how data handling decisions are made. These recommendations share an acknowledgment, however, that repositories cannot resolve all issues of downstream misuse — they can establish standards, document intended use, and direct communities to appropriate resources, but they cannot control what happens once data leave their systems.

The second assumption of the open data movement is that data are neutral, that the data along with basic metadata will speak for themselves. This position shades into what is called information determinism: the assumption that political and social problems stem from insufficient data, and that more or better data are always the remedy (Srinivasan et al., 2017). Information determinism is a "sort of category error in that it attributes agency to data which, by [themselves], [they do] not possess" (Soden et al., 2025, p. 2). The sharing of research objects should not be construed as a way to transform situated human artifacts into "neutral products that can be interpreted irrespective of their provenance and the ways in which they have been processed" (Leonelli, 2023, pp. 52–53). Although this position is naive, it is all too common. We can recognize that all models of the environment are abstractions and although it may surprise some ACC skeptics, that includes the most unprocessed data. It is the abstraction that makes the models useful. As Soden and colleagues note, Borges' cartographers in "On Exactitude in Science," illustrate the point: a map drawn at 1:1 scale, identical in size to the territory it represents, is useless (Borges, 1946 as cited in Soden et al. 2025). However, every act of abstraction involves choices about what to include, what to leave out, and how to represent what remains. Those choices reflect priorities, values, and social position, whether the person making them recognizes it or not. The ACC skeptic community recognizes that data are not neutral; they

treat curatorial decisions—which stations to include, which adjustments to apply, which version to publish—as choices that reflect the priorities of whoever made them. This is, in my opinion, an accurate description of how data are made. However, where I read non-neutrality as an ordinary feature of situated knowledge production, the community reads it as errors or even evidence of intent to deceive. The assumption that data speak for themselves is not only wrong, but it also leaves repositories poorly equipped to respond when a community identifies that non-neutrality and interprets it as fraud. The validity, relevance, and significance of research objects must be reassessed as science and the world change. They are made meaningful through situated human activity, and those activities carry values, epistemologies, and political stakes that no documentation standard can fully anticipate. What contextual, non-neutral engagement looks like in practice is illustrated by Leonelli's account of a researcher consulting an online database: rather than simply using it as a neutral data source, the researcher takes time to familiarize themselves with how the database is set up and the conditions under which the data were produced, discovering which assumptions underpin the data and whether those assumptions make sense in a new context of use (Leonelli, 2023, pp. 57–58). Scholars have called for approaches that attend more carefully to the decisions shaping environmental data, but these remain underdeveloped (Muller & Strohmayer, 2022; Soden et al., 2025, p. 2; Soden & Kauffman, 2019).

The third assumption of the open data movement is that the primary threat to open data is restriction. In other words that maximizing findability, accessibility, interoperability, and reusability is an unqualified good, and that barriers to access are the main obstacle to legitimate reuse. The FAIR principles have become the dominant framework for open research data management (Wilkinson et al., 2016), and they are frequently read as an endorsement of this assumption. FAIR encodes genuine and valuable commitments while also including the need for rich metadata, clear provenance, and usage licenses. My concern is not with FAIR as written but with FAIR as implemented, and with the broad advocacy that treats maximal openness as self-evidently good.

The ACC skeptic case shows what happens when the accessibility commitments are applied without the contextual infrastructure that gives data meaning in its original community. Interoperability without context enables exactly the cross-dataset comparisons that obscure why datasets are kept separate. Reusability without community norms enables reuse that the

producing community would not recognize as legitimate. Alternatives do exist. Governance frameworks built by and for specific communities show what that nuance in openness can look like in practice. The CARE principles, Collective Benefit, Authority to Control, Responsibility, and Ethics, were developed explicitly by the authors as a complement to FAIR, centering Indigenous communities that require nuanced and informed openness rather than more of it (Carroll et al., 2020). CARE reflects a broader recognition that open data governance cannot be reduced to maximizing access.

The governance gaps this dissertation documents are not between open and closed. It is between data that are open and data that are open with the contextual, methodological, and communal infrastructure needed to sustain their integrity across reuse contexts (Borgman & Groth, 2025; Pasquetto et al., 2019). If we design repositories assuming that reuse is primarily scientific, we may be designing for a reuse pattern that does not reflect the full reality of who is using the data and why. Data will inevitably be reused in ways repositories do not anticipate and may not agree with. This will be difficult, if not impossible, to avoid through any act of design or curation alone. Repositories should be designed for a broader range of reuse. Different user groups bring different expertise, goals, and relationships to the data. Data and tools, even if not publicized, will be found by non-experts. Not every reuse case can be predicted, but designing with diversity in mind is possible.

These three assumptions, that openness is sufficient, that data are neutral, and that restriction is the primary threat, are not independent. They form a governing logic of open data that this dissertation's findings complicate. Each fails by treating data as stable, fungible objects rather than as artifacts whose meanings are produced and sustained through infrastructure. Open infrastructures differ from other types of infrastructure precisely in this respect: openness introduces interpretive pluralism. This means more actors, more epistemologies, more potential for both misuse and innovation. Their sustainability depends more on trust, documentation, governance, and community norms than do proprietary systems, and misuse exposes their inner workings, values, and vulnerabilities in ways closed systems often mask. The ACC skeptic case illustrates what this looks like in practice. Data often leave a repository stripped of the curatorial context, methodological caveats, and community norms that gave them meaning in their original community. What travels is the dataset. What stays behind is the infrastructure of interpretation.

As Leonelli argues, treating shareability as the main problem, and interpretation as someone else's, is appealing in a world of distributed expertise, but the trustworthiness of these platforms still depends on a collective willingness to keep scrutinizing whether they are fit for purpose and whether they support epistemic diversity and justice (Leonelli, 2023, p. 58). These dynamics cannot all be designed away. Some can be mitigated through better documentation, more careful interface design, and more reflexive governance. Others are structural, built into the conditions of openness itself, and no intervention will resolve them. But understanding them more precisely is a prerequisite for responding to them more wisely, which is what the following sections attempt.

7.5.2 Designing for trust across differences

Trust in a repository cannot be manufactured through certification alone (Bak, 2016; M. S. Moss, 2011; Yakel et al., 2024; Yoon, 2014). Bak and Moss have both criticized audit culture as creating a false sense of security where mechanisms that are meant to signal trustworthiness are ineffective at actually moving trustors to trust (Bak, 2016; M. S. Moss, 2011). Trust is relational and situated, built or eroded through specific interactions between reusers and infrastructure over time (Yakel et al., 2024; Yoon, 2014). Repositories have often behaved as though demonstrating the integrity of their preservation processes will automatically produce trustors' faith in their institutions, but that trust is no longer automatic, if it ever was (Yakel et al., 2024). This dissertation documents a community that approaches official repositories with deep, structurally motivated distrust. This distrust is shaped both by the community's epistemological commitments, which make institutional authority itself a source of suspicion, and by infrastructural decisions that provide grounds for critique.

This is not unique to ACC skeptics. Communities in cities have expressed distrust in data collection and representation (Yoon & Copeland, 2020). Indigenous communities have expressed deep distrust of archival institutions, viewing them as repositories of materials produced through surveillance and oppression (McKemmish et al., 2011). Academics have a long history of distrusting the archives they depend on (Force & Wiles, 2021; K. E. Warren, 2016). Previous studies have linked data reusers' trust in repositories with organizational attributes, past experiences, repository processes such as documentation and quality checking, and users' perceptions of repository roles (Yoon, 2014). ACC skeptics invoke these same factors when discussing their distrust. Some research suggests that thorough and transparent documentation

enhances trust and reduces violations (Yoon, 2017). However, even with improved documentation, it is unlikely ACC skeptics would stop critiquing repositories, but it may give them less fodder for skepticism, and changes what a repository can point to when responding. Yoon (2017) notes, "trust is a continuum: it is gained, lost, and sometimes regained" (Yoon, 2017 as cited in Yakel et al., 2024, p. 14).

Many of the following recommendations are grounded in interest convergence, leaning into shared values (Jonas & Hanrahan, 2022). ACC skeptics, mainstream scientists, and other data reusers share more common ground in what they want from open data infrastructures than is often acknowledged. In times of institutional and political uncertainty, prioritizing those shared values and working within existing systems where possible is both pragmatic and principled. NASA's continued work on broadening access to Earth data illustrates this, even without an explicit environmental justice focus, progress on shared infrastructural values remains possible within existing systems. First, invest in basic website maintenance: persistent URLs, proper redirects, and stable archiving. These are low-cost decisions with high-stakes consequences. The community's use of the Wayback Machine to recover removed data is not only a skeptic practice — it is a symptom of infrastructural fragility that affects all reusers. Second, document data changes clearly. Unexplained updates feed exactly the kind of distrust that skeptical narratives exploit. At the same time, avoid the dark transparency trap: more versions and more visibility are not always better. Burton's proliferation of dataset versions illustrates how radical transparency can itself become a source of confusion. Be explicit about which dataset is authoritative and why. Third, document not just what data are but what they can and cannot support analytically. Concrete examples of appropriate and inappropriate reuse give curators a reference standard when misuse occurs, give good-faith reusers the context they need, and may redirect skeptical reusers toward critique that is constructive rather than corrosive. These are not accommodations to skeptics but basic infrastructure investments that benefit every reuser. What ACC skeptics say they want from repositories is worth taking seriously, not because their analytical conclusions are always correct but because their usability complaints are often legitimate.

Given that trust is not a binary condition but an evolving process, repositories should develop trust across multiple dimensions: transparency, usability, and reuser engagement. In data repositories, trust in data, trust in repository staff, and trust in institutional processes form an inextricable web (Yakel et al., 2024). This involves providing clear data provenance, facilitating

communication between data producers, intermediaries, and reusers, and offering mechanisms for user feedback (Lin et al., 2020; Yakel et al., 2024). Repositories must move beyond one-size-fits-all models of trust. Different reusers have different trust requirements which require different interventions. Epistemic empathy is a useful frame: interactions with communities skeptical of institutional authority require staff who can engage with distrust without dismissing it or capitulating to it, who can acknowledge legitimate usability concerns while being clear about the limits of what repositories can and should do (Eadon, 2020). This means not enforcing dominant epistemologies but facilitating engagement and acknowledging alternative knowledge systems.

Data intermediaries offer one mechanism for implementing these recommendations without requiring repositories to take explicit positions on contested scientific questions (Yoon et al., 2018). Data intermediaries are agents positioned between two other agents in a data supply-demand chain who facilitate the reuse of data (Schalkwyk et al., 2015; Yoon et al., 2018). Positioned between data producers and reusers, they can help skeptical reusers understand the limits of what data support, connect them to documentation they may not have found, and surface usability problems that repositories can address. Intermediaries are also well positioned to guide skeptical communities toward more appropriate data reuse practices and more constructive forms of critique, not by policing what conclusions reusers reach, but by providing the contextual infrastructure that makes rigorous engagement with data more likely.

7.5.3 Designing for and responding to potential data misuse[5]

Data misuse cannot be fully prevented through design, but it can be mitigated and used as impetus for change. As Section 7.4 argued, data misuse accusations arise where communities holding different norms draw on the same data, and no universal definition of data misuse would resolve those collisions. The recommendations that follow work with that condition rather than against it. Not all contested data reuse is equally nebulous. As documented in Chapter 6, some practices are recognized as problematic by both the mainstream climate science community and the ACC community's own stated standards, such as cherry-picking temporal endpoints, or substituting local measurements for global patterns. These are more amenable to design intervention because they do not require the repository to take a side. For instance, a tool that

[5] Portions of this section, including the interventions listed below, are adapted from Pasquetto et al. (2024). Because this work was collaborative and I was not the lead author, I identify passages drawn from it in this section.

defaults to displaying full trends rather than user-selected endpoints enforces a standard both communities already claim to hold. Design in these cases can operationalize shared norms rather than adjudicating contested ones. Others involve more complicated norm clashes where communities disagree, and where design can clarify conditions without necessarily adjudicating outcomes. The recommendations that follow work across both categories, but with different expectations for each.

Several interventions for data misuse are possible and worth prioritizing, adapted from Pasquetto et al. (2024):

1. Invest in professional data curation: staff who can carefully document data provenance, formats, granularity, and quality following established best practices and disciplinary standards (Pasquetto et al., 2024).
2. Consult misinformation experts, fact-checkers, and professional science communicators to develop a close understanding of potential reuser communities. This helps in designing effective communication strategies before and after misuse events (Pasquetto et al., 2024).
3. Adopt different forms of data description beyond standard metadata. Data papers, README files, and in some cases narrative documentation give reusers the contextual information they need to understand what data represent and how they were collected (Pasquetto et al., 2024).
4. Encourage self-reflexive reuse practices, what Leonelli et al. (2021) call methodological data fairness, by providing concrete examples of what such practices look like for a given dataset (Pasquetto et al., 2024). Dataset landing pages can display links to papers that use the dataset appropriately, providing reference examples of proper use. Documentation can include explicit statements of what analytical questions the data can and cannot answer.
5. Work with communities affected by data harm to decide what values and research questions should drive reuse, ensuring that those most affected by harmful reuse have a voice in how data are made available and for what purposes (Pasquetto et al., 2024).
6. Before releasing a dataset publicly, engage in threat modeling: brainstorm potential misuse cases with diverse stakeholders before the data become available (Pasquetto et al., 2024). Anticipating how data could be misused is more effective than responding after misuse has already occurred. Not every case can be predicted, but structured pre-release consultation is more likely to surface risks than waiting for controversy to emerge.

7. Be explicit about the temporal dimensions of data and their appropriate use. Document changes in the repository's own understanding of what constitutes misuse and maintain ongoing conversation with reuser communities about new and emerging ways data reuse can cause harm.
8. Where tools afford misuse, through flexibility that makes cherry-picking easy or interfaces that strip contextual caveats, design choices should weigh that affordance against the benefits of flexibility, sometimes trading flexibility for integrity.

These interventions will not prevent all contested reuse, nor are they meant to police users. They set reference standards that curators and other reusers can invoke when potential misuse occurs as evidence of what data represent and what they do not, recorded before the controversy arose. They may also redirect good-faith reusers in the ACC skeptic data community toward constructive critique.

Contested reuse is not only a tension or threat; it is also a driver of infrastructural adaptation. Across domains, concerns about various types of misuse and the harms it can cause have prompted the development of new governance mechanisms designed to preserve openness while mitigating harm. These include things like improved metadata documentation, guidelines for appropriate reuse, and mechanisms for identifying and addressing misuse (Pasquetto et al., 2024). Some repositories now employ prebunking and debunking strategies, such as labeling datasets with recommended interpretative guidelines or linking to publications that demonstrate appropriate use (Pasquetto et al., 2024). These responses do not eliminate misuse wholesale; rather, they reconfigure infrastructures to make certain forms of misuse less prevalent or more visible, accountable, and contestable. For instance, in conservation biology, researchers have debated how making species-location data openly available may enable poaching, lead to misinterpretation by government agencies, or introduce less-understood geoprivacy harms (Kaehrle & Eschenfelder, 2025a). While there are still improvements to be made, these tensions and potential harms have shaped major design decisions in platforms like iNaturalist, which deliberately introduced location uncertainty to mitigate harm. Here, potential misuse directly drives infrastructural change.

These adaptations reveal that misuse is not only addressed at the level of individual behavior, but through structural interventions that reshape how data can be used, interpreted, and evaluated. These mechanisms redistribute responsibility across researchers, reviewers,

repositories, and platforms. They also surface underlying epistemic values that had previously been implicit or unevenly enforced. Rather than treating misuse as an anomaly to be eliminated, these interventions acknowledge it as a predictable consequence of openness and respond by redesigning infrastructures to support transparency, and accountability. In doing so, misuse serves as a moment of infrastructural inversion, exposing where existing arrangements fail to align practices, values, and expectations.

When potential misuse events arise, a structured set of questions can shift the focus from solely policing users to understanding the structural conditions that allowed the situation to emerge. These questions foreground what infrastructures often struggle to surface: who gets to interpret data, whose epistemology defines proper use, and who bears the harm when interpretations diverge.

1. Should this be defined as misuse at all? Before making that determination, identify which communities are in contact and where the friction points are. The question is less whether a violation occurred and more which norms are clashing and why. If harm is the basis for the determination, name whose definition of harm is being applied.
2. Which parts of the infrastructure facilitated it: the documentation, a lack of provenance, an overly flexible interface, or an assumption that all users share the same norms?
3. Who has the authority to make the determination that data misuse occurred, and is that process transparent? Misuse determinations concentrate power, and the governance of those decisions matters. Multi-stakeholder governance, involving curators, reusers, and communities affected by data harm, is essential for making these determinations legitimate and durable.
4. How should the infrastructure adapt? Misuse is not simply an error to prevent. It is a signal about where the infrastructure needs more safeguards, more context, or new forms of community engagement.
5. What are the resource implications of responding? Many interventions require funding that institutional structures rarely support.
6. How should the repository communicate with the reuser community after a contested reuse event? The response shapes how the community understands the data and the repository's role going forward. Platforms that govern contested content, such as community-moderated forums, offer design models worth examining here. Strengthening provenance,

documentation, and institutional support also helps protect individual curators facing reputational harm.

7. Is this a one-time event or a pattern? Single instances and systematic practices from individual actors may warrant different responses.
8. What precedent does this response set? Responses to misuse shape future reuser behavior and repository culture in ways that extend far beyond the immediate event.

Repositories and curators are not in a position to act as arbiters of truth. But they are not neutral either. When a repository names a reuse as misuse, it does so from its own position, in the sense developed in Section 7.4, and the recommendations above are designed to make that position explicit and defensible rather than to disguise it as neutrality. Curatorial policies shape not just data reuse but public perceptions of scholarship. Acknowledging that is not a retreat from the commitment to openness; it is a more honest account of what openness involves and leads to. Focusing on misuse is not a slippery slope toward restricting data access. Safeguards are not blatant restrictions; most of them preserve openness while adding the context and governance that make openness sustainable. Infrastructures must support reflexive reuse, encouraging users to engage with uncertainty, bias, and contextual limits, and community-engaged governance is essential, especially for those affected by data harm (Pasquetto et al., 2024). Repositories that have clearly documented what their data represent and what the data cannot support are better positioned to respond to misuse publicly, to support curators facing reputational harm, and to maintain the trust of the broader scientific community.

7.5.4 The role of information professionals with skeptic communities

Another question this dissertation raises, and these recommendations make more pertinent, for information professionals is one it cannot fully answer: what role, if any, should repositories and data intermediaries play with communities actively using open data infrastructure to contest the science that infrastructure supports? Leonelli argues that "the idea of openness is quintessentially linked to that of learning as going beyond one's boundaries" (Leonelli, 2023, p. 54). If we take that to be true, then engagement with communities that actively challenge the boundaries of mainstream science is inevitable and, perhaps, what open science values ask of them. The more important question is what that engagement should look like, and where its limits are.

