Getting Better All the Time: Metadata Enhancements in the ICPSR Catalog

Megan Chenoweth, Senior Associate Librarian, ICPSR
Mike Shallcross, Senior Associate Archivist, ICPSR
Jeannette Jackson, Director of Strategic Initiatives, ICPSR
Jared Lyle, Director of Metadata and Preservation, ICPSR

ICPSR

Today's Presentation

Goal

- Preview changes coming to ICPSR metadata
- Provide practical tips for getting the most out of metadata improvements

Intended Audience

- Data depositors
- Data users
- People who help ICPSR depositors and users



About Metadata

- Definition: “Structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource” (NISO)
- Multiple types of metadata
  - Technical metadata
  - Administrative metadata
  - **Descriptive metadata** for searching and describing information resources
- Selected examples of descriptive metadata at ICPSR:
  - Title
  - Principal investigator(s)
  - Subject terms
  - Summary
  - Funding agency(ies)



Metadata Enhancements: A High-Level Overview

- Improved metadata consistency across studies
- Expanded support for persistent identifiers (PIDs)
- New required metadata fields
- Expanded use of controlled vocabularies
- Enhanced geographic coverage and nationally representative studies



Why These Enhancements Matter

To Depositors

- Easier to describe data deposits
- Consistent with FAIR data principles *(findability, accessibility, interoperability, reusability)*
- Clearer requirements

To Users

- Better search
- More filters
- Improved discoverability



Community-Informed Changes

- Informed by widely used metadata standards and best practices
- Connects us with communities that support metadata standards
  - Data Documentation Initiative
  - DataCite
- Makes ICPSR metadata more efficient and interoperable
  - Supports sharing in library and repository catalogs





Background on ICPSR's Software Modernization

Research Data Ecosystem (RDE) goals:

- Modernize ICPSR's 60 year old software system
- Build tools to support the entire research data lifecycle for social science data
- Improve user experience
- Increase sustainability for ICPSR software platform ensuring viability into the future
- In service of creating FAIR data

RDE software platform will allow you to:

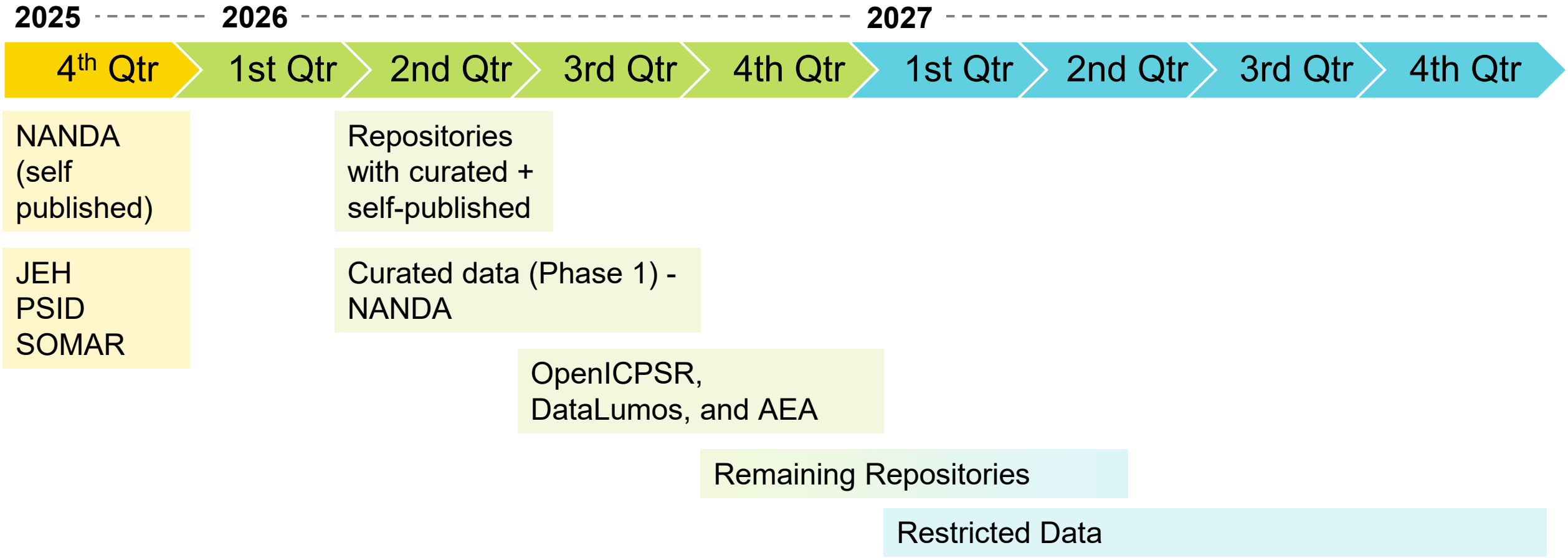
- Deposit and access data
- Explore data to determine suitability for use
- Discover and use different data types including video and geospatial data