This is a genuine question rather than any sort of settled recommendation. ACC skeptics are using open data infrastructures whether the field engages with them or not. The question is how the field responds. One position is that the field's obligation is to provide well-documented, accessible data to all comers, and that how it is used is ultimately the reuser's responsibility. Another is that silence in the face of documented misuse is itself a policy choice, and one with consequences for how scholarship is perceived and how public trust in data is sustained. The tension between these positions cannot be resolved by appeal to neutrality. Open science projects cannot be universally beneficial in their repercussions, like any other social transformation it is "necessarily divisive," and value judgments are unavoidable in building open research infrastructures, however inclusive the technology promises to be (Leonelli, 2023, p. 63). Repositories must make decisions on what they choose to make open and to whom. Silence itself is a policy choice.

> Open data practitioners must make "choices among possible visions of who can and should participate in research efforts, whose perspectives can and should be voiced and whose ideas of research could and should be supported…There is always a decision made around who will benefit and who may lose out from specific research initiatives, and the more explicit and reflexive that choice is, the better" (Leonelli, 2023, p. 59).

This points toward a different kind of response than simply managing or restricting skeptical reuse. If choices about who benefits are unavoidable, one answer is to be more deliberate about which reuse communities repositories actively invest in cultivating. Rather than centering resources on responding to ACC skeptics, repositories might direct energy toward building the capacity of communities whose data reuse aligns more closely with the values open data infrastructure was built to serve, environmental justice advocates, community scientists, public health researchers, and others who are often underserved by existing infrastructures. This is consistent with the recommendations in 7.5.2, which were general recommendations that would serve all reusers, including skeptical reusers. The question here is where to direct scarcer resources, including relational investment, and the answer is toward communities where engagement can be reciprocal. This is not a way of avoiding the harder question. It is a reframing of it: the question is not only how to respond to adversarial reuse but which communities to proactively support.

The harder cases are those where the audit orientation is deployed not to improve infrastructure but to discredit it, and where the harm is epistemic, political, and temporal. These cases require institutional courage as well as design ingenuity. The interventions documented in 7.5.3 matter most precisely here, clear documentation, established curation standards, and pre-release threat modeling give repositories the infrastructure through which a response becomes possible when things get harder. That is not a guarantee against data misuse. It is the foundation on which a response can be built.

Chapter 8 Conclusion

This dissertation began with a tension at the heart of the open data movement: the same openness that promises to broaden participation in science also places scientific data in the hands of communities whose conclusions diverge from those of the scientists who produced the data and the repositories that steward them. ACC skeptics are one such community. Across the preceding chapters, I traced how this technically and rhetorically effective community, working at the margins of mainstream climate science, reuses open climate data and builds a knowledge infrastructure of its own. What emerged was not a portrait of incompetent or merely oppositional reusers but of a community organized around an audit orientation, entangled with the mainstream climate science infrastructure it contests, and committed to an extreme version of the very scientific values that the mainstream climate science KI also relies upon. This chapter draws those threads together, reflects on what they do and do not show, acknowledges the study's limitations, and points toward the questions it leaves open.

8.1 Final Thoughts

ACC skeptic data reuse is organized around an audit orientation: community members treat open climate data not as a settled source but as material for scrutinizing the data, claims, and interpretive decisions of mainstream climate science. This orientation runs through the stages of data reuse I examined and is also embedded in the knowledge infrastructure the community builds and maintains. This alternative knowledge infrastructure does not stand apart from the mainstream one but relies on it. The relationship between the two knowledge infrastructures is one of symbiosis in the ecological sense, by turns parasitic, competitive, and occasionally mutualistic. And the community's audit orientation rests on values mainstream science shares – transparency, objectivity, interoperability – but extends them until they begin to work against the institutions that uphold them. The dissertation also reframes accusations of data misuse as a site of infrastructural friction. Harm-centered accounts of data misuse still hold, but only from a situated position, never a neutral one.

Taken together, these findings complicate the typical image of the ACC skeptic as simply anti-science. As my collaborators and I have argued, members of these communities, and others like them, are better understood not as rejecting science outright, but as asking for more of it.

> …[M]embers of these communities are deeply committed to "understanding more" about the phenomena they are investigating. They are not anti-science; they are, in a sense, asking for "more science." It remains unclear whether and under what conditions opponents would be willing to change their beliefs if such additional science would end up contradicting their core claims, as climate change scientists did when they corrected errors identified by climate change skeptics in their models (Edwards, 2010). (Pasquetto et al., 2026, p. 12).

This is not to claim that ACC skeptics' conclusions are correct, nor that their actions are harmless. Chapter 7 showed how their extreme application of scientific values can obstruct the science they claim to defend. Members of a community can be genuinely committed to a version of scientific values and, by pressing those values past their limits, still produce effects that obstruct science. This dissertation shows that both can coexist. This account does not address bad actors or organizations that perform these values to purposefully obstruct science or avoid regulation.

I want to be careful about my own position in this account. In the discussion, I was critical of the community, and I stand by those critiques. But every community has flaws. I have tried to portray ACC skeptics realistically rather than as caricatures, neither dismissing them nor endorsing their conclusions. At the same time, my reading of this community is shaped by my own standpoint, and my practical recommendations especially reflect it. What I have framed as friction, as misuse, or as a value pushed too far are interpretations made from a particular position. A reader or researcher situated differently might weigh the same evidence I saw differently. I offer this account as a careful but situated interpretation: not a view from nowhere, and not an attempt at aperspectival objectivity.

The entanglement I have described between the two KIs can lead to a kind of coevolution, where the mainstream climate science KI reacts and changes in response to the pressure of the alternative. I did not study how mainstream climate science reacts to skeptical scrutiny. Luckily, others have. Ramírez-i-Ollé (2019) documented one such consequence in what she calls inside-out skepticism, in which some climate scientists internalize the critical

perspective of skeptical auditors during their own analysis, rationalizing each methodological step rather than black-boxing it and moving toward more transparent and statistically rigorous standards. The pressure of being audited has, in some areas, pushed climate science toward more explicit documentation, more defensible standards, and more reflexivity.

That possibility points toward a different way of thinking about openness itself. Rather than treating openness as the mere availability of data, Leonelli (2023) proposes a process-oriented philosophy of openness, one understood through "judicious connections among researchers," connections grounded in "relationality and trust, among widely diverse individuals and groups" oriented toward collective agency rather than the transfer of objects alone (Leonelli, 2023, p. 43).

In some ways, though, the ACC skeptic case shows the inverse of this vision. It shows what happens when data travel without the relationships, documentation, and trust that gave them meaning in their original community. The dissertation makes the case, by negative example, for an open science perhaps worth building. Those who build and steward open data infrastructures must decide for themselves whether investing in relationships with skeptics is worth it, or whether their focus belongs elsewhere. I leave that decision up to them. I just suggest they make the choice consciously.

8.2 Limitations

This study has limitations, several of which I named in earlier chapters and which also motivate the work that follows. The most consequential concern method and scope.

Member checking was targeted rather than comprehensive. I returned the factual accounts of the Chapter 4 data journeys to the reusers they describe, confirming the accuracy of what they did rather than how I interpreted it, and shared quoted passages with one interviewee who requested them. I did not return findings or analysis to the full set of participants. With more time, I would have done this more fully and iteratively, both to strengthen the account's validity and to give the community a fuller voice in how it was represented. Additionally in retrospect, a broader IRB protocol would have allowed me to incorporate direct messages and email correspondence with participants**.** This is where much of my participation took place, and leaving that interaction out makes the story less ethnographic and omits some of its most revealing interactions.

My fieldwork window also coincided with a pause in major in-person skeptic gatherings. The Heartland conference, among the most significant of these events, resumed in 2026, after my data collection had closed. In-person observation would have added a dimension this study could not capture: the embodied, face-to-face dynamics of a community I encountered almost entirely online.

Finally, I collected more material than I could feasibly code qualitatively. A digital ethnography of a community this prolific invites over-collection, and in hindsight I would scope the corpus more realistically from the outset and pair the qualitative analysis with quantitative methods suited to the volume.

8.3 Future Work

These limitations lead directly into future work. The first is a quantitative extension of this work. Given the sheer volume of content the community produces online, and the understanding of its practices and infrastructure that this study has built, a computational analysis could now complement the qualitative account. It would test how far the patterns identified here generalize across the community, and trace the movement of data and claims at scale that a close reading cannot reach. This may be somewhat difficult as social media companies continue to restrict data access, but the community is so active on blogs and alternative websites that access is still feasible. I would also like to examine how members of the community react to generative AI, and whether they extend to it the same skepticism they apply to climate science.

A second direction follows from the finding of the audit orientation. The language of the "audit" is not confined to skeptics, as the discussion shows. It circulates widely in contemporary data discourse, across finance, governance, and algorithmic accountability. Whether the audit orientation documented here is a feature of this community or a broader disposition is a question for future research. Specifically, I want to conduct a digital ethnography of environmental justice advocates. Both groups reuse similar climate datasets and often challenge scientific authority, yet they mobilize data toward different goals.

Taken with the rest of this dissertation, this future work begins from the premise that we should take communities like this one seriously. They are prolific and effective data reusers and infrastructure builders rather than anomalies. They are perhaps a predictable outcome of open data. Taking them seriously is a precondition for a fuller understanding of how open data actually works.

Appendices

Appendix A Semi-structured Interview Protocol

Semi-structured Interview Protocol
Length: 60 minutes

Informed Consent Script Before Interview

Thanks so much for taking the time to talk with me today. I want to start by giving you an overview of the project and to make sure you understand what the research entails.

Again, I am a PhD candidate at the University of Michigan, conducting research for my dissertation on how people use publicly available data to critique scientific claims related to the climate. I really appreciate your time and insights on this topic.

[If they have already submitted the consent/interview preferences form]

You've already submitted the consent and interview preferences form. Thanks so much for that.

[review choices and answer questions]

Just a reminder, your participation is completely voluntary. If you ever want to skip any question or even stop participating, that's totally fine.

Do you have any questions about the consent form?

[if they have not already submitted the consent/interview preferences form]

Do you have any questions about the consent form before we move on to the consent and interview preferences at the bottom?

[review consent and optional consent choices, guide through choosing/signing, answer questions]

Just a reminder, your participation is completely voluntary. If you ever want to skip any question or even stop participating, that's fine.

Excellent. If anything comes to mind later, please contact me or my faculty advisor.

1. **Introduction**
    1. Okay, now onto the interview, the main purpose of this interview is to walk me through your research project about [ADD HERE] and how you work with data. I've read your publications, but I want to hear from you directly. I'll also ask you some overview questions at the end, but only if we have time. As I ask questions, feel free to share your screen if you want to show me images or documents related to your research. Any last questions before we begin?
    2. **[If they consented to video or audio recording]:** Okay, and if you're ready, I'll start the recording now. Is that alright? [Wait for affirmative]. I just started the recording. Thank you for agreeing to be recorded.
2. **Walkthrough of data reuse**

    [Specific questions asked are based on the individual's research project and what I am missing from the data journey.]

    1. **[Research idea] - Can you start by telling me about how you got the idea for the project?**
        1. What made you take the step to conduct your own research on the climate? [Goal - articulating both practices and motivations]
            1. How did you learn how to do this?
            2. How do you keep learning or try to get better at doing your research?
        2. Do you have any collaborators in this project? Who are they?
    2. **[Locating and Choosing Data] -** From your project, I identified # data sources [Read and/or show the list]. Is this correct?
        1. Why did you choose these specific data sources?
        2. How did you locate or find the data you reused?
        3. [Potential follow-ups]

1. Are there specific file formats, like CSV files, that you prefer when downloading datasets?

3. **[Assessing Data - Trustworthiness]**
    1. How did you assess the quality or usefulness of the data?
        1. [Potential follow-ups]
            1. What specific things did you look for to decide if your different data sources were trustworthy?
            2. Was there anything specific about the [data provider/source] that made you see it as trustworthy or not trustworthy?
            3. When do you typically question the trustworthiness of data?
            4. Are there specific things you look for in data that would make you trust it more?

4. **[Processing/cleaning] -** How did you prepare your data for analysis? [Clarification prompt, if needed]: For example, did you need to clean it, change anything, or narrow it down?
    1. [Potential follow-ups]
        1. Why did you need to do this?
        2. Did you face any challenges while doing it?

5. **[Analysis] -** How did you analyze the data? Can you walk me through the steps?
    1. [Potential follow-ups]
        1. Did you use any resources—like books, articles, websites, or help from colleagues—to help with your analysis?
        2. Did any colleagues review your analysis or give you feedback? What kind of feedback did they give?
        3. What were your main conclusions from the analysis?
        4. As you know data can be analyzed in a variety of ways. Did you think about other ways to interpret your data or results while you were analyzing them?
            1. [Potential follow-ups]

1. Can you explain what those other interpretations were, and why you went with the one you chose?
2. How did you decide your interpretation was the right one?

6. **[Publish/share] -** How did you share your research findings?
    1. [Potential follow-ups]
        1. How do you respond to reviews or feedback on your work? [Point out specific discussions in the comments if relevant]
        2. How do you handle requests from others to share your data or code? [Point out specific discussions in the comments if relevant]
            1. What things do you think about when deciding whether to share or not?
7. **[Preserve] -** What do you usually do with your data and code after the project is done?
    1. [Potential follow-ups]
        1. [If they say they keep the data] How do you store it?
            1. What do you name the files?
        2. Do you add any extra info to the files—like where it came from, the format, or what project it's for?
        3. Do you ever go back to the data? [If yes] Why?
8. **[Usability Questions]**
    1. Looking back at the data you used, is there anything that would've made it easier to use?
        1. Are there any tools or features—like visualizations, API access, or export options—that would've helped?
    2. Were there any challenges with using [open data source] that you had to figure out?
        1. Did you ever need to contact someone at the [data repository] for help? Why or why not?
9. **[Closing questions for the walkthrough]** - Is there anything else about your research or experience with the project that you'd like to share?

3. **Climate Change Beliefs**

   Transition: Now, I'm going to ask you some more general questions about how you find and evaluate research in climate science.

   1. How do you stay up to date with research?
   2. How do you evaluate others' climate change research?
      1. [Goal - The question is meant to investigate how interviewees evaluate research. Investigate their scientific knowledge and data literacy.]
   3. Be as honest as you want. What do you think are the biggest strengths or weaknesses of mainstream climate science?
      1. So, it sounds like you have a lot of critiques of climate science. Can you say more about those?

   Transition: Next, I'd like to discuss a slightly different topic. A lot of public datasets sometimes get updated, changed, or even removed for different reasons.

4. **Recent Repository Changes:** Have you ever run into problems when data was changed or removed from a data repository?
   1. [if yes] Can you tell me what happened?
      1. Were you able to to work around it or solve the issue? [If yes] How?
      2. Were you able to overcome them? If so, how?
   2. [if no]
      1. Even though you haven't had this happen, what do you think about the idea of data being changed or removed?
      2. If data is changed or taken down, what kind of documentation or explanation would you want to see?
   3. The Trump administration is changing some of these federal agencies' structure and their data. Does that change your trust in this data? How?
5. **Demographic Question**

   [only ask if not already known]

   1. Can you tell me about your current job or what you do day to day?

1. [Follow up if they mention they are retired] What kind of work did you do before you retired?

2. Can you tell me a bit about your educational background?

6. Closing

1. Is there anyone else you think I should talk to who could help me better understand how open climate data is reused?
2. Is there anything else you think I should know that I haven't asked about?

Debriefing Statement After Interview

Okay, we are all done then. Thank you so much for taking the time to talk to me today. Feel free to contact me if you have questions in the future or think of anything else you'd like to share. My contact information is on your copy of the consent form. Do you have any other questions for me before we sign off?

Appendix B Codebook

Name	Description
Accusations of data misuse	Applied where the community's stated standards came into tension with the epistemic norms of mainstream climate science knowledge infrastructure.
Accusations of data misuse - believer to skeptic	
Accusations of data misuse - skeptic to believer	
Accusations of data misuse - skeptic to skeptic	
Archival research	
Internet Archive	https://archive.org/
New England Historical Society	https://www.newenglandhistoricalsociety.com/
Newspapers.com	Newspapers.com
Retrobibliothek	https://www.retrobibliothek.de/retrobib/seite.html?id=119669&imageview=true
The National Library of Australia - TROVE Database	
The New York Times Archive	https://timesmachine.nytimes.com/browser Also called the Times Machine
Washington Post Archives	Uses Washington Post Free Article Preview https://www.washingtonpost.com/wp-srv/world/digitalarchive/index.html
Climate change beliefs	
Admits temperature is rising or warm	
Believe global average temperature is impossible to measure	
Believe world is warming but not because of CO2	

Name	Description
Believes in climate change but it's not a catastrophe	
Believes in greenhouse effect	
Climate is changing but not because of humans	
Lukewarmers	
Warming is not necessarily caused by greenhouse gases	
Warming-increased CO2 is occurring but it is good	This includes discussion of warming or CO2 levels increasing.
Worried about 'real' environmental issues	Use this code when they are talking about caring about other environmental issues.
Data repositories	Marks reuse of an external data repository or archive. The individual repositories coded here have been collapsed in this codebook to avoid duplication.
Data Reusers	
Believers	
Skeptics	
How do other members of the community interact with the artifacts skeptics produce from the data	
Active comment section with reuser	Support, questions, and critiques Back and forth between author and commenters Mimics peer review in a way
Commentor asking questions	
Commentor pointing out potential errors in reuse	
Commentor points out errors in code	
Commentor requests data and code from reuser to reproduce findings	
Commentor tries to reproduce findings	
Commentors know one another	

Name	Description
Retractions	
Reuser changing research based on comments	
Reuser provides code based off commentor request	
Reuser provides reuse instructions off of commentor's request	
Reuser refuses to give data and code to commentor	
Modify skeptics' visualizations	Ex: Adding trend line.
Personal attacks	
Reuse skeptics' visualizations	
Using visualizations without reference	
How do they acquire data	
.csv	
Contact repository or government	This is used when data are not publicly accessible.
Data producer refuses to provide data	
Data upon request	Directly contacted the authors of a paper to get data.
Export vizualization as PNG or JPEG	
How do they cite data	
Data Papers	
data repository mention	
Dataset mention	
Dataset not mentioned	
FTP	
Link to data	
Links to story about data rather than data	
How do they discuss open	

Name	Description
data	
Data are the highest form of evidence	Data-driven decisions
Disprove my beliefs with data	
Importance of reproducibility	Importance of reproducibility for research integrity.
Only trust yourself	Take nobody's word for it
Skeptic and believer disagreement with data	Skeptic and believer are disagreeing about data and its analysis on the post. Either the skeptic and/or the believer are using data to support their arguments.
Super long twitter thread	
How do they evaluate open data	
What skeptics accuse climate scientists of doing with data	
Adjust data to support theory	Homogenize, adjust, manipulate
Can't distinguish trend when you take into account natural variability	
Cherrypicking observational data	Accuse scientists' of cherrypicking observational data to include in their analysis. For example, accuse scientists of only using satellite data (beginning in 1979) for sea ice levels when they think other data (historic ship logbooks) from earlier periods is available that would show sea ice is not melting.
Climategate	
Computational thinking is better than scientific thinking	Basically saying climate scientists are bad at computer science, statistics, etc.
Correlation is not causation	
Critique of data smoothing	Signal vs. noise
Do not blame believers	Instead they blame the system or human nature.
Honest Mistakes	Skeptics characterize the errors they identify in climate science as honest mistakes rather than deliberate or nefarious acts.
Logical fallacy - Appeal to authority	

Name	Description
Manipulating visualizations to support conclusions	
Not doing the ground truthing of data	
What skeptics accuse climate science of as a whole	
Accuse climate science of being a pseudoscience, cult, or religion	
Censorship	
Climate is too complex and cannot be or is not yet understood	
Climate science is not fit for any policy purpose	
Confirmation bias	
Corrupted incentives	
Lack of data and statisitical rigor in climate science	
Lack of transparency	Use when skeptics accuse data repositories or creators, or climate science in general of lack of transparency
Media conflates weather with climate change	
Only older people can speak truth to power	
Positive publication bias	
Rewrite theory so that it fits with unexpected observational data	
Some models are untestable	
Evaluate data repositories	

Name	Description
Accuse repository of changing or removing data to support theory	Comparing different versions of data
Agencies not posting data that doesn't fit with ACC	
Assessing database qualities	
Believe agency's data, not their interpretation of it	
Calling out data repository employees by name	Provide contact information
Compare data repository at two different times	Using internet archive etc.
Comparing two different copies of the same version of a dataset	
Contacting data repository manager to discuss dataset	
Difficult to get to the actual data	
Discuss future and past curation of data	
Distrust certain data sources	
Examine data documentation	
Mentioning previous repository errors	
Publish correspondance with data repository employees	
Questions actions or motivations of data repository employees	
Evaluate datasets	

Name	Description
Compare to data created by ACC skeptics	
Comparing different versions of dataset	
Contact author directly to understand data	
Critique data instruments	Surface temperature stations (urban heat island effect).
Instruments are more sensitive today	Instruments are more sensitive today so they record more observations that were missed during previous times. This affects observational trends.
Natural disaster rate increase caused by increase in reporting not frequency	
Temperature stations in wrong areas	
Critique infilled or 'ghost' temperature data	
Critiquing proxies for temperature	
Data are superior to models	
Data are superior to studies	Data are superior to studies because they are not as subject to manipulation.
Data itself is flawed	
Difficult to accurately measure ocean temperature	
Limitations of available data	
Mention geographical and temporal limitations of existing observational data	
Critique short time period of satellite observations	

Name	Description
Model outputs are not data	
Projection won't happen so data doesn't matter	
Provenance	
Raw data is superior	
Satellite data is better than surface data	
Satellites are inferior to tide gauges	Satellites are not good measurement of sea level rise because they can only measure the open ocean. They cannot measure the coast, which is the only place that matters. I have not evaluated if this is true.
Uncertainty of measurements	
Using GenAI to evaluate datasets	
Evaluate journal articles	
Rating quality of studies with self-made scale	
Comparing different disciplines data editing practices	
Shortcomings of models	
Use FOI Requests	
Use reverse image search to determine visualization's provenance	
How do they find data	
Find data in peer reviewed journal articles	
Read journal articles	
General knowledge of where data exists	
Google Image Search	Use Google Image Search for visualizations.
Use contacts to find data	
Use data they have reused before	

Name	Description
Use Generative AI	
Use Google Search to check for datasets	
How do they get research ideas	
Noticing discrepancies in climate science research	
Respond to articles or reports	
Topics that are important to people	
How do they publish their research	
Published in high impact journal	
How do they reuse data	
Data mistakes they make	Applied in cases where practices diverged from the community's own stated standards and epistemic norms.
Adding uncertainty to the conversation by analyzing same data in various ways	
Cherrypicking observational data	
Correlation is not causation	
Going against dataset's appropriate use	
Haven't updated data to match repository changes	
Link Rot	Link rot is the phenomenon of hyperlinks tending over time to cease to point to their originally targeted file, web page, or server due to that resource being relocated to a new address or becoming permanently unavailable. A link that no longer resolves at the intended target may be called broken or dead.
Local vs. Global	ACC skeptics suggest that local trends are representative of global patterns.
Look at maximum temperature rather than average	

Name	Description
Misuse of database tool	
No statistical significance	
Suggest trends should match exactly	This code should be used when ACC skeptics suggest that if acceleration or trends don't match exactly then a relationship between the two is not occurring.
What they are doing with the data	
Add projections to graph to compare to observations	
Analyze the same dataset in different ways	
Comparing and critiquing similar datasets	Also how they evaluate data. Often compare to datasets created by skeptics.
Comparing CO2 Ice Core data with other CO2 datasets	
Comparing different model projections	
Correlation between two variables	
Create database	
Create dataset and include interpretation of that data	
Create own dataset from global temperature sources	HAADCRUT4, GISTEMP, UAH, RSS
Data backed talking points for skeptics	
Examining which equations best fit the data	Linear, polynomial, etc.
Ground truthing	
Identify trends or make predictions	
Creating visualizations combining datasets to show patterns	

Name	Description
Instruct others how to reuse data	
Linear regression	
List of data repositories and tools	
Provide data to interested readers	
Reproduce an article's findings	
Statistical test of significance between CO2 and Temperature	
Testing their equations against data	For instance: “The carbon cycle equations we built accurately replicate IPCC’s natural carbon cycle data shown in Figure 2.”
Update diagrams with new data	
Use data to check if todays temperature is uniquely warm	
Use data to refute IPCC claims	
Use data to show more rapidly warming periods	
Use data to test model projections	
Use temperature record to make predictions	Example: projection of global surface temperatures assuming linear relationship
Using data to test IPCC predictions	
Tools used to reuse data	
Excel charts	
Interactive database visualization	
Javascript	
Outside interactive data visualization tool	WeatherBELL Analytics

Name	Description
Prefer pre-made visualization over creating themselves	
R programming	
RStudio	
Reusing own excel formulas	
Scraping website	
Use GenAI for data analysis	
Use repository tool and post with instructions and results	Include how to reproduce
How do they portray or visualize the data	
Copy and paste spreadsheet	How they present data to the readers.
Direct download of data	
Highlight something in the graph using PowerPoint	
Modify journal article graph with observational data	
Post code used for data analysis	
Self-made visualization	No interactive visualization tools used. Instead, common tools like excel or RStudio.
Use paint to export Excel files into .jpg or .png	
Visualizing same data in different ways	
Mentioning limitations in study	
Admit to errors in analysis	
Admit uncertainty in analysis	Or perhaps introduce it

Name	Description
What data do they choose	
Choose processed data by ACC skeptics rather than raw data	
Reuse data from journal articles	
Use data from skeptics' experiments	
Use data with licensing restrictions	
Documentation with reuse	
Instruct others how to reproduce their figures	
Problems they encounter	
Changes in dataset causing issues in reuse process	
Data processing	
Interoperability issues	Facing issues when needing to combine datasets for data reuse.
Discussion among coauthors	
Quality assessment data reuse	
Research topics	
Alternate causes of observed climate change topics	
El Nino causes warming	El Niño causes warming and accounts for many of the warming trends that are attributed to ACC.
Sun's effect on climate	Research topic
Sunspots effects on climate	
Climate sensitivity	
Cloud topics	

Name	Description
Problems with clouds in models	
Relationship between cloud and ice albedo	
CO2 topics	
Causes of CO2 increase	Typically do not attribute the increase to humans
CO2 follows temperature	
CO2 Levels	
Greenhouse Gases Act as Coolants	
Relationship between CO2 and Temperature	
Complexity topics	
Chaos makes prediction impossible	
Uncertainty of the future climate	
Energy consumption and fossil fuel use topics	
Air pollution goes down despite increasing fossil fuels	
Clean water has gone up with increasing fossil fuels	
Energy Consumption	
Inadequacy of Renewable Energy Sources	
Income goes up with fossil fuel consumption	

Name	Description
Life expectancy goes up with fossil fuel consumption	
Ice, snow, and sea topics	
No trend in Arctic sea ice	
Sea ice levels	
Sea Level Rise	Global sea level has been increasing since the mid-1800s and there has only been a minor recent acceleration.
Total snow mass	
Natural disasters and quality of life topics	
California Drought	
Climate-related deaths going down	
Disaster reporting standards	
Food Production	
Heatwaves	Heatwaves are a normal part of the climate, and are not increasing in frequency or intensity.
Lake Mead Water Levels	Water levels are lowering but rose for a while so lowering is inevitable. There will always be some areas in drought.
Malaria	If global warming caused uptick in malaria we would be seeing it. Instead we see a decline in deaths.
Natural disaster rates	
Relationship between rainfall and wildfire	
Water Levels Lake Tahoe	Lake Tahoe has reached the maximum allowable water level through 2017-2019. Previous years were in the middle of a natural drought.
Weather disproving ACC	
Wildfires	Wildfires are far less frequent and severe than the 20th century. When occasional upticks occur its still less land.
Temperature topics	
Homogenization mixes good and bad climate data	

Name	Description
from temperature stations	
Manipulation of temperature record	
Temperature increase even if it does occur won't matter for us	Colder climates like Canada
Temperature increase is just recovery from the Little Ice Age	
Temperature is cooling	
Temperature not increasing	
Temperature record	
Weather disproving ACC	
How do they select data	
Choose most comprehensive dataset to avoid interoperability issues	
Choose temporally longest data source	
Evaluate different datasets to see which is higher quality	
Look for specific geographical ranges in datasets	
Preference for Russian model	This denies the future catastrophe
How do they share data	
Data available upon request	
Made data and code publicly available	
Use Generative AI to write papers	

Name	Description
How do they stay up to date with climate science topics	
Google Search	
ResearchGate.net	
Use blogs to stay up to date	
Use Generative AI	
Use Google Scholar	
Use personal connections to stay up to date	
Use social media to stay up to date	
Want to find real reasons for observed environmental changes	
How do they store the data	
Data backed up in multiple locations	
File Naming Conventions	
Save in folder	
How do they try to convince others	
Convince believers of one small error rather than changing world view completely	
Find middle ground	
Teach the scientific method	
Interviewee background and how they got involved in climate science	
Computer Science Background	
Conservative or libertarian political background	

Name	Description
Did own research	Not just reusing data but conducting their own research.
Ecology Background	
Legal background	
Must look at the data to understand the science	
Physics Background	
Proposed legislation or regulations lead to investigation	Skeptics began to reuse data or look into climate science in response to proposed legislation or regulations that would impose on individual freedom.
Related professional industry.	Got involved in climate science research because they were associated with industry that was garnering negative attention from environmental organizations. Not gas/oil.
Serendipity	Became interested in climate change skepticism through chance, rather than through a related job, industry or politics.
Mainstream epistemic values	
Challenge existence of a consensus	
Credential dropping	Scientific support from Drs, or dropping one's own credentials.
Critiquing skeptics	
Diversity	
Emphasize their independence	
Portray information as objective	
Pragmatic Pessimism	
Scientific method	
Transparency	
Public Comments	
Misinformation researcher techniques	
Accusing climate change believers of common misinformation techniques	
Blame lack of data literacy	
Garbage-In, Garbage-Out	

Name	Description
Platforms	Marks the platform on which skeptic content was published or circulated (e.g., blogs, Reddit, X, YouTube, organizational publications). The individual platforms and sites coded here have been collapsed in this codebook for brevity.
Relationships between platforms	Marks instances where content moved between platforms, such as cross-posting from a blog to Reddit or X. The individual platform-to-platform links coded here have been collapsed in this codebook for brevity.
STS themes	
Accuse scientists of groupthink	
Article using situated knowledges language that skeptics use	
Citizen Science	Also seen in Wood for Trees and Surface Stations site
Critique of peer review	
IPCC Peer Review Critique	
Discuss importance of open science and data	This can include a discussion of the benefits of transparency, open science, or the internet.
Disparaging scientific expertise	
Expertise is distributed, legitimacy is not	
Federal administrative changes will affect data	
Manipulating statistics	
Mentioning limitations of their data analysis	
Notes things that are not objective in their own website	
Objectivity as epistemic ideal	
One plus truth equals a majority	
Purpose of science - questioning consensus	
Relationship between policy and science	
Science is everchanging	
Social constructedness of data	

Name	Description
Social constructedness of models	
Social constructedness of science	
STS citations	
Tools	Marks use of an external computational, analysis, or visualization tool. The individual tools coded here have been collapsed in this codebook for brevity.
Trump and scientific data removal	
Change interpretation of data on sites	
Decrease redundancies in agencies	
Don't change existing data	
Don't remove the data	
More explanation of why data may be incorrect but keep the data up	
Only fund high quality data	
Recognize that he is destroying infrastructure on purpose.	
Separate funding science from climate change	
Tear it all down	
We still need the data for weather	Even if they don’t believe the data can show climate change, they do think it should still be collected because we need it to record weather.
Trust development	
Having multiple datasets say the same thing	
Reputation of data provider	
Suspicious outliers	Begin to question accuracy of data because of suspicious outliers
Trusted it in the past so they trust it now	

Name	Description
What do skeptics want from repositories	
Better search engines	
Critique of netCDF	
Document changes	
GitHub with so many small files is confusing	
High resolution visualizations	
Improved navigability	
Interactions with repositories	There are some other sub-codes that should be copied into this high-level code when completing writing and analysis.
Contact repository to offer them data	
Interoperable file formats	
Keep all versions of the dataset	This is sometimes previously coded as keep raw data.
Less propaganda	Delivering a message to do something rather than just presenting the data and letting others decide what to do.
Make limitations of existing data clear	
Make raw or unprocessed data available	If processing/cleaning/adjustments have to occur, still provide the unadjusted data to users.
More charts and visualizations	
More documentation	
More objective data = better	
Redirect pages	Skeptics want more redirect pages so if things get changed they know where to go.
Requesting more curation from data repositories	
Stop using abbreviations or jargon	
Subsetting features	
Too much metadata	

Name	Description
Use most simple file format possible	.csv or text format
Various export options	
Want repository to open the data processing black box	
What do they say would make them change their beliefs	
New data might convince me to change my opinion	
Why do they reuse data	
Commissioned by US government	
Data use for policy argument	
Z-Extras	
2020 Election Data	
ACC Skeptic asking interviewer questions	
Climate change is the way for elites to control us	
DOE 2025 Climate Report	
Existing expertise	
Goals	
Internal Review	
Short time frame	
Environmental stewardship in the Bible	
Generative AI	
Generative AI is left-wing biased	
How to conduct own experiments	
Interesting image	
Interesting quote	
Interesting reuse case	

Name	Description
Media catastrophizing	
See denialist as a derogatory word	
Tie to Covid-19	
Want to be viewed as equal to climate scientists	

Appendix C Public Datasets Found

#	Publicly-Available Datasets
1	Alberta Electric System Operator Data
2	Australian Government Bureau of Meteorology - Rainfall History
3	Berkeley Earth - Monthly Global Average Surface Temperature
4	BP - Statistical Review World Energy
5	Centre for Research on the Epidemiology of Disasters EM-DAT The International Disaster Database
6	Climatic Research Unit (CRU) - North Atlantic Oscillation (NAO) Data Files
7	Climatic Research Unit (CRU) - Southern Oscillation Index (SOI)
8	DOE ESS-DIVE - Global, Regional, and National Fossil-Fuel CO2 Emissions: 1751-2017
9	DOE ESS-DIVE - Tropospheric and Lower Stratospheric Temperature Anomalies Based on Global Radiosonde Network Data (1958 - 2005)
10	DOE ESS-DIVE (formerly CDIAC) - Law Dome Ice Core CH4 Data (Etheridge et al.)
11	EPA - Air Pollutant Emissions Trends Data
12	EPA - Heat Wave Index (Climate Change Indicators)
13	Ercot Grid Information - Grid and Market Conditions
14	EUMETSAT Ocean and Sea Ice (OSI) SAF - Sea-ice index
15	European Energy Exchange Transparency Platform - Power Data
16	Federal Statistical Office of Germany Genesis Online Database - Monthly Electricity Production Data
17	Food and Agriculture Organization of the United Nations (FAO) - FAOSTAT Agricultural Production Statistics
18	Government of Canada Climate & Weather - Historical Data
19	Index Mundi - Commodity Food Price Index Monthly Price - Index Number
20	Japan Meteorolgical Agency - Number of Typhoons
21	Japan Meteorological Agency (JMA) Tokyo Climate Center WMO Regional Climate Center in RA II (Asia) - Global Average Surface Temperature Anomalies
22	Hessisches Landesamt für Naturschutz, Umwelt und Geologie (HLNUG) - Linden/Giessen Continuous CO2 Monitoring Data
23	Met Office Climate and Climate Change - UK Temperature, rainfall and sunshine time series

24	Met Office Hadley Centre observations datasets - HadAT
25	Met Office Hadley Centre observations datasets - HadCET
26	Met Office Hadley Centre observations datasets - HadSST
27	Met Office Hadley Centre observations datasets - UKMO EN4
28	Met Office Hadley Centre observations datasets – HADCRUT
29	NASA Earth Data - Active Fire Data
30	NASA Goddard Institute for Space Studies (GISS) - Surface Temperature Analysis (GISTEMP)
31	NASA Goddard Institute for Space Studies (GISS) - Ice Core CO2 Data (pre-1958)
32	National Oceanography Centre Permanent Service for Mean Sea Level (PSMSL)
33	National Snow & Ice Data Center (NSIDC) - Multisensor Analyzed Sea Ice Extent - Northern Hemisphere (MASIE-NH)
34	National Snow & Ice Data Center (NSIDC) - Sea Ice Index
35	NOAA Atlantic Oceanographic and Meteorological Laboratory - Hurricane Data
36	NOAA Global Monitoring Laboratory (GML) Earth System Research Laboratories - Background CO2 Data (Barrow, South Pole)
37	NOAA Global Monitoring Laboratory (GML) Earth System Research Laboratories - Monthly Average Mauna Loa CH4
38	NOAA Global Monitoring Laboratory (GML) Earth System Research Laboratories - Monthly Average Mauna Loa CO2
39	NOAA Global Monitoring Laboratory (GML) Earth System Research Laboratories - Water Vapor
40	NOAA National Centers for Environmental Information (NCEI) - Global Historical Climatology Network Monthly (GHCN-M)
41	NOAA National Weather Service Climate Prediction Center - Cold & Warm Episodes by Season
42	NOAA National Weather Service Climate Prediction Center - NCEP/NCAR Reanalysis Data (CDAS)
43	NOAA National Centers for Environmental Information (NCEI) - Global Surface Temperature Anomalies
44	NOAA National Centers for Environmental Information (NCEI) - Law Dome Ice Core CO2 Data (Etheridge et al.)
45	NOAA National Centers for Environmental Information (NCEI) - NOAA Climate Data Record (CDR) of Total Solar Irradiance (TSI), NRLTCI
46	NOAA National Centers for Environmental Information (NCEI) - Radiosonde Atmospheric Temperature Products for Assessing Climate (RATPAC)
47	NOAA National Centers for Environmental Information (NCEI) - Sea Ice and Snow Cover Extent

48	NOAA National Centers for Environmental Information (NCEI) - US Climate Reference Network (USCRN)
49	NOAA Physical Sciences Laboratory - Climate Timeseries AMO (Atlantic Multidecadal Oscillation) Index
50	NOAA Physical Sciences Laboratory - NCEP-NCAR Reanalysis Monthly Means and Other Derived Variables
51	NOAA Physical Sciences Laboratory - NOAA Interpolated Outgoing Longwave Radiation (OLR)
52	NOAA Physical Sciences Laboratory - Pacific Decadal Oscillation (PDO)
53	NOAA Space Weather Prediction Center - Predicted Sunspot Number and Radio Flux
54	NOAA Tides and Currents Stations - Sea Level Trends
55	Polish Institute of Meteorology and Water Management (IMGW-PIB) - Tide Gauge Sea Level Data (Świnoujście)
56	Remote Sensing System Microve Climate Data Center (MCDC) - RSS Upper Air Temperature Products
57	Royal Observatory of Belgium Solar Influences Data and Analysis Centre (SIDC)- Sunspot number series
58	The University of Alabama in Huntsville Earth System Science Center - Global Temperature Report
59	The World Bank Data Catalog - World Development Indicators
60	University of California, Irvine CHRS RainSphere - Historical Satellite Observation
61	University of California, San Diego - Argo 2.5m Global Ocean Temperature Anomaly Float Measurements
62	University of California, San Diego Scripps CO2 Program - Ice-Core Merged Products
63	University of California, San Diego Scripps CO2 Program - Atmospheric CO2 Records (Keeling)
64	University of Colorado Boulder Solar Radion & Climate Experiment - Total Solar Irradiance Data
65	University of Colorado Center for Astrodynamics - Global Mean Sea Level (GMSL) Data
66	University of Groningen Growth and Development Centre - Maddison Historical Statistics
67	University of Hawaii Sea Level Center - Tide Gauge Data
68	US Energy Information Administration - Crude Oil Production
69	US National Interagency Fire Center - Total Wildland Fires and Acres
70	USGS National Water Information System - Surface Water
71	World Health Organization (WHO) - Global Malaria Deaths

Bibliography

Abatzoglou, J. T. (2013). Development of gridded surface meteorological data for ecological applications and modelling. International Journal of Climatology, 33(1), 121–131. https://doi.org/10.1002/joc.3413

About Us. (n.d.-a). The Heartland Institute. Retrieved March 26, 2026, from https://heartland.org/about-us/

About Us. (n.d.-b). CO2 Coalition. Retrieved March 31, 2026, from https://co2coalition.org/about-us/

Acker, A. (2015). How cells became records: Standardization and infrastructure in tissue culture. Archival Science, 15(1), 1–24. https://doi.org/10.1007/s10502-013-9213-x

Akin, H., Hardy, B. W., & Jamieson, K. H. (2020). Countering Identity-protective Responses to Climate Change Data. Environmental Communication, 14(8), 1111–1126. https://doi.org/10.1080/17524032.2020.1776359

Aklin, M., & Urpelainen, J. (2014). Perceptions of scientific dissent undermine public support for environmental policy. Environmental Science & Policy, 38, 173–177. https://doi.org/10.1016/j.envsci.2013.10.006

Altheide, D. L. (1987). Reflections: Ethnographic content analysis. Qualitative Sociology, 10(1), 65–77. https://doi.org/10.1007/BF00988269

Anderson, C. (2008, June 23). The End of Theory: The Data Deluge Makes the Scientific Method Obsolete. Wired, 16(07). http://www.wired.com/science/discoveries/magazine/16-07/pb_theory

Anti-Eviction Mapping Project. (2021). Anti-Eviction Mapping Project. https://antievictionmap.com

Arzberger, P., Schroeder, P., Beaulieu, A., Bowker, G., Casey, K., Laaksonen, L., Moorman, D., Uhlir, P., & Wouters, P. (2004). An International Framework to Promote Access to Data. Science, 303(5665), 1777–1778. https://doi.org/10.1126/science.1095958

Baack, S. (2015). Datafication and empowerment: How the open data movement re-articulates notions of democracy, participation, and journalism. Big Data & Society, 2(2), 2053951715594634. https://doi.org/10.1177/2053951715594634

Bak, G. (2016). Trusted by whom? TDRs, standards culture and the nature of trust. Archival Science, 16(4), 373–402. https://doi.org/10.1007/s10502-015-9257-1

Ball, A. (2012, February 13). Review of Data Management Lifecycle Models. Bath, UK: University of Bath.

Bandy, J. (2021). Problematic Machine Behavior: A Systematic Literature Review of Algorithm Audits. Proceedings of the ACM on Human-Computer Interaction, 5(CSCW1), 1–34. https://doi.org/10.1145/3449148

Bates, J. (2017). The politics of data friction. Journal of Documentation, 74(2), 412–429. (world). https://doi.org/10.1108/JD-05-2017-0080

Bates, J., Lin, Y.-W., & Goodale, P. (2016). Data journeys: Capturing the socio-material constitution of data objects and flows. Big Data & Society, 3(2), 2053951716654502. https://doi.org/10.1177/2053951716654502

Beaulieu, A. (2010). Research Note: From co-location to co-presence: Shifts in the use of ethnography for the study of knowledge. Social Studies of Science, 40(3), 453–470. https://doi.org/10.1177/0306312709359219

Beck, E.-G. (2007). 180 Years of Atmospheric CO2 Gas Analysis by Chemical Methods. Energy & Environment, 18(2).

Beck, E.-G. (2022). Reconstruction of Atmospheric CO2 Background Levels since 1826 from Direct Measurements near Ground. Science of Climate Change, 2.2, 148–211. https://doi.org/10.53234/scc202112/16

Berkhout, F., Tuinstra, W., Tàbara, J. D., Jäger, J., Chabay, I., de Wit, B., Langlais, R., Mills, D., Moll, P., Otto, I. M., Petersen, A., Pohl, C., & van Kerkhoff, L. (2013). Opening up knowledge systems for better responses to global environmental change. Environmental Science & Policy, Special Issue: Responding to the Challenges of Our Unstable Earth (RESCUE), 28, 60–70. https://doi.org/10.1016/j.envsci.2012.11.008

Biagioli, M., & Pottage, A. (2022, March 19). Dark Transparency: Hyper-Ethics at Trump's EPA. Los Angeles Review of Books. https://lareviewofbooks.org/article/dark-transparency-hyper-ethics-at-trumps-epa

Biernacki, P., & Waldorf, D. (1981). Snowball Sampling: Problems and Techniques of Chain Referral Sampling. Sociological Methods & Research, 10(2), 141–163. https://doi.org/10.1177/004912418101000205

Bietz, M. J., & Lee, C. P. (2009). Collaboration in Metagenomics: Sequence Databases and the Organization of Scientific Work. In I. Wagner, H. Tellioğlu, E. Balka, C. Simone, & L. Ciolfi (Eds.), ECSCW 2009 (pp. 243–262). Springer London. https://doi.org/10.1007/978-1-84882-854-4_15

Birt, L., Scott, S., Cavers, D., Campbell, C., & Walter, F. (2016). Member Checking: A Tool to Enhance Trustworthiness or Merely a Nod to Validation? Qualitative Health Research, 26(13), 1802–1811. https://doi.org/10.1177/1049732316654870

Bolsen, T., & Druckman, J. N. (2018). Do partisanship and politicization undermine the impact of a scientific consensus message about climate change? Group Processes & Intergroup Relations, 21(3), 389–402. https://doi.org/10.1177/1368430217737855

Bonney, R., Cooper, C. B., Dickinson, J., Kelling, S., Phillips, T., Rosenberg, K. V., & Shirk, J. (2009). Citizen Science: A Developing Tool for Expanding Science Knowledge and Scientific Literacy. BioScience, 59(11), 977–984. https://doi.org/10.1525/bio.2009.59.11.9

Borges, J. L. (1946). On exactitude in science. In Collected Fictions (A. Hurley, Trans.). In Collected Fictions. Penguin Books.

Borgman, C. L. (2015). Big data, little data, no data: Scholarship in the networked world. The MIT Press.

Borgman, C. L., Darch, P. T., Sands, A. E., Pasquetto, I. V., Golshan, M. S., Wallis, J. C., & Traweek, S. (2015). Knowledge infrastructures in science: Data, diversity, and digital libraries. International Journal on Digital Libraries, 16(3–4), 207–227. https://doi.org/10.1007/s00799-015-0157-z

Borgman, C. L., Darch, P. T., Sands, A. E., Wallis, J. C., & Traweek, S. (2014). The Ups and Downs of Knowledge Infrastructures in Science: Implications for Data Management. Proceedings of the 2014 IEEE/ACM Joint Conference on Digital Libraries (JCDL), 257–266. https://doi.org/10.1109/JCDL.2014.6970177

Borgman, C. L., & Groth, P. (2025). From Data Creator to Data Reuser: Distance Matters. Harvard Data Science Review, 7(2). https://doi.org/10.1162/99608f92.35d32cfc

Borgman, C. L., Sands, A. E., Darch, P. T., & Golshan, M. S. (2016). The durability and fragility of knowledge infrastructures: Lessons learned from astronomy. Proceedings of the Association for Information Science and Technology, 53, 1–10. https://doi.org/10.1002/pra2.2016.14505301057

Borgman, C. L., Scharnhorst, A., & Golshan, M. S. (2019). Digital data archives as knowledge infrastructures: Mediating data sharing and reuse. Journal of the Association for Information Science and Technology, 70(8), 888–904. https://doi.org/10.1002/asi.24172

Borgman, C. L., Wallis, J. C., & Enyedy, N. (2007). Little science confronts the data deluge: Habitat ecology, embedded sensor networks, and digital libraries. International Journal on Digital Libraries, 7(1–2), 17–30.

Borgman, C. L., & Wofford, M. F. (2021). From Data Processes to Data Products: Knowledge Infrastructures in Astronomy. Harvard Data Science Review, (3.3). https://doi.org/10.1162/99608f92.4e792052

Borgman, C. L., Wofford, M. F., Golshan, M. S., Darch, P. T., & Scroggins, M. J. (2020). Collaborative Ethnography at Scale: Reflections on 20 years of data integration. https://escholarship.org/uc/item/5bb8b1tn

Boulton, G., Babini, D., Hodson, S., Li, J., Marwala, T., Musoke, M. G. N., Uhlir, P. F., & Wyatt, S. (2015). Open Data in a Big Data World: An International Accord [Outcome of Science International 2015 meeting]. ICSU, IAP, ISSC, TWAS. https://twas.org/sites/default/files/open-data-in-big-data-world_short_en.pdf

Bourdieu, P. (2008). Outline of a Theory of Practice. In The New Social Theory Reader (2nd ed.). Routledge.

Bowker, G. C. (1994). Science on the Run: Information Management and Industrial Geophysics at Schlumberger, 1920-1940. MIT Press.

Bowker, G. C., & Star, S. L. (1999). Sorting Things Out: Classification and Its Consequences. MIT Press.

Bowser, A., Cooper, C., De Sherbinin, A., Wiggins, A., Brenton, P., Chuang, T.-R., Faustman, E., Haklay, M. (Muki), & Meloche, M. (2020). Still in Need of Norms: The State of the Data in Citizen Science. Citizen Science: Theory and Practice, 5(1), 18. https://doi.org/10.5334/cstp.303

Bowser, A., Shilton, K., Preece, J., & Warrick, E. (2017). Accounting for Privacy in Citizen Science: Ethical Research in a Context of Openness. Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing, 2124–2136. https://doi.org/10.1145/2998181.2998305

boyd, danah. (2015). Making Sense of Teen Life: Strategies for Capturing Ethnographic Data in a Networked Era. In E. Hargittai & C. Sandvig (Eds.), Digital Research Confidential (pp. 79–102). The MIT Press. https://doi.org/10.7551/mitpress/9386.003.0006

Brassard, J., & Singh, B. (2007). Effects of climate change and CO2 increase on potential agricultural production in Southern Québec, Canada. Climate Research, 34, 105–117. https://doi.org/10.3354/cr034105

Brohan, P., Kennedy, J. J., Harris, I., Tett, S. F. B., & Jones, P. D. (2006). Uncertainty estimates in regional and global observed temperature changes: A new data set from 1850. Journal of Geophysical Research, 111(D12), D12106. https://doi.org/10.1029/2005JD006548

Brown, S., Davidovic, J., & Hasan, A. (2021). The algorithm audit: Scoring the algorithms that score us. Big Data & Society, 8(1), 2053951720983865. https://doi.org/10.1177/2053951720983865

Bui, M. (2023). Data and/as activism: Community-based racial justice data repertoires. New Media & Society, 14614448231206118. https://doi.org/10.1177/14614448231206118

Burkhardt, M., van Geenen, D., Gerlitz, C., Hind, S., Kaerlein, T., Lämmerhirt, D., & Volmar, A. (Eds.). (2022). Interrogating Datafication: Towards a Praxeology of Data. transcript Verlag.

Burnett, H. S. (2025, January 10). Quit Fretting, The Hill, Apple Production Is Not Declining from Climate Change. ClimateRealism. https://climaterealism.com/2025/01/quit-fretting-the-hill-apple-production-is-not-declining-from-climate-change/

Burton, D. (2017, January 21). New website gives you the real deal on sea level rise and rates. Watts Up With That? https://wattsupwiththat.com/2017/01/21/new-website-gives-you-the-real-deal-on-sea-level-rise-and-rates/

Burton, D. (2021). TLIB Version Control (Version 18) [Computer software]. Burton Systems Software. https://www.burtonsys.com/overview.html

Burton, D. (2023, January 24). Documentation. https://sealevel.info/data/00readme.txt

Burton, D. (2026a). Historical CH4 mixing ratios from Law Dome (Antarctica) and Summit (Greenland) ice cores. [Dataset]. Sealevel.info. https://sealevel.info/EthCH498B.txt

Burton, D. (2026b, January 15). Mean Sea Level at Swinoujscie, Poland. https://sealevel.info/MSL_graph.php?id=110-092

Burton, D. (2026c, January 15). Mean Sea Level at Swinoujscie, Poland JSON Data File. https://sealevel.info/data/110-092_data.json

Burton, D. (2026d, March 17). Sea-level and climate change information. SeaLevel Info. http://sealevel.info/

Candela, A. (2019). Exploring the Function of Member Checking. The Qualitative Report. https://doi.org/10.46743/2160-3715/2019.3726

Carroll, S. R., Garba, I., Figueroa-Rodríguez, O. L., Holbrook, J., Lovett, R., Materechera, S., Parsons, M., Raseroka, K., Rodriguez-Lonebear, D., Rowe, R., Sara, R., Walker, J. D., Anderson, J., & Hudson, M. (2020). The CARE Principles for Indigenous Data Governance. Data Science Journal, 19(1), Article 1. https://doi.org/10.5334/dsj-2020-043

Cavalier, D., & Kennedy, E. B. (2016). The rightful place of science: Citizen science. Consortium for Science, Policy & Outcomes.

CDIAC. (n.d.). Cdiac.ess-dive.lbl.gov/ftp/trends/atm_meth/EthCH498B.txt. Retrieved May 26, 2026, from https://cdiac.ess-dive.lbl.gov/ftp/trends/atm_meth/EthCH498B.txt

Cera, M. (2023). Digital ethnography: Ethics through the case of QAnon. Frontiers in Sociology, 8. https://doi.org/10.3389/fsoc.2023.1119531

Challinor, A. J., Watson, J., Lobell, D. B., Howden, S. M., Smith, D. R., & Chhetri, N. (2014). A meta-analysis of crop yield under climate change and adaptation. Nature Climate Change, 4(4), 287–291. https://doi.org/10.1038/nclimate2153

Chanen, C., & Soden, R. (2025). Data Practices and Double-Binds of Toronto's Climate Governance. Proc. ACM Hum.-Comput. Interact., 9(2), CSCW057:1-CSCW057:25. https://doi.org/10.1145/3710955

Chao, T. C., Cragin, M. H., & Palmer, C. L. (2015). Data Practices and Curation Vocabulary (DPCVocab): An empirically derived framework of scientific data practices and curatorial processes. Journal of the Association for Information Science and Technology, 66(3), 616–633. https://doi.org/10.1002/asi.23184

Charmaz, K. (1983). The grounded theory method. In Contemporary field research. Little, Brown.

Charmaz, K. (2006). Constructing Grounded Theory: A Practical Guide through Qualitative Analysis. SAGE.

Choudry, A. (2013). Activist research practice: Exploring research and knowledge production for social action. Socialist Studies, 9(1), 128–151. https://doi.org/10.18740/S4G01K

Choudry, A. (2014). (Almost) everything you always wanted to know about activist research but were afraid to ask: What activist researchers say about theory and methodology. Contention: The Multidisciplinary Journal of Social Protest, 1(2), 75–88.

Christin, A. (2020). What Data Can Do: A Typology of Mechanisms. International Journal of Communication, 14(0), Article 0.

Christy, J. R., & Spencer, R. W. (2026). Global Temperature Report: April 2026 with Version 6.1 (Version 6.1) [Dataset]. The Earth System Science Center at The University of Alabama in Huntsville. The Earth System Science Center at The University of Alabama in Huntsville

Cifor, M., & Rawson, K. J. (2023). Mediating Queer and Trans Pasts: The Homosaurus as Queer Information Activism. Information, Communication & Society, 26(11), 2168–2185. https://doi.org/10.1080/1369118X.2022.2072753

Climate Working Group. (2025). A Critical Review of Impacts of Greenhouse Gas Emissions on the U.S. Climate. Department of Energy. https://www.energy.gov/sites/default/files/2025-07/DOE_Critical_Review_of_Impacts_of_GHG_Emissions_on_the_US_Climate_July_2025.pdf

Coady, S. A., Mensah, G. A., Wagner, E. L., Goldfarb, M. E., Hitchcock, D. M., & Giffen, C. A. (2017). Use of the National Heart, Lung, and Blood Institute Data Repository. New England Journal of Medicine, 376(19), 1849–1858. https://doi.org/10.1056/NEJMsa1603542

Cohler, J. (2026). The Physical Tether Theorem: A Rigorous Formal Synthesis of Representational Measurement Theory, Operationalism, and Classical Thermodynamics. Preprint, 1–15.

Committee on Best Practices for a Future Open Code Policy for NASA Space Science, Space Studies Board, Division on Engineering and Physical Sciences, & National Academies of Sciences, Engineering, and Medicine. (2018). Open Source Software Policy Options for NASA Earth and Space Sciences (p. 25217). National Academies Press. https://doi.org/10.17226/25217

Comms, E. (2025, December 10). EPA Scrubs Information About Climate Change Indicators and Impacts. Environmental Data and Governance Initiative. https://envirodatagov.org/epa-scrubs-information-about-climate-change-indicators-and-impacts/

CoreTrustSeal. (2024). https://www.coretrustseal.org/

Cragin, M. H., Palmer, C. L., Carlson, J. R., & Witt, M. (2010). Data sharing, small science and institutional repositories. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 368(1926), 4023–4038. https://doi.org/10.1098/rsta.2010.0165

Crooks, R. (2019). Accesso Libre: Equity of Access to Information through the Lens of Neoliberal Responsiblization. Journal of Critical Library and Information Studies, 2(1). https://doi.org/10.24242/jclis.v2i1.91

Crooks, R., & Currie, M. (2021). Numbers will not save us: Agonistic data practices. The Information Society, 37(4), 201–213. https://doi.org/10.1080/01972243.2021.1920081

Cukier, K. N. (2010, February 25). Data, Data Everywhere. The Economist. http://www.economist.com/node/15557443

Curty, R. G., Crowston, K., Specht, A., Grant, B. W., & Dalton, E. D. (2017). Attitudes and norms affecting scientists' data reuse. PLOS ONE, 12(12), e0189288. https://doi.org/10.1371/journal.pone.0189288

Dagiral, E., & Peerbaye, A. (2016). Making Knowledge in Boundary Infrastructures: Inside and Beyond a Database for Rare Diseases. Science & Technology Studies, Special Issue on Knowledge Infrastructures: Part II, 29(2), 44–61.

Daston, L. (1992). Objectivity and the Escape from Perspective. Social Studies of Science, 22(4), 597–618. https://doi.org/10.1177/030631292022004002

Daston, L., & Galison, P. (2010). Objectivity (Illustrated edition). Zone Books.

Data Rescue Project. (2026, May 11). Data Rescue Project. https://www.datarescueproject.org/

Dayaratna, K. D., McKitrick, R., & Michaels, P. J. (2020). Climate sensitivity, agricultural productivity and the social cost of carbon in FUND. Environmental Economics and Policy Studies, 22(3), 433–448. https://doi.org/10.1007/s10018-020-00263-w

Dayaratna, K., & McKitrick, R. (2023). Reply to comment on "climate sensitivity, agricultural productivity and the social cost of carbon in fund" by Philip Meyer. Environmental Economics and Policy Studies, 25(2), 291–298. https://doi.org/10.1007/s10018-023-00364-2

Dean, J. P., & Whyte, W. F. (1958). How Do You Know If the Informant is Telling the Truth? Human Organization, 17(2), 34–38.

DiCicco-Bloom, B., & Crabtree, B. F. (2006). The qualitative research interview. Medical Education, 40(4), 314–321. https://doi.org/10.1111/j.1365-2929.2006.02418.x

D'Ignazio, C., & Klein, L. F. (2020). Data Feminism. MIT Press.

Dillon, L., Lave, R., Mansfield, B., Wylie, S., Shapiro, N., Chan, A. S., & Murphy, M. (2019). Situating Data in a Trumpian Era: The Environmental Data and Governance Initiative. Annals of the American Association of Geographers, 109(2), 545–555. https://doi.org/10.1080/24694452.2018.1511410

Donovan, K. (2012). Seeing Like a Slum: Towards Open, Deliberative Development (SSRN Scholarly Paper No. 2045556). https://papers.ssrn.com/abstract=2045556

Dourish, P. (2006). Implications for Design. CHI 2006 Proceedings, 10.

Dunlap, R. E., & Jacques, P. J. (2013). Climate Change Denial Books and Conservative Think Tanks: Exploring the Connection. American Behavioral Scientist, 57(6), 699–731. https://doi.org/10.1177/0002764213477096

Durkin, M., & O'Toole, M. (Directors). (2023, March 23). Climate: The Movie (The Cold Truth) [Video recording].

Eadon, Y. M. (2020). (Not) Part of the System: Resolving Epistemic Disconnect Through Archival Reference. KNOWLEDGE ORGANIZATION, 47(6), 441–460. https://doi.org/10.5771/0943-7444-2020-6-441

Earth Science Data Systems, N. (2021, February 25). Earthdata Cloud Evolution | NASA Earthdata [Program]. Earth Science Data Systems, NASA. https://www.earthdata.nasa.gov/about/earthdata-cloud-evolution

Earth Science Data Systems, N. (2022, August 24). Understanding Needs to Broaden Outside Use of NASA Data (UNBOUND) [Basic Page]. Earth Science Data Systems, NASA. https://www.earthdata.nasa.gov/esds/esds-program-components/unbound

EDGI_Team. (2025, February 12). EPA Removes EJScreen from Its Website. Environmental Data and Governance Initiative. https://envirodatagov.org/epa-removes-ejscreen-from-its-website/

Edwards, P. N. (2010). A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming. MIT Press.

Edwards, P. N. (2011). History of climate modeling. Wiley Interdisciplinary Reviews: Climate Change, 2(1), 128–139. https://doi.org/10.1002/wcc.95

Edwards, P. N. (2019). Knowledge infrastructures under siege: Climate data as memory, truce, and target. In D. Bigo, E. F. Isin, & E. S. Ruppert (Eds.), Data politics: Worlds, subjects, rights (pp. 21–42). Routledge, Taylor & Francis Group.

Edwards, P. N., Jackson, S. J., Chalmers, M. K., Bowker, G. C., Borgman, C. L., Ribes, D., Burton, M., & Calvert, S. (2013). Knowledge Infrastructures: Intellectual Frameworks and Research Challenges. 1–40. http://hdl.handle.net/2027.42/97552.

Edwards, P. N., Mayernik, M. S., Batcheller, A. L., Bowker, G. C., & Borgman, C. L. (2011). Science friction: Data, metadata, and collaboration. Social Studies of Science, 41(5), 667–690.

Emerson, R. M., Fretz, R. I., & Shaw, L. L. (2011). Writing ethnographic fieldnotes (Second edition.). The University of Chicago Press.

Engelbeen, F., Hannon, R., & Burton, D. A. (2024). The Human Contribution to Atmospheric Carbon Dioxide. Open Science Framework. https://doi.org/10.31219/osf.io/het6n

Enke, N., Thessen, A., Bach, K., Bendix, J., Seeger, B., & Gemeinholzer, B. (2012). The user's view on biodiversity data sharing—Investigating facts of acceptance and requirements to realize a sustainable use of research data—. Ecological Informatics, Data Platforms in Integrative Biodiversity Research, 11, 25–33. https://doi.org/10.1016/j.ecoinf.2012.03.004

Eschle, C., & Maiguashca, B. (2010). Making Feminist Sense of the Global Justice Movement. Rowman & Littlefield Publishers.

Espinosa, R., Garriga, L., Zubcoff, J. J., & Mazón, J.-N. (2014). Linked Open Data mining for democratization of big data. 2014 IEEE International Conference on Big Data (Big Data), 17–19. https://doi.org/10.1109/BigData.2014.7004479

Essex, C., McKitrick, R., & Andresen, B. (2007). Does a Global Temperature Exist? Journal of Non-Equilibrium Thermodynamics, 32(1), 1–27. https://doi.org/10.1515/JNETDY.2007.001

Essien, E. O. (2025). Climate Change Disinformation on Social Media: A Meta-Synthesis on Epistemic Welfare in the Post-Truth Era. Social Sciences, 14(5), 304. https://doi.org/10.3390/socsci14050304

Etheridge, D. M., Barnola, J.-M., Morgan, V. I., Steele, L. P., Langenfelds, R. L., & Francey, R. J. (1998). Historical CO2 Records from the Law Dome DE08, DE08-2, and DSS Ice Cores (1006 A.D.-1978 A.D) [Dataset]. Environmental System Science Data Infrastructure for a Virtual Ecosystem; Carbon Dioxide Information Analysis Center (CDIAC); Oak Ridge National Laboratory (ORNL), Oak Ridge, TN (United States). https://doi.org/10.3334/CDIAC/ATG.011

Etheridge, D. M., Steele, L. P., Langenfelds, R. L., Francey, R. J., Barnola, J.-M., & Morgan, V. I. (1996). Natural and anthropogenic changes in atmospheric CO2 over the last 1000 years from air in Antarctic ice and firn. Journal of Geophysical Research: Atmospheres, 101(D2), 4115–4128. https://doi.org/10.1029/95JD03410

Etheridge, D. M., Steele, L. P., Langenfelds, R. L., Francey, R. J., Barnola, J.-M., & Morgan, V. I. (2001). Law Dome Atmospheric CO2 Data (Version Data Contribution Series #2001-083) [Dataset]. IGBP PAGES/World Data Center for Paleoclimatology. https://www.ncei.noaa.gov/pub/data/paleo/icecore/antarctica/law/law_co2.txt

Etheridge, D. M., Steele, L. P., Francey, R. J., & Langenfelds, R. L. (2002). Historic CH4 Records from Antarctic and Greenland Ice Cores, Antarctic Firn Data, and Archived Air Samples from Cape Grim, Tasmania [Dataset]. Environmental System Science Data Infrastructure for a Virtual Ecosystem; Carbon Dioxide Information Analysis Center (CDIAC), Oak Ridge National Laboratory (ORNL), Oak Ridge, TN (United States). https://doi.org/10.3334/CDIAC/ATG.030

Eubanks, V. (2018). Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor. Macmillan + ORM.

Executive Office of the President. (2023). Revitalizing Our Nation's Commitment to Environmental Justice for All (p. 11) [Executive Order]. https://www.federalregister.gov/documents/2023/04/26/2023-08955/revitalizing-our-nations-commitment-to-environmental-justice-for-all

Fall, S., Watts, A., Nielsen-Gammon, J., Jones, E., Niyogi, D., Christy, J. R., & Pielke Sr., R. A. (2011). Analysis of the impacts of station exposure on the U.S. Historical Climatology Network temperatures and temperature trends. Journal of Geophysical Research: Atmospheres, 116(D14). https://doi.org/10.1029/2010JD015146

Faniel, I. M., Austin, A., Kansa, S. W., Kansa, E., Jacobs, J., & France, P. (2020). Identifying Opportunities for Collective Curation During Archaeological Excavations. International Journal of Digital Curation, 15(1), Article 1. https://doi.org/10.2218/ijdc.v15i1.699

Faniel, I. M., & Jacobsen, T. E. (2010). Reusing Scientific Data: How Earthquake Engineering Researchers Assess the Reusability of Colleagues' Data. Journal of Computer Supported Cooperative Work, 19(3–4), 355–375. https://doi.org/10.1007/s10606-010-9117-8

Faniel, I. M., Kansa, E. C., Kansa, S. W., Barrera-Gomez, J., & Yakel, E. (2013). The Challenges of Digging Data: A Study of Context in Archaeological Data Reuse.

Proceedings of the 13th ACM/IEEE-CS Joint Conference on Digital Libraries, JCDL '13, 295–304. https://doi.org/10.1145/2467696.2467712

Faniel, I. M., Kriesberg, A., & Yakel, E. (2012). Data reuse and sensemaking among novice social scientists. Proceedings of the American Society for Information Science and Technology, 49(1), 1–10. https://doi.org/10.1002/meet.14504901068

Faniel, I. M., Kriesberg, A., & Yakel, E. (2016). Social scientists' satisfaction with data reuse. Journal of the Association for Information Science and Technology, 67(6), 1404–1416. https://doi.org/10.1002/asi.23480

Faniel, I. M., & Yakel, E. (2017). Practices Do Not Make Perfect: Disciplinary Data Sharing and Reuse Practices and Their Implications for Repository Data Curation. In L. R. Johnston (Ed.), Curating Research Data, Volume One: Practical Strategies for Your Digital Repository (pp. 103–126). Association of College and Research Libraries. http://www.oclc.org/research/publications/2017/practices-do-not-make-perfect.html

FAOSTAT. (n.d.). Retrieved March 27, 2026, from https://www.fao.org/faostat/en/#data/QCL/metadata

Faundeen, J. L., Burkey, T. E., Carlino, J. A., Govoni, D. L., Henkel, H. S., Holl, S. L., Hutchison, V. B., Martin, E., Montgomery, E. T., Ladino, C. C., Tessler, S., & Zolly, L. S. (2014). The United States Geological Survey Science Data Lifecycle Model (Open-File Report Nos. 2013–1265; Open-File Report, p. 12). U.S. Geological Survey.

Federer, L. M. (2019). Who, what, when, where, and why? Quantifying and understanding biomedical data reuse. https://doi.org/10.13016/60jd-9hux

Field, D., Sansone, S.-A., Collis, A., Booth, T., Dukes, P., Gregurick, S. K., Kennedy, K., Kolar, P., Kolker, E., Maxon, M., Millard, S., Mugabushaka, A.-M., Perrin, N., Remacle, J. E., Remington, K., Rocca-Serra, P., Taylor, C. F., Thorley, M., Tiwari, B., & Wilbanks, J. (2009). 'Omics Data Sharing. Science, 326(5950), 234–236. https://doi.org/10.1126/science.1180598

Findlay, C. (2013). People, records and power: What archives can learn from WikiLeaks 1. Archives and Manuscripts, 41(1), 7–22. https://doi.org/10.1080/01576895.2013.779926

Finn, M. (2018). Documenting aftermath: Information infrastructures in the wake of disasters. MIT Press.

Fiske, A., Degelsegger-Márquez, A., Marsteurer, B., & Prainsack, B. (2023). Value-creation in the health data domain: A typology of what health data help us do. BioSocieties, 18(3), 473–497. https://doi.org/10.1057/s41292-022-00276-6

Force, D., & Wiles, B. (2021). "Quietly Incomplete": Academic Historians, Digital Archival Collections, and Historical Research in the Web Era. Journal of Contemporary Archival Studies, 8(1). https://elischolar.library.yale.edu/jcas/vol8/iss1/18

Frank, R. D., Kriesberg, A., Yakel, E., & Faniel, I. M. (2015). Looting hoards of gold and poaching spotted owls: Data confidentiality among archaeologists & zoologists. Proceedings of the Association for Information Science and Technology, 52(1), 1–10. https://doi.org/10.1002/pra2.2015.145052010037

Frank, R. D., & Wylie, J. (2023). Audit Team Communication and Risk in Trustworthy Digital Repository Certification. Proceedings of the Association for Information Science and Technology, 60(1), 145–155. https://doi.org/10.1002/pra2.776

Freudenburg, W. R., Gramling, R., & Davidson, D. J. (2008). Scientific Certainty Argumentation Methods (SCAMs): Science and the Politics of Doubt. Sociological Inquiry, 78(1), 2–38. https://doi.org/10.1111/j.1475-682X.2008.00219.x

Friedberg, B. (2020). Investigative Digital Ethnography: Methods for Environmental Modeling. In Media Manipulation Casebook. https://mediamanipulation.org/research/investigative-digital-ethnography-methods-environmental-modeling

Friedlingstein, P., O'Sullivan, M., Jones, M. W., Andrew, R. M., Bakker, D. C. E., Hauck, J., Landschützer, P., Le Quéré, C., Luijkx, I. T., Peters, G. P., Peters, W., Pongratz, J., Schwingshackl, C., Sitch, S., Canadell, J. G., Ciais, P., Jackson, R. B., Alin, S. R., Anthoni, P., … Zheng, B. (2023). Global Carbon Budget 2023. Earth System Science Data, 15(12), 5301–5369. https://doi.org/10.5194/essd-15-5301-2023

Gabrielsen, A. M. (2020). Openness and trust in data-intensive science: The case of biocuration. Medicine, Health Care and Philosophy, 23(3), 497–504. https://doi.org/10.1007/s11019-020-09960-5

Galletta, A., & Cross, W. E. (2013). Mastering the Semi-Structured Interview and Beyond: From Research Design to Analysis and Publication. New York University Press. http://ebookcentral.proquest.com/lib/umichigan/detail.action?docID=1187368

Geertz, C. (1973). The Interpretation of Cultures. Basic Books.

Geertz, C. (1998). Deep hanging out. The New York Review of Books, 45(16), 69–72.

Geiger, R. S., & Ribes, D. (2011). Trace Ethnography: Following Coordination through Documentary Practices. 2011 44th Hawaii International Conference on System Sciences, 1–10. https://doi.org/10.1109/HICSS.2011.455

Giddens, A. (1979). Central Problems in Social Theory: Action, Structure, and Contradiction in Social Analysis. University of California Press.

Gilbert, S., Vitak, J., & Shilton, K. (2021). Measuring Americans' Comfort With Research Uses of Their Social Media Data. Social Media + Society, 7(3), 20563051211033824. https://doi.org/10.1177/20563051211033824

GISTEMP Team. (2026). GISS Surface Temperature Analysis (GISTEMP), version 4. NASA Goddard Institute for Space Studies. Journal of Geophysical Research: Atmospheres. https://doi.org/10.1029/2023JD040179

Gitelman, L. (Ed.). (2013). "Raw Data" Is an Oxymoron. MIT Press.

Glaser, B. G., & Strauss, A. L. (1967). The discovery of grounded theory; strategies for qualitative research. Aldine Publishing.

Gramling, C. (2007, October). Geotimes—October 2007—Error in NASA climate data sparks debate. http://www.geotimes.org/oct07/article.html?id=nn_climate.html

Gray, J. (2018). Three Aspects of Data Worlds. Krisis | Journal for Contemporary Philosophy, 38(1), 4–17. https://doi.org/10.21827/krisis.38.1.38827

Green, A. G., & Gutmann, M. P. (2007). Building partnerships among social science researchers, institution-based repositories and domain specific data archives. OCLC Systems & Services: International Digital Library Perspectives, 23(1), 35–53. https://doi.org/10.1108/10650750710720757

Gregory, K., Groth, P., Scharnhorst, A., & Wyatt, S. (2020). Lost or Found? Discovering Data Needed for Research. Harvard Data Science Review. https://doi.org/10.1162/99608f92.e38165eb

Gregory, K. M., Groth, P., Scharnhorst, A., & Wyatt, S. (2023). The Mysterious User of Research Data: Knitting Together Science and Technology Studies with Information and Computer Science. In K. Bijsterveld & A. Swinnen (Eds.), Interdisciplinarity in the Scholarly Life Cycle: Learning by Example in Humanities and Social Science Research (pp. 191–211). Springer International Publishing. https://doi.org/10.1007/978-3-031-11108-2_11

Gregory, K. M., Koesten, L., Schuster, R., Möller, T., & Davies, S. (2024). Data Journeys in Popular Science: Producing Climate Change and COVID-19 Data Visualizations at Scientific American. Harvard Data Science Review, 6(2). https://doi.org/10.1162/99608f92.141c99cf

Gregory, K., & Strecker, D. (2026). Thinking like a business: Reconfiguring relationships to sustain open data infrastructures (arXiv:2604.27580). arXiv. https://doi.org/10.48550/arXiv.2604.27580

Guillou, L. (2020, March 23). Covid-19: How unprecedented data sharing has led to faster-than-ever outbreak research | Research and Innovation. Horizon: The EU Research & Innovation Magazine. https://projects.research-and-innovation.ec.europa.eu/en/horizon-magazine/covid-19-how-unprecedented-data-sharing-has-led-faster-ever-outbreak-research

Gurstein, M. B. (2011). Open data: Empowering the empowered or effective data use for everyone? First Monday, 16(2). http://firstmonday.org/ojs/index.php/fm/article/view/3316

Gutiérrez, M., & Milan, S. (2019). Playing with data and its consequences. First Monday. https://firstmonday.org/ojs/index.php/fm/article/view/9554

H. Sterling Burnett. (n.d.). The Heartland Institute. Retrieved March 26, 2026, from https://heartland.org/about-us/who-we-are/h-sterling-burnett/

Hammersley, M., & Atkinson, P. (2019). Ethnography: Principles in Practice (4th ed.). Routledge. https://doi.org/10.4324/9781315146027

Hanson, B., Sugden, A., & Alberts, B. (2011). Making Data Maximally Available. Science, 331(6018), 649–649. https://doi.org/10.1126/science.1203354

Haraway, D. (1988). Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective. Feminist Studies, 14(3), 575. https://doi.org/10.2307/3178066

Harding, S. (1992). Rethinking standpoint epistemology: What is "strong objectivity?" The Centennial Review, 36(3), 437–470.

Harding, S. (1995). "Strong objectivity": A response to the new objectivity question. Synthese, 104(3), 331–349. https://doi.org/10.1007/BF01064504

Hardy, B. W., & Jamieson, K. H. (2017). Overcoming Endpoint Bias in Climate Change Communication: The Case of Arctic Sea Ice Trends. Environmental Communication, 11(2), 205–217. https://doi.org/10.1080/17524032.2016.1241814

Harvey, P. (2016). Infrastructures and Social Complexity: A Companion (1st ed.). Routledge. https://doi.org/10.4324/9781315622880

Hemphill, L., Pienta, A., Lafia, S., Akmon, D., & Bleckley, D. A. (2022). How do properties of data, their curation, and their funding relate to reuse? Journal of the Association for Information Science and Technology, 73(10), 1432–1444. https://doi.org/10.1002/asi.24646

Hine, C. (2006). Databases as Scientific Instruments and Their Role in the Ordering of Scientific Work. Social Studies of Science, 36(2), 269–298. (ISI:000237058900005). https://doi.org/10.1177/0306312706054047

Hine, C. (2007). Connective Ethnography for the Exploration of e-Science. Journal of Computer-Mediated Communication, 12(2), 618–634. https://doi.org/10.1111/j.1083-6101.2007.00341.x

Hine, C. (2020). Ethnography for the Internet: Embedded, Embodied and Everyday. Routledge. https://doi.org/10.4324/9781003085348

Hinojosa, L., Riedy, R., Polman, J., Swanson, R., Nuessle, T., & Garneau, N. (2021). Expanding Public Participation in Science Practices Beyond Data Collection. Citizen Science: Theory and Practice, 6(1). https://doi.org/10.5334/cstp.292

Hoeppe, G. (2019). Mediating Environments and Objects as Knowledge Infrastructure. Computer Supported Cooperative Work (CSCW), 28(1), 25–59. https://doi.org/10.1007/s10606-018-9342-0

Hoeppe, G. (2021). Encoding Collective Knowledge, Instructing Data Reusers: The Collaborative Fixation of a Digital Scientific Data Set. Computer Supported Cooperative Work (CSCW), 30(4), 463–505. https://doi.org/10.1007/s10606-021-09407-2

Holdren, J. P. (2013, February 22). Increasing Access to the Results of Federally Funded Scientific Research. Executive Office of the President, Office of Science and Technology Policy. https://www.usaid.gov/sites/default/files/documents/1865/NW2-CCBY-HO2-Public_Access_Memo_2013.pdf

Holgate, S. J., Matthews, A., Woodworth, P. L., Rickards, L. J., Tamisiea, M. E., Bradshaw, E., Foden, P. R. F., Gordon, K. M., Jevrejeva, S., & Pugh, J. (2012). New Data Systems and Products at the Permanent Service for Mean Sea Level. Journal of Coastal Research, 29(3), 493. https://doi.org/10.2112/JCOASTRES-D-12-00175.1

Hornsey, M. J., & Lewandowsky, S. (2022). A toolkit for understanding and addressing climate scepticism. Nature Human Behaviour, 6(11), 1454–1464. https://doi.org/10.1038/s41562-022-01463-y

House, N. A. V. (2004). Science and Technology Studies and Information Studies. Annual Review of Information Science and Technology (ARIST), 38, 3–86.

Huang, Y.-H., Rose, P. W., & Hsu, C.-N. (2015). Citing a Data Repository: A Case Study of the Protein Data Bank. PLOS ONE, 10(8), e0136631. https://doi.org/10.1371/journal.pone.0136631

Huberman, M., & Miles, M. B. (2002). The Qualitative Researcher's Companion. SAGE.

IMGW-PIB – Instytut Meteorologii i Gospodarki Wodnej. (n.d.). Retrieved May 26, 2026, from https://imgw.pl/en/

Infrastructure for Integration in Structural Sciences' (I2S2) idealised scientific research activity lifecycle model. (2011). https://www.ukoln.ac.uk/projects/I2S2/documents/I2S2-ResearchActivityLifecycleModel-110407.pdf

Interagency Working Group on Social Cost of Carbon, United States Government (IGW). (2013). Technical Support Document: Technical update of the social cost of carbon for regulatory impact analysis Under Executive Order 12866. https://www.ourenergypolicy.org/wp-content/uploads/2013/06/social_cost_of_carbon_for_ria_2013_update.pdf

Interior Department Restores Gold Standard Science Across Federal Decision-Making | U.S. Department of the Interior. (2025, September 4). [Press Release]. https://www.doi.gov/pressreleases/interior-department-restores-gold-standard-science-across-federal-decision-making

International Panel on the Information Environment (IPIE), Elbeyi, E., Bruhn Jensen, K., Aronczyk, M., Asuka, J., Ceylan, G., Cook, J., Erdelyi, G., Ford, H., Milani, C., Mustafaraj, E., Ogenga, F., Yadin, S., Howard, P., & Valenzuela, S. (2025). Information Integrity about Climate Science: A Systematic Review (R. Brulle, J. Jacquet, S. Lewandowsky, & T. Roberts, Eds.). International Panel on the Information Environment (IPIE). https://doi.org/10.61452/BTZP3426

Irwin, A. (2002). Citizen science: A study of people, expertise and sustainable development. Routledge. https://api.taylorfrancis.com/content/books/mono/download?identifierName=doi&identifierValue=10.4324/9780203202395&type=googlepdf

Jackson, S. J. (2014). Rethinking Repair. In T. Gillespie, P. J. Boczkowski, & K. A. Foot (Eds.), Media technologies: Essays on communication, materiality, and society (pp. 221–240). MIT Press. https://doi.org/10.7551/mitpress/9780262525374.003.0011

Jackson, S. J. (2016). Speed, Time, Infrastructure: Temporalities of Breakdown, Maintenance, and Repair. In J. Wajcman & N. Dodd (Eds.), The Sociology of Speed: Digital, Organizational, and Social Temporalities (pp. 169–186). Oxford University Press.

James, I., & Haggerty, N. (2025, July 1). Trump administration shuts down U.S. website on climate change. Los Angeles Times. https://www.latimes.com/environment/story/2025-07-01/trump-us-climate-website

Jensen, C. B. (2014). Continuous Variations: The Conceptual and the Empirical in STS. Science, Technology, & Human Values, 39(2), 192–213. https://doi.org/10.1177/0162243913502866

Jonas, R. M., & Hanrahan, B. V. (2022). Designing for Shared Values: Exploring Ethical Dilemmas of Conducting Values Inclusive Design Research. Proceedings of the ACM on Human-Computer Interaction, 6(CSCW2), 1–20. https://doi.org/10.1145/3555182

Kaehrle, M., & Eschenfelder, K. (2025). Justifying Biodiversity Data Access Restrictions: A Global Comparison of Data Policies. Proceedings of the Association for Information Science and Technology, 62(1), 324–334. https://doi.org/10.1002/pra2.1259

Kaltenbrunner, W. (2015). Infrastructural Inversion as a Generative Resource in Digital Scholarship. Science as Culture, 24(1), 1–23. https://doi.org/10.1080/09505431.2014.917621

Kansa, E. C. (2012). Openness and archaeology's information ecosystem. World Archaeology, 44(4). https://doi.org/10.1080/00438243.2012.737575

Karasti, H., & Baker, K. S. (2008). Digital data practices and the Long Term Ecological Research program growing global. International Journal of Digital Curation, 3(2), 42–58. https://doi.org/10.2218/ijdc.v3i2.57

Karasti, H., Millerand, F., Hine, C. M., & Bowker, G. C. (2016). Knowledge infrastructures: Part I. Science & Technology Studies, 29(1), 2–12. https://doi.org/10.23987/sts.55406

Kassen, M. (2021). Understanding motivations of citizens to reuse open data: Open government data as a philanthropic movement. Innovation, 23(1), 44–70. https://doi.org/10.1080/14479338.2020.1738940

Key Perspectives Ltd. (2010). Data dimensions: Disciplinary differences in research data sharing, reuse and long term viability. SCARP Synthesis Study. Digital Curation Centre. http://www.dcc.ac.uk/scarp

Khan, N., Thelwall, M., & Kousha, K. (2019). Data citation and reuse practice in biodiversity—Challenges of adopting a standard citation model. In 1. International Society for Scientometrics and Informetrics. https://wlv.openrepository.com/handle/2436/623005

Knorr-Cetina, K. (1999). Epistemic Cultures: How the Sciences Make Knowledge. Harvard University Press.

Koelsch, L. E. (2013). Reconceptualizing the Member Check Interview. International Journal of Qualitative Methods, 12(1), 168–179. https://doi.org/10.1177/160940691301200105

Krippendorff, K. (2018). Content Analysis: An Introduction to Its Methodology. SAGE Publications.

Kuutti, K., & Bannon, L. J. (2014). The turn to practice in HCI: Towards a research agenda. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '14, 3543–3552. https://doi.org/10.1145/2556288.2557111

Kvale, S., & Brinkmann, S. (2009). InterViews: Learning the Craft of Qualitative Research Interviewing. SAGE.

La, V. P., Nguyen, M. H., & Vuong, Q. H. (2024). Climate Change Denial Theories, Skeptical Arguments, and the Role of Science Communication. https://doi.org/10.32388/9WXW9S

Lacy, S., Watson, B. R., Riffe, D., & Lovejoy, J. (2015). Issues and Best Practices in Content Analysis. Journalism & Mass Communication Quarterly, 92(4), 791–811. https://doi.org/10.1177/1077699015607338

Lagoze, C. (2014). eBird: Curating Citizen Science Data for Use by Diverse Communities. International Journal of Digital Curation, 9(1), 71–82. https://doi.org/10.2218/ijdc.v9i1.302

Lämmerhirt, D., Micheli, M., & Schade, S. (2024). Exploring the practices of "data-driven innovation" in the European public sector. Data & Policy, 6, e24. https://doi.org/10.1017/dap.2024.6

Lan, X. (2026). Mauna Loa CH4 Annual Mean Data [Dataset]. NOAA Global Monitoring Laboratory. https://gml.noaa.gov/webdata/ccgg/trends/ch4/ch4_annmean_gl.txt

Lan, X., & Keeling, R. (2026). Mauna Loa CO2 Annual Mean Data [Dataset]. NOAA Global Monitoring Laboratory. https://gml.noaa.gov/webdata/ccgg/trends/co2/co2_annmean_mlo.txt

Lan, X., Tans, P., Thoning, K., & NOAA Global Monitoring Laboratory. (2023). Trends in globally-averaged CO2 determined from NOAA Global Monitoring Laboratory measurements. [Dataset]. NOAA GML. https://doi.org/10.15138/9N0H-ZH07

Landecker, H. (2007). Culturing Life: How Cells Became Technologies. Harvard University Press. https://doi.org/10.4159/9780674039902

Latour, B. (1987). Science in action: How to follow scientists and engineers through society. Harvard University press.

Latour, B., & Woolgar, S. (2013). Laboratory Life: The Construction of Scientific Facts. In Laboratory Life. Princeton University Press. https://doi.org/10.1515/9781400820412

Lay, J. (2022). Climate Disinformation Database. DeSmog. https://www.desmog.com/climate-disinformation-database/

Leahey, E., Entwisle, B., & Einaudi, P. (2003). Diversity in Everyday Research Practice: The Case of Data Editing. Sociological Methods & Research, 32(1), 64–89. https://doi.org/10.1177/0049124103253461

Lee, C., Yang, T., Inchoco, G. D., Jones, G. M., & Satyanarayan, A. (2021). Viral Visualizations: How Coronavirus Skeptics Use Orthodox Data Practices to Promote Unorthodox Science Online. Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, 1–18. https://doi.org/10.1145/3411764.3445211

Lehtiniemi, T., & Ruckenstein, M. (2019). The social imaginaries of data activism. Big Data & Society. https://journals.sagepub.com/doi/full/10.1177/2053951718821146

Leiserowitz, A. A., Maibach, E. W., Roser-Renouf, C., Smith, N., & Dawson, E. (2013). Climategate, Public Opinion, and the Loss of Trust. American Behavioral Scientist, 57(6), 818–837. https://doi.org/10.1177/0002764212458272

Leiserowitz, A., & Smith, N. (2010). Knowledge of Climate Change across Global Warming's Six Americas (New Haven: Yales Project on Climate Change Communication).

Lenssen, N., Schmidt, G. A., Hendrickson, M., Jacobs, P., Menne, M. J., & Ruedy, R. (2024). A NASA GISTEMPv4 Observational Uncertainty Ensemble. Journal of Geophysical

Research: Atmospheres, 129(17), e2023JD040179. https://doi.org/10.1029/2023JD040179

Leonelli, S. (2013a). Global Data for Local Science: Assessing the Scale of Data Infrastructures in Biological and Biomedical Research. BioSocieties, 8(4), 449–465. https://doi.org/10.1057/biosoc.2013.23

Leonelli, S. (2013b). Why the current insistence on open access to scientific data? Big data, knowledge production, and the political economy of contemporary biology. Bulletin of Science, Technology & Society, 33(1–2), 6–11. https://doi.org/10.1177/0270467613496768

Leonelli, S. (2014). What difference does quantity make? On the epistemology of Big Data in biology. Big Data & Society, 1(1), 2053951714534395. https://doi.org/10.1177/2053951714534395

Leonelli, S. (2016). Locating ethics in data science: Responsibility and accountability in global and distributed knowledge production systems. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 374(2083), 20160122. https://doi.org/10.1098/rsta.2016.0122

Leonelli, S. (2020). Learning from Data Journeys. In S. Leonelli & N. Tempini (Eds.), Data Journeys in the Sciences (pp. 1–24). Springer International Publishing. https://doi.org/10.1007/978-3-030-37177-7_1

Leonelli, S. (2023). Philosophy of Open Science (1st ed.). Cambridge University Press. https://doi.org/10.1017/9781009416368

Leonelli, S., Lovell, R., Wheeler, B. W., Fleming, L., & Williams, H. (2021). From FAIR data to fair data use: Methodological data fairness in health-related social media research. Big Data & Society, 8(1), 20539517211010310. https://doi.org/10.1177/20539517211010310

Leonelli, S., & Tempini, N. (Eds.). (2020). Data Journeys in the Sciences. Springer Nature. https://doi.org/10.1007/978-3-030-37177-7

Lewandowsky, S., Ballard, T., Oberauer, K., & Benestad, R. (2016). A blind expert test of contrarian claims about climate data. Global Environmental Change, 39, 91–97. https://doi.org/10.1016/j.gloenvcha.2016.04.013

Lewandowsky, S., Mann, M. E., Brown, N. J. L., & Friedman, H. (2016). Science and the Public: Debate, Denial, and Skepticism. Journal of Social and Political Psychology, 4(2), Article 2. https://doi.org/10.5964/jspp.v4i2.604

Lewandowsky, S., Oreskes, N., Risbey, J. S., Newell, B. R., & Smithson, M. (2015). Seepage: Climate change denial and its effect on the scientific community. Global Environmental Change, 33, 1–13. https://doi.org/10.1016/j.gloenvcha.2015.02.013

Lin, D., Crabtree, J., Dillo, I., Downs, R. R., Edmunds, R., Giaretta, D., De Giusti, M., L'Hours, H., Hugo, W., Jenkyns, R., Khodiyar, V., Martone, M. E., Mokrane, M., Navale, V., Petters, J., Sierman, B., Sokolova, D. V., Stockhause, M., & Westbrook, J. (2020). The TRUST Principles for digital repositories. Scientific Data, 7(1), 144. https://doi.org/10.1038/s41597-020-0486-7

Lincoln, Y. S., & Guba, E. G. (1985). Naturalistic Inquiry. SAGE.

Linden Air Gauge | Measurement Data Portal. (n.d.). [Dataset]. Retrieved May 28, 2026, from https://www.hlnug.de/messwerte/datenportal/messstelle/2/1/1005

Lloyd, E. A. (2012). The role of 'complex' empiricism in the debates about satellite data and climate models. Studies in History and Philosophy of Science Part A, Structures and Strategies in Ancient Greek and Roman Technical Writing, 43(2), 390–401. https://doi.org/10.1016/j.shpsa.2012.02.001

Lloyd, J. (2026, June 1). Who Will Keep Research Data Infrastructure Open and Running? | Real Numbers. Issues in Science and Technology. https://issues.org/open-data-research-infrastructure-gibson-thaney/

Lowenberg, D., Chodacki, J., Fenner, M., Kemp, J., & Jones, M. B. (2019). Open Data Metrics: Lighting the Fire (Version 1). Zenodo. https://doi.org/10.5281/ZENODO.3525349

Luka, M. E., & Millette, M. (2018). (Re)framing Big Data: Activating Situated Knowledges and a Feminist Ethics of Care in Social Media Research. Social Media + Society, 4(2), 205630511876829. https://doi.org/10.1177/2056305118768297

Lupton, D. (2015). Personal Data Practices in the Age of Lively Data (SSRN Scholarly Paper ID 2636709). Social Science Research Network. https://doi.org/10.2139/ssrn.2636709

MacNeil, H. (2001). Trusting Records in a Postmodern World. Archivaria, 36–47.

Madill, A., & Sullivan, P. (2018). Mirrors, portraits, and member checking: Managing difficult moments of knowledge exchange in the social sciences. Qualitative Psychology, 5(3), 321–339. https://doi.org/10.1037/qup0000089

Malterud, K., Siersma, V. D., & Guassora, A. D. (2016). Sample Size in Qualitative Interview Studies: Guided by Information Power. Qualitative Health Research, 26(13), 1753–1760. https://doi.org/10.1177/1049732315617444

Mapping Police Violence. (2021). Mapping Police Violence. https://mappingpoliceviolence.org

Margolis, R., Derr, L., Dunn, M., Huerta, M., Larkin, J., Sheehan, J., Guyer, M., & Green, E. D. (2014). The National Institutes of Health's Big Data to Knowledge (BD2K) initiative: Capitalizing on biomedical big data. Journal of the American Medical Informatics Association, 21(6), 957–958. https://doi.org/10.1136/amiajnl-2014-002974

Marwick, A., & Partin, W. C. (2022). Constructing Alternative Facts: Populist Expertise and the QAnon Conspiracy. SocArXiv. https://doi.org/10.31235/osf.io/ru4b8

Mash, M., Rosenthal, S., & Simmons, R. (2021). DSWorkFlow: A Framework for Capturing Data Scientists' Workflows. Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems, CHI EA '21, 1–7. https://doi.org/10.1145/3411763.3451683

Matthews, P. (2015). Why Are People Skeptical about Climate Change? Some Insights from Blog Comments. Environmental Communication, 9(2), 153–168. https://doi.org/10.1080/17524032.2014.999694

McCright, A. M., & Dunlap, R. E. (2011a). Cool dudes: The denial of climate change among conservative white males in the United States. Global Environmental Change, 21(4), 1163–1172. https://doi.org/10.1016/j.gloenvcha.2011.06.003

McCright, A. M., & Dunlap, R. E. (2011b). The Politicization of Climate Change and Polarization in the American Public's Views of Global Warming, 2001–2010. The Sociological Quarterly, 52(2), 155–194. https://doi.org/10.1111/j.1533-8525.2011.01198.x

McKemmish, S., Faulkhead, S., & Russell, L. (2011). Distrust in the archive: Reconciling records. Archival Science, 11(3), 211–239. https://doi.org/10.1007/s10502-011-9153-2

McKitrick, R. (n.d.). Your Environment: The Complete Canadian Environmental Record. Retrieved June 9, 2026, from https://www.yourenvironment.ca/

McKitrick, R. (2025). Extended crop yield meta-analysis data do not support upward SCC revision. Scientific Reports, 15(1), 5575. https://doi.org/10.1038/s41598-025-90254-2

Meeuwig, J. J., Harcourt, R. G., & Whoriskey, F. G. (2015). When Science Places Threatened Species at Risk. Conservation Letters, 8(3), 151–152. https://doi.org/10.1111/conl.12185

Menne, M. J., Williams, C. N., Gleason, B. E., Rennie, J. J., & Lawrimore, J. H. (2018). The Global Historical Climatology Network Monthly Temperature Dataset, Version 4. Journal of Climate, 31(24), 9835–9854. https://doi.org/10.1175/JCLI-D-18-0094.1

Meyer, P. (2023). Comment on 'Climate sensitivity, agricultural productivity and the social cost of carbon in FUND.' Environmental Economics and Policy Studies, 25(2), 285–290. https://doi.org/10.1007/s10018-022-00355-9

Mirowski, P. (2018). The future(s) of open science. Social Studies of Science, 48(2), 171–203. https://doi.org/10.1177/0306312718772086

Moore, F. C., Baldos, U., Hertel, T., & Diaz, D. (2017). New science of climate change impacts on agriculture implies higher social cost of carbon. Nature Communications, 8(1), 1607. https://doi.org/10.1038/s41467-017-01792-x

Morice, C. P., Kennedy, J. J., Rayner, N. A., & Jones, P. D. (2012). Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set: THE HADCRUT4 DATASET. Journal of Geophysical Research: Atmospheres, 117(D8), n/a-n/a. https://doi.org/10.1029/2011JD017187

Morice, C. P., Kennedy, J. J., Rayner, N. A., Winn, J. P., Hogan, E., Killick, R. E., Dunn, R. J. H., Osborn, T. J., Jones, P. D., & Simpson, I. R. (2021). An Updated Assessment of Near-Surface Temperature Change From 1850: The HadCRUT5 Data Set. Journal of Geophysical Research: Atmospheres, 126(3). https://doi.org/10.1029/2019JD032361

Mosbergen, D., McKay, B., Kusisto, L., & Whitton, B. (2025, January 31). Vaccine Information, Transgender References Disappear From Federal Websites. Wall Street Journal. https://www.wsj.com/health/healthcare/federal-agencies-purge-transgender-references-from-their-websites-0651103e

Moss, E., & Lyle, J. (2018). Opaque data citation: Actual citation practice and its implication for tracking data use. http://deepblue.lib.umich.edu/handle/2027.42/142393

Moss, M. S. (2011). Is it a question of trust or why are we afraid to go to Nineveh? Archival Science, 11(3), 409–425. https://doi.org/10.1007/s10502-011-9149-y

Muller, M., Lange, I., Wang, D., Piorkowski, D., Tsay, J., Liao, Q. V., Dugan, C., & Erickson, T. (2019). How Data Science Workers Work with Data: Discovery, Capture, Curation, Design, Creation. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, CHI '19, 1–15. https://doi.org/10.1145/3290605.3300356

Muller, M., & Strohmayer, A. (2022). Forgetting Practices in the Data Sciences. Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, CHI '22, 1–19. https://doi.org/10.1145/3491102.3517644

Muravyov, D. (2024). Doubt to be certain: Epistemological ambiguity of data in the case of grassroots mapping of traffic accidents in Russia. Social Movement Studies, 23(3), 338–354. https://doi.org/10.1080/14742837.2022.2128327

NASA GISS. (2007). Global Mean CO2 Mixing Ratios (ppm): Observations [Dataset]. NASA GISS. https://data.giss.nasa.gov/modelforce/ghgases/Fig1A.ext.txt

Nelson, T. (n.d.). Tom Nelson Podcast. YouTube. Retrieved May 31, 2026, from https://www.youtube.com/channel/UCfZS_wFmJCXqPr4MYtAIN6w

Neuman, W. L. (2014). Social research methods: Qualitative and quantitative approaches (Seventh edition, Pearson new international edition). Pearson.

Newman, I., & Benz, C. R. (1998). Qualitative-quantitative Research Methodology: Exploring the Interactive Continuum. SIU Press.

Nielsen, M. A. (2011). Reinventing discovery: The new era of networked science. Princeton University Press.

NOAA Center for Operational Oceanographic Products and Services. (2026). Local Relative Sea Level Trends [Dataset]. NOAA Tides and Currents. https://tidesandcurrents.noaa.gov/sltrends/sltrends.html

NOAA CO-OPS. (n.d.). Swinoujscie, Poland Monthly Mean Sea Levels [Dataset]. NOAA Tides and Currents. Retrieved May 26, 2026, from https://sealevel.info/viewtxt.php?https://tidesandcurrents.noaa.gov/sltrends/data/110-092_meantrend.csv

NOAA CO-OPS. (2026). CO-OPS Data Retrieval API [Dataset]. NOAA Tides and Currents. https://api.tidesandcurrents.noaa.gov/api/prod/

Obama, B. (2013, May 9). Executive Order—Making Open and Machine Readable the New Default for Government Information. https://obamawhitehouse.archives.gov/the-press-office/2013/05/09/executive-order-making-open-and-machine-readable-new-default-government-

O'Brien, M., Duerr, R., Taitingfong, R., Martinez, A., Vera, L., Jennings, L. L., Downs, R. R., Antognoli, E., Brink, T. T., Halmai, N. B., David-Chavez, D., Carroll, S. R., Hudson, M., & Buttigieg, P. L. (2024). Earth Science Data Repositories: Implementing the CARE Principles. Data Science Journal, 23. https://doi.org/10.5334/dsj-2024-037

Office of The Director, National Institutes of Health. (2023). Final NIH Policy for Data Management and Sharing (NOT-OD-21-013:). National Institutes of Health. https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-013.html

O'Grady, M., & Mangina, E. (2024). Citizen scientists—Practices, observations, and experience. Humanities and Social Sciences Communications, 11(1), 469. https://doi.org/10.1057/s41599-024-02966-x

O'Leary, L. (2025, February 4). Unfortunately, the Economy Runs On the Data Trump Is Trying to Delete. Slate. https://slate.com/business/2025/02/donald-trump-data-deletion-jobs-report-economy-public-health.html

OOI. (2026, June 18). Announcement on NSF OOI Descoping. Ocean Observatories Initiative. https://oceanobservatories.org/2026/05/announcement-on-ooi-descoping/

Open Knowledge Foundation. (n.d.). Open Definition 2.1—Open Definition—Defining Open in Open Data, Open Content and Open Knowledge. Retrieved April 18, 2023, from https://opendefinition.org/od/2.1/en/

Oreskes, N., & Conway, E. M. (2011). Merchants of Doubt: How a Handful of Scientists Obscured the Truth on Issues from Tobacco Smoke to Global Warming. Bloomsbury Publishing USA.

OSTP. (2022). Desirable Characteristics of Data Repositories for Federally Funded Research. Executive Office of the President of the United States. https://doi.org/10.5479/10088/113528

Oviedo-García, M. Á. (2021). Journal citation reports and the definition of a predatory journal: The case of the Multidisciplinary Digital Publishing Institute (MDPI). Research Evaluation, 30(3), 405–419a. https://doi.org/10.1093/reseval/rvab020

Page Not Found | US EPA. (n.d.). Retrieved May 27, 2026, from https://www.epa.gov/climate-indicators/climate-change-indicators-heat-waves#%20

Painter, J., & Ashe, T. (2012). Cross-national comparison of the presence of climate scepticism in the print media in six countries, 2007–10. Environmental Research Letters, 7(4), 044005. https://doi.org/10.1088/1748-9326/7/4/044005

Palmer, C. L., & Cragin, M. H. (2008). Scholarship and disciplinary practices. Annual Review of Information Science and Technology, 42(1), 163–212. https://doi.org/10.1002/aris.2008.1440420112

Palmer, C. L., Teffeau, L. C., & Pirman, C. M. (2009). Scholarly Information Practices in the Online Environment: Themes from the Literature and Implications for Library Service Development.

Park, H., You, S., & Wolfram, D. (2018). Informal data citation for data sharing and reuse is more common than formal data citation in biomedical fields. Journal of the Association for Information Science and Technology, 69(11), 1346–1354. https://doi.org/10.1002/asi.24049

Parker, W. S. (2020). Evaluating Data Journeys: Climategate, Synthetic Data and the Benchmarking of Methods for Climate Data Processing. In S. Leonelli & N. Tempini (Eds.), Data Journeys in the Sciences (pp. 191–206). Springer International Publishing. https://doi.org/10.1007/978-3-030-37177-7_10

Parmiggiani, E. (2015). Integration by infrastructuring: The case of subsea environmental monitoring in oil and gas offshore operations [NTNU]. https://ntnuopen.ntnu.no/ntnu-xmlui/bitstream/handle/11250/2358470/Elena_Parmiggiani.pdf?sequence=4

Parmiggiani, E., Monteiro, E., & Hepsø, V. (2015). The Digital Coral: Infrastructuring Environmental Monitoring. Computer Supported Cooperative Work (CSCW), 24(5), 423–460. https://doi.org/10.1007/s10606-015-9233-6

Parsons, M. A. (2011). Making data useful for modelers to understand complex Earth systems. Earth Science Informatics, 4(4), 197–223. https://doi.org/10.1007/s12145-011-0089-0

Pasquetto, I. V., Borgman, C. L., & Wofford, M. F. (2019). Uses and Reuses of Scientific Data: The Data Creators' Advantage. Harvard Data Science Review, 1(2). https://doi.org/10.1162/99608f92.fc14bf2d

Pasquetto, I. V., Cullen, Z., Thomer, A., & Wofford, M. (2024). What is research data "misuse"? And how can it be prevented or mitigated? Journal of the Association for Information Science and Technology, n/a(n/a). https://doi.org/10.1002/asi.24944

Pasquetto, I. V., Randles, B. M., & Borgman, C. L. (2017). On the Reuse of Scientific Data. Data Science Journal, 16. https://doi.org/10.5334/dsj-2017-008

Pasquetto, I. V., Wofford, M., Jordan, C., & Thomer, A. (2026). Opposing consensus science through scholarly practices: The role of claims maintenance. Journal of the Association for Information Science and Technology, n/a(n/a). https://doi.org/10.1002/asi.70074

Pei, L., Olgado, B. S., & Crooks, R. (2022). Narrativity, Audience, Legitimacy: Data Practices of Community Organizers. Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems, CHI EA '22, 1–6. https://doi.org/10.1145/3491101.3519673

PEN America. (2019, May 24). Prepare for Online Harassment. Online Harassment Field Manual. https://onlineharassmentfieldmanual.pen.org/prepare-for-online-harassment/

PEN America. (2023). PEN America's Digital Safety Checklist [Oct 2023]. Google Docs. https://docs.google.com/document/d/1AFJHDUO9zvz3hDK6NP1e14lhTjzBGcQNKPGpdfd5Mv4/edit?usp=embed_facebook

Peters, I., Kraker, P., Lex, E., Gumpenberger, C., & Gorraiz, J. I. (2017). Zenodo in the Spotlight of Traditional and New Metrics. Frontiers in Research Metrics and Analytics, 2. https://doi.org/10.3389/frma.2017.00013

Phillips, W. (2018). The Oxygen of Amplification: Better Practices for Reporting on Extremists, Antagonists, and Manipulators Online. Data and Society.

Pienta, A. M., Alter, G. C., & Lyle, J. A. (2010). The Enduring Value of Social Science Research: The Use and Reuse of Primary Research Data. 1–31.

Pink, S., Horst, H., Postill, J., Hjorth, L., Lewis, T., & Tacchi, J. (2015). Digital Ethnography: Principles and Practice. SAGE.

Piwowar, H. A., & Vision, T. J. (2013). Data reuse and the open data citation advantage. PeerJ, 1, e175. https://doi.org/10.7717/peerj.175

Plantin, J.-C., Lagoze, C., & Edwards, P. N. (2018). Re-integrating scholarly infrastructure: The ambiguous role of data sharing platforms. Big Data & Society, 5(1), 2053951718756683. https://doi.org/10.1177/2053951718756683

Plantin, J.-C., Lagoze, C., Edwards, P. N., & Sandvig, C. (2016). Infrastructure studies meet platform studies in the age of Google and Facebook: New Media & Society. (Sage UK: London, England). https://doi.org/10.1177/1461444816661553

Powell, J. L. (2015). Climate Scientists Virtually Unanimous: Anthropogenic Global Warming Is True. Bulletin of Science, Technology & Society, 35(5–6), 121–124. https://doi.org/10.1177/0270467616634958

Power, M. (1997). The Audit Society: Rituals of Verification. OUP Oxford.

Prasad, A. (2019). Denying Anthropogenic Climate Change: Or, How Our Rejection of Objective Reality Gave Intellectual Legitimacy to Fake News. Sociological Forum, 34(S1), 1217–1234. https://doi.org/https://doi.org/10.1111/socf.12546

Press, A. (2025, May 8). Noaa to stop tracking cost of climate crisis-fueled disasters: 'Major loss.' The Guardian. https://www.theguardian.com/us-news/2025/may/08/noaa-climate-crisis-extreme-weather-disasters-trump

Preston, S., Rajagopalan, K., Yourek, M., Kalcsits, L., & Singh, D. (2024). Changing climate risks for high-value tree fruit production across the United States. Environmental Research Letters, 19(12), 124092. https://doi.org/10.1088/1748-9326/ad90f4

PSMSL. (n.d.-a). Data and Station Information for SWINOUJSCIE. Retrieved March 25, 2026, from https://psmsl.org/data/obtaining/stations/2.php

PSMSL. (n.d.-b). Referencing the Tide Gauge Data Set. PSMSL. Data Repository. Retrieved March 24, 2026, from https://psmsl.org/data/obtaining/reference.php

PSMSL. (n.d.-c). Swinoujscie, Poland Unadjusted Monthly Sea Levels [Dataset]. Retrieved May 26, 2026, from https://psmsl.org/data/obtaining/rlr.monthly.data/2.rlrdata

Radin, J. (2017). "Digital Natives": How Medical and Indigenous Histories Matter for Big Data. Osiris, 32(1), 43–64. https://doi.org/10.1086/693853

Rajan, K. S. (2021). Multisituated: Ethnography as Diasporic Praxis. Duke University Press.

Rakova, B., & Dobbe, R. (2023). Algorithms as Social-Ecological-Technological Systems: An Environmental Justice Lens on Algorithmic Audits. 2023 ACM Conference on Fairness, Accountability, and Transparency, 491–491. https://doi.org/10.1145/3593013.3594014

Ramírez-i-Ollé, M. (2019). Into the woods: An epistemography of climate change. Manchester University Press. https://www.manchesterhive.com/view/9781526140999/9781526140999.xml

Rannard, G. (2020, January 7). Australia fires: Misleading maps and pictures go viral. https://www.bbc.com/news/blogs-trending-51020564

Rayburn, A. J., & Thomer, A. K. (2022). The craft of database curation: Taking cues from quiltmaking. https://hdl.handle.net/2142/113740

Reckwitz, A. (2002). Toward a Theory of Social Practices: A Development in Culturalist Theorizing. European Journal of Social Theory, 5(2), 243–263. https://doi.org/10.1177/13684310222225432

Reilly, P., & Trevisan, F. (2016). Researching protest on Facebook: Developing an ethical stance for the study of Northern Irish flag protest pages. Information, Communication & Society, 19(3), 419–435. https://doi.org/10.1080/1369118X.2015.1104373

Researcher Support Consortium. (2024a). For Researchers: Mitigating Risk and Coping During an Attack. https://researchersupport.org/researchers-mitigate-risk-and-cope-during-an-attack/

Researcher Support Consortium. (2024b). Toolkit for Institutions (No. 2). Researcher Support Consortium. https://researchersupport.org/wp-content/uploads/2024/09/RSC_Toolkit_V2.pdf

Restoring Gold Standard Science. (2025, May 29). Federal Register. https://www.federalregister.gov/documents/2025/05/29/2025-09802/restoring-gold-standard-science

Reyenga, P. J., Howden, S. M., Meinke, H., & McKeon, G. M. (1999). Modelling global change impacts on wheat cropping in south-east Queensland, Australia. Environmental Modelling & Software, 14(4), 297–306. https://doi.org/10.1016/S1364-8152(98)00081-4

Ribes, D., & Lee, C. P. (2010). Sociotechnical Studies of Cyberinfrastructure and e-Research: Current Themes and Future Trajectories. Computer Supported Cooperative Work (CSCW), 19(3), 231–244. https://doi.org/10.1007/s10606-010-9120-0

Ricker, B., Cinnamon, J., & Dierwechter, Y. (2020). When open data and data activism meet: An analysis of civic participation in Cape Town, South Africa. The Canadian Geographer / Le Géographe Canadien, 64(3), 359–373. https://doi.org/10.1111/cag.12608

Rimfrost Experience. (n.d.). Retrieved June 7, 2026, from http://rimfrost.no/rimfrost/ourexperience.html

Robinson-García, N., Jiménez-Contreras, E., & Torres-Salinas, D. (2016). Analyzing data citation practices using the data citation index. Journal of the Association for Information Science and Technology, 67(12), 2964–2975. https://doi.org/https://doi.org/10.1002/asi.23529

Rohde, R., Muller, R. A., Jacobsen, R., Muller, E., & Wickham, C. (2013). A New Estimate of the Average Earth Surface Land Temperature Spanning 1753 to 2011. Geoinformatics & Geostatistics: An Overview, 01(01). https://doi.org/10.4172/2327-4581.1000101

Rosenzweig, C., & Parry, M. L. (1994). Potential impact of climate change on world food supply. Nature, 367(6459), 133–138. https://doi.org/10.1038/367133a0

Rubin, H. J., & Rubin, I. S. (2011). Qualitative Interviewing: The Art of Hearing Data. SAGE.

Ryan, P. D. (2019, January 14). H.R.4174 - 115th Congress (2017-2018): Foundations for Evidence-Based Policymaking Act of 2018 [Legislation]. (2017/2018). http://www.congress.gov/

Sands, A. E., Borgman, C. L., Wynholds, L. A., & Traweek, S. (2012, October 29). Follow the data: How astronomers use and reuse data [Poster]. ASIS&T 75th Annual Meeting, Baltimore, MD.

Sandvig, C., Hamilton, K., Karahalios, K., & Langbort, C. (2014). An algorithm audit. Data and Discrimination: Collected Essays, 6–10.

Santarsiero, R. (2025, September 30). Disappearing Data, Part II: Distorted Science and Deregulation | National Security Archive. National Security Archive. https://nsarchive.gwu.edu/briefing-book/climate-change-transparency-project/2025-09-30/disappearing-data-part-ii-distorted

Saunders, B., Sim, J., Kingstone, T., Baker, S., Waterfield, J., Bartlam, B., Burroughs, H., & Jinks, C. (2018). Saturation in qualitative research: Exploring its conceptualization and operationalization. Quality & Quantity, 52(4), 1893–1907. https://doi.org/10.1007/s11135-017-0574-8

Sayogo, D. S., & Pardo, T. A. (2013). Exploring the determinants of scientific data sharing: Understanding the motivation to publish research data. Government Information Quarterly, ICEGOV 2011 Supplement, 30, S19–S31. https://doi.org/10.1016/j.giq.2012.06.011

Schalkwyk, F. V., Canares, M., Chattapadhyay, S., & Andrason, A. (2015). Open Data Intermediaries in Developing Countries. 640993 Bytes. https://doi.org/10.6084/M9.FIGSHARE.1449222

Schatzki, T. R. (Ed.). (2005). The practice turn in contemporary theory (Online-Ausg). Taylor and Francis.

Schreier, M. (2012). Qualitative Content Analysis in Practice. SAGE Publications Ltd. https://doi.org/10.4135/9781529682571

Schwartz-Shea, P. (2020). "Member-Checking": Not a Panacea, Sometimes a Quagmire. https://doi.org/10.5281/ZENODO.3946815

Scientific Reports. (2026, August 6). Nature. https://www.nature.com/srep/

Seta, G. de. (2020). Three lies of digital ethnography. Journal of Digital Social Research, 2(1), Article 1. https://doi.org/10.33621/jdsr.v2i1.24

Shankar, K., Eschenfelder, K. R., & Downey, G. (2016). Studying the History of Social Science Data Archives as Knowledge Infrastructure. Science & Technology Studies, 29(2). http://ojs.tsv.fi/index.php/sts/article/view/55691

Shankar, K., Jeng, W., Thomer, A., Weber, N., & Yoon, A. (2021). Data curation as collective action during Covid-19. Journal of the Association for Information Science and Technology, 72(3), 280–284. https://doi.org/10.1002/asi.24406

Sharman, A. (2014). Mapping the climate sceptical blogosphere. Global Environmental Change, 26, 159–170. https://doi.org/10.1016/j.gloenvcha.2014.03.003

Shove, E., Watson, M., & Pantzar, M. (2012). The Dynamics of Social Practice: Everyday Life and how it Changes. 1–208.

Simonsen, J., Karasti, H., & Hertzum, M. (2020). Infrastructuring and Participatory Design: Exploring Infrastructural Inversion as Analytic, Empirical and Generative. Computer Supported Cooperative Work (CSCW), 29(1–2), 115–151. https://doi.org/10.1007/s10606-019-09365-w

Singer, E. (2025, February 2). Thousands of U.S. Government Web Pages Have Been Taken Down Since Friday. The New York Times. https://www.nytimes.com/2025/02/02/upshot/trump-government-websites-missing-pages.html

Singh, M. K. (2025). Men deny more than they believe about climate change on Twitter (X). PLOS ONE, 20(2), e0303007. https://doi.org/10.1371/journal.pone.0303007

Soden, R., Agrawaal, T. S., Lord, A., Chanen, C., Flawn, L., Seaifan, Z., Classens, M., & Easterbrook, S. (2025). Climate Data Practices: A Research Approach for HCI and Climate Justice. ACM Trans. Comput.-Hum. Interact., 32(4), 34:1-34:27. https://doi.org/10.1145/3719346

Soden, R., & Kauffman, N. (2019). Infrastructuring the Imaginary: How Sea-Level Rise Comes to Matter in the San Francisco Bay Area. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, 1–11. https://doi.org/10.1145/3290605.3300516

Solomun, S., Monroy, I. B., Bugiel, J., Chan, E., Gowd, N., Hayes, H. A., Jayme, N. H., Kim, S., Tollefson, H., & Ross, C. (2025). Climate Obstruction: The State and Spread of Climate Disinformation in Canada. Centre for Media, Technology, and Democracy. mediatechdemocracy.com/climate-obstruction-report

Spencer, R. W., Christy, J. R., & Braswell, W. D. (2025). Urban Heat Island Effects in U.S. Summer Surface Temperature Data, 1895–2023. Journal of Applied Meteorology and Climatology, 64(7), 717–728. https://doi.org/10.1175/JAMC-D-23-0199.1

Spradley, J. P. (1980). Participant Observation. Holt, Rinehart and Winston.

Srinivasan, J., Finn, M., & Ames, M. (2017). Information determinism: The consequences of the faith in information. The Information Society, 33(1), 13–22. https://doi.org/10.1080/01972243.2016.1248613

Star, S. L. (1983). Simplification in Scientific Work: An Example from Neuroscience Research. Social Studies of Science, 13(2), 205–228. https://doi.org/10.1177/030631283013002002

Star, S. L., & Ruhleder, K. (1996). Steps Toward an Ecology of Infrastructure: Design and Access for Large Information Spaces. Information Systems Research, 7(1), 111–134. https://doi.org/10.1287/isre.7.1.111

Stone, W. (2025, February 6). Some federal health websites restored, others still down, after data purge. NPR. https://www.npr.org/sections/shots-health-news/2025/02/06/nx-s1-5288113/cdc-website-health-data-trump

Strathern, M. (Ed.). (2000). Audit cultures: Anthropological studies in accountability, ethics, and the academy. Routledge.

Stvilia, B., Hinnant, C. C., Wu, S., Worrall, A., Lee, D. J., Burnett, K., Burnett, G., Kazmer, M. M., & Marty, P. F. (2013). Studying the data practices of a scientific community. Proceedings of the 13th ACM/IEEE-CS Joint Conference on Digital Libraries, JCDL '13, 425–426. https://doi.org/10.1145/2467696.2467781

Surface Stations. (2021). surfacestations.org

Takahashi, T., Sutherland, S. C., Sweeney, C., Poisson, A., Metzl, N., Tilbrook, B., Bates, N., Wanninkhof, R., Feely, R. A., Sabine, C., Olafsson, J., & Nojiri, Y. (2002). Global sea–air CO2 flux based on climatological surface ocean pCO2, and seasonal biological and temperature effects. Deep Sea Research Part II: Topical Studies in Oceanography, 49(9–10), 1601–1622. https://doi.org/10.1016/S0967-0645(02)00003-6

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., Manoff, M., & Frame, M. (2011). Data Sharing by Scientists: Practices and Perceptions. PLOS ONE, 6(6), e21101. https://doi.org/10.1371/journal.pone.0021101

Tenopir, C., Dalton, E. D., Allard, S., Frame, M., Pjesivac, I., Birch, B., Pollock, D., & Dorsett, K. (2015). Changes in Data Sharing and Data Reuse Practices and Perceptions among Scientists Worldwide. PLOS ONE, 10(8), e0134826. https://doi.org/10.1371/journal.pone.0134826

Tenopir, C., Palmer, C. L., Metzer, L., van der Hoeven, J., & Malone, J. (2011). Sharing data: Practices, barriers, and incentives. Proceedings of the American Society for Information Science and Technology, 48(1), 1–4. https://doi.org/10.1002/meet.2011.14504801026

Testimony of John R. Christy: Hearing before the Science, Space and Technology, House (2017). https://science.house.gov/_cache/files/5/6/56b2c90e-acc2-4cab-bb10-a510d3cb43ac/AD54FE912F5E3094C8B391DA314D1E4C.hhrg-115-sy-wstate-jchristy-20170329.pdf

Thomer, A. K. (2022). Integrative data reuse at scientifically significant sites: Case studies at Yellowstone National Park and the La Brea Tar Pits. Journal of the Association for Information Science and Technology, 73(8), 1155–1170. https://doi.org/10.1002/asi.24620

Thomer, A. K., Akmon, D., York, J. J., Tyler, A. R. B., Polasek, F., Lafia, S., Hemphill, L., & Yakel, E. (2022). The Craft and Coordination of Data Curation: Complicating Workflow Views of Data Science. Proceedings of the ACM on Human-Computer Interaction, 6(CSCW2), 414:1-414:29. https://doi.org/10.1145/3555139

Thomer, A. K., & Wickett, K. M. (2020). Relational data paradigms: What do we learn by taking the materiality of databases seriously? Big Data & Society, 7(1), 2053951720934838. https://doi.org/10.1177/2053951720934838

Thomer, A. K., Wofford, M. F., Lenard, M. C., Dominguez Vidana, S., & Goring, S. J. (2023). Revealing Earth science code and data-use practices using the Throughput Graph Database. In X. Ma, M. Mookerjee, L. Hsu, & D. Hills (Eds.), Recent Advancements in Geoinformatics and Data Science. Geological Society of America. https://doi.org/10.1130/2022.2558(10)

Tide Gauge Data. (2026). [Dataset]. Permanent Service for Mean Sea Level. http://www.psmsl.org/data/obtaining/

Tiidenberg, K. (2018). Research Ethics, Vulnerability, and Trust on the Internet. In J. Hunsinger, L. Klastrup, & M. M. Allen (Eds.), Second International Handbook of Internet Research (pp. 1–15). Springer Netherlands. https://doi.org/10.1007/978-94-024-1202-4_55-1

Tom Nelson. (n.d.). CO2 Coalition. Retrieved April 9, 2026, from https://co2coalition.org/teammember/tom-nelson/

Tran, A.-T., Boone, A., Le Dantec, C. A., & DiSalvo, C. (2022). Careful Data Tinkering. Proceedings of the ACM on Human-Computer Interaction, 6(CSCW2), 1–29. https://doi.org/10.1145/3555532

Treen, K., Williams, H., O'Neill, S., & Coan, T. G. (2022). Discussion of Climate Change on Reddit: Polarized Discourse or Deliberative Debate? Environmental Communication, 16(5), 680–698. https://doi.org/10.1080/17524032.2022.2050776

Tucci, G., Bastos, J. G., & Carneiro, B. (2025). Contesting Climate Futures: Mapping misinformation narratives and networks on Telegram. https://doi.org/10.13140/rg.2.2.27370.22724

Udasin, S. (2025, January 6). Climate change is hampering US apple quality and output: Study [Text]. The Hill. https://thehill.com/policy/energy-environment/5069348-climate-change-impact-apple-growth-study/

US Environmental Protection Agency. (2023). Supplementary Material for the Regulatory Impact Analysis for the Final Rulemaking, "Standards of Performance for New, Reconstructed, and Modified Sources and Emissions Guidelines for Existing Sources: Oil and Natural Gas Sector Climate Review", EPA Report on the Social Cost of Greenhouse Gases: Estimates Incorporating Recent Scientific Advances, November.

US EPA, O. (2025, July 18). EPA Announces Reduction in Force, Reorganization Efforts to Save Taxpayers Nearly Three-Quarters of a Billion Dollars [News Release]. https://www.epa.gov/newsreleases/epa-announces-reduction-force-reorganization-efforts-save-taxpayers-nearly-three

U.S. Heatwaves. (n.d.). Retrieved March 27, 2026, from https://climateataglance.com/climate-at-a-glance-u-s-heatwaves/

Van House, N. A., Butler, M. H., & Schiff, L. R. (1998). Cooperative knowledge work and practices of trust: Sharing environmental planning data sets. Proceedings of the 1998 ACM Conference on Computer Supported Cooperative Work, 335–343. https://doi.org/10.1145/289444.289508

Van Tuyl, S., & Whitmire, A. (2018). Investigation of Non-Academic Data Management Practices to Inform Academic Research Data Management. Research Ideas and Outcomes, 4, e30829. https://doi.org/10.3897/rio.4.e30829

Vera, L. A., Dillon, L., Wylie, S., Ohayon, J. L., Lemelin, A., Brown, P., Sellers, C., Walker, D., & the Environmental Data and Governance Initiative. (2018). Data Resistance: A social movement organizational autoethnography of the environmental data and governance initiative. Mobilization: An International Quarterly, 23(4), 511–529. https://doi.org/10.17813/1086-671X-24-4-511

Vera, L. A., Walker, D., Murphy, M., Mansfield, B., Siad, L. M., & Ogden, J. (2019). When Data Justice and Environmental Justice Meet: Formulating a Response to Extractive Logic through Environmental Data Justice. Information, Communication and Society, 22(7), 1012–1028. https://doi.org/10.1080/1369118X.2019.1596293

Verhulst, S. G. (2021). Reimagining data responsibility: 10 new approaches toward a culture of trust in re-using data to address critical public needs. Data & Policy, 3, e6. https://doi.org/10.1017/dap.2021.4

Vertesi, J., & Dourish, P. (2011). The value of data: Considering the context of production in data economies. Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work, CSCW '11, 533–542. https://doi.org/10.1145/1958824.1958906

VijayaVenkataRaman, S., Iniyan, S., & Goic, R. (2012). A review of climate change, mitigation and adaptation. Renewable and Sustainable Energy Reviews, 16(1), 878–897. https://doi.org/10.1016/j.rser.2011.09.009

Vitak, J., Shilton, K., & Ashktorab, Z. (2016). Beyond the Belmont Principles: Ethical Challenges, Practices, and Beliefs in the Online Data Research Community. 939–951. https://doi.org/10.1145/2818048.2820078

Wallis, J. C., Pepe, A., Mayernik, M. S., & Borgman, C. L. (2008). An Exploration of the Life Cycle of eScience Collaboratory Data. https://hdl.handle.net/2142/15122

Wallis, J. C., Rolando, E., & Borgman, C. L. (2013). If we share data, will anyone use them? Data Sharing and reuse in the long tail of science and technology. PLoS ONE, 8. https://doi.org/10.1371/journal.pone.0067332

Wallis, J. C., Wynholds, L. A., Borgman, C. L., Sands, A. E., & Traweek, S. (2012). Data, data use, and inquiry: A new point of view on data curation. https://escholarship.org/uc/item/24x7c6pq

Wang, X., Duan, Q., & Liang, M. (2021). Understanding the process of data reuse: An extensive review. Journal of the Association for Information Science and Technology, 72(9), 1161–1182. https://doi.org/10.1002/asi.24483

Wang, Y. (2024). Do participants lie? Imposter participants in online qualitative research. Qualitative Research Journal. https://doi.org/10.1108/QRJ-06-2024-0130

Warren, C. A. B. (2001). Qualitative Interviewing. In J. Galletta & J. A. Holstein (Eds.), Handbook of Interview Research (pp. 83–102). SAGE Publications, Inc. https://doi.org/10.4135/9781412973588

Warren, K. E. (2016). We Need These Bodies, But Not Their Knowledge: Black Women in the Archival Science Professions and Their Connection to the Archives of Enslaved Black Women in the French Antilles. Library Trends, 64(4), 776–794.

Weber, N. M., Thomer, A. K., Mayernik, M. S., Dattore, B., Ji, Z., & Worley, S. (2013). The Product and System Specificities of Measuring Curation Impact. International Journal of Digital Curation, 8(2), Article 2. https://doi.org/10.2218/ijdc.v8i2.286

Weddig, C. (2022). Climate Change Denial & Skepticism: A Review of the Literature. MediaWell, Social Science Research Council. https://doi.org/10.35650/MW.3043.d.2022

White House Office of Science and Technology Policy (OSTP). (2022). Desirable Characteristics of Data Repositories for Federally Funded Research. Executive Office of the President of the United States. https://doi.org/10.5479/10088/113528

Whitmarsh, L. (2011). Scepticism and uncertainty about climate change: Dimensions, determinants and change over time. Global Environmental Change, Special Issue on The Politics and Policy of Carbon Capture and Storage, 21(2), 690–700. https://doi.org/10.1016/j.gloenvcha.2011.01.016

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, 3, 160018. https://doi.org/10.1038/sdata.2016.18

Wofford, M. F. (2022). Parasitic Knowledge Infrastructures: Data Reuse by Anthropogenic Climate Change Skeptics. Proceedings of the Association for Information Science and Technology, 59(1), 837–839. https://doi.org/10.1002/pra2.743

Wofford, M. F., & Thomer, A. K. (2023). Curating for Contrarian Communities: Data Practices of Anthropogenic Climate Change Skeptics. Proceedings of the Association for Information Science and Technology, 60(1), 442–455. https://doi.org/10.1002/pra2.802

Wofford, M. F., Thomer, A. K., Hemphill, L., Polasek, K., & Yakel, E. (2025). Valuing curation infrastructures. Journal of the Association for Information Science and Technology. https://doi.org/10.1002/asi.70015

Wood for Trees: Home. (n.d.). Retrieved June 7, 2026, from https://woodfortrees.org/

Wu, Y., Bai, R., & Wang, F. (2024). An assessment of whether educated non-researcher audiences understand how to reuse research data. The Electronic Library, 42(6), 949–969. https://doi.org/10.1108/EL-10-2023-0241

Yakel, E., Faniel, I. M., Kriesberg, A., & Yoon, A. (2013). Trust in Digital Repositories. International Journal of Digital Curation, 8(1), Article 1. https://doi.org/10.2218/ijdc.v8i1.251

Yakel, E., Faniel, I. M., & Polasek, K. (2024). Trust. Digital Preservation: A Critical Vocabulary. https://digital-preservation-a-critical-vocabulary.pubpub.org/pub/mfby9zaj/release/1

Yoon, A. (2014). End users' trust in data repositories: Definition and influences on trust development. Archival Science, 14(1), 17–34.

Yoon, A. (2017). Data reusers' trust development. Journal of the Association for Information Science and Technology, 68(4), 946–956. https://doi.org/10.1002/asi.23730

Yoon, A., & Copeland, A. (2020). Toward community-inclusive data ecosystems: Challenges and opportunities of open data for community-based organizations. Journal of the Association for Information Science and Technology, 71(12), 1439–1454. https://doi.org/10.1002/asi.24346

Yoon, A., Copeland, A., & McNally, P. J. (2018). Empowering communities with data: Role of data intermediaries for communities' data utilization. Proceedings of the Association for Information Science and Technology, 55(1), 583–592. https://doi.org/10.1002/pra2.2018.14505501063

Yoon, A., & Kim, Y. (2017). Social scientists' data reuse behaviors: Exploring the roles of attitudinal beliefs, attitudes, norms, and data repositories. Library & Information Science Research, 39(3), 224–233. https://doi.org/10.1016/j.lisr.2017.07.008

Zent, M., Yong, S., Bala, D., Chancellor, S., Konstan, J. A., Terveen, L., & Yarosh, S. (2025). Beyond the Individual: A Community-Engaged Framework for Ethical Online Community Research (arXiv:2503.13752). arXiv. https://doi.org/10.48550/arXiv.2503.13752

Zervas, C. E., & Center for Operational Oceanographic Products and Services (U.S.). (2009). Sea level variations of the United States, 1854-2006 (NOAA Technical Report NOS CO-OPS 53). NOAA. https://repository.library.noaa.gov/view/noaa/2565

Zhao, Y. (2024). TikTok and Researcher Positionality: Considering the Methodological and Ethical Implications of an Experimental Digital Ethnography. International Journal of Qualitative Methods, 23, 16094069231221374. https://doi.org/10.1177/16094069231221374

Zimmerman, A. S. (2003). Data sharing and secondary use of scientific data: Experiences of ecologists [Ph.D., University of Michigan]. https://www.proquest.com/docview/287907131/abstract/AFAB7F64515D4657PQ/1

Zimmerman, A. S. (2008). New Knowledge from Old Data: The Role of Standards in the Sharing and Reuse of Ecological Data. Science, Technology, & Human Values, 33(5), 631–652. https://doi.org/10.1177/0162243907306704

Zuiderwijk, A., & Janssen, M. (2014). Open data policies, their implementation and impact: A framework for comparison. Government Information Quarterly, 31(1), 17–29. https://doi.org/10.1016/j.giq.2013.04.003