Metadata is the underlying foundation for FAIR data at ICPSR, it enables

- Discovery
- Access
- Interoperability

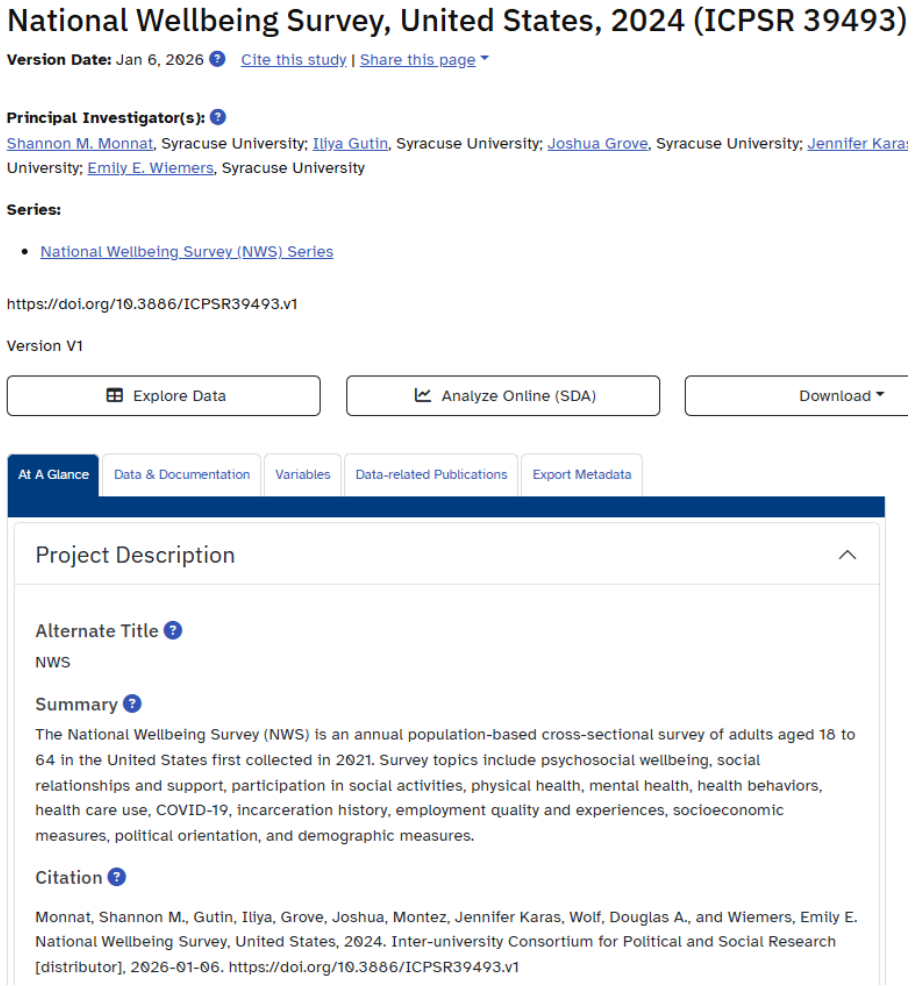
Timeline for Migration to Modernized Platform

Migration order is determined by how technology is being developed. Users will see more consistent metadata and an improved discovery experience



Enhancement #1: Improved Metadata Consistency Across Studies

Current State: Different Metadata for Curated Versus Self-Published Studies



National Wellbeing Survey, United States, 2024 (ICPSR 39493)

Version Date: Jan 6, 2026 Cite this study | Share this page

Principal Investigator(s):
Shannon M. Monnat, Syracuse University; Iliya Gutin, Syracuse University; Joshua Grove, Syracuse University; Jennifer Karas University; Emily E. Wiemers, Syracuse University

Series:

- National Wellbeing Survey (NWS) Series

https://doi.org/10.3886/ICPSR39493.v1

Version V1

Explore Data | Analyze Online (SDA) | Download

At A Glance | Data & Documentation | Variables | Data-related Publications | Export Metadata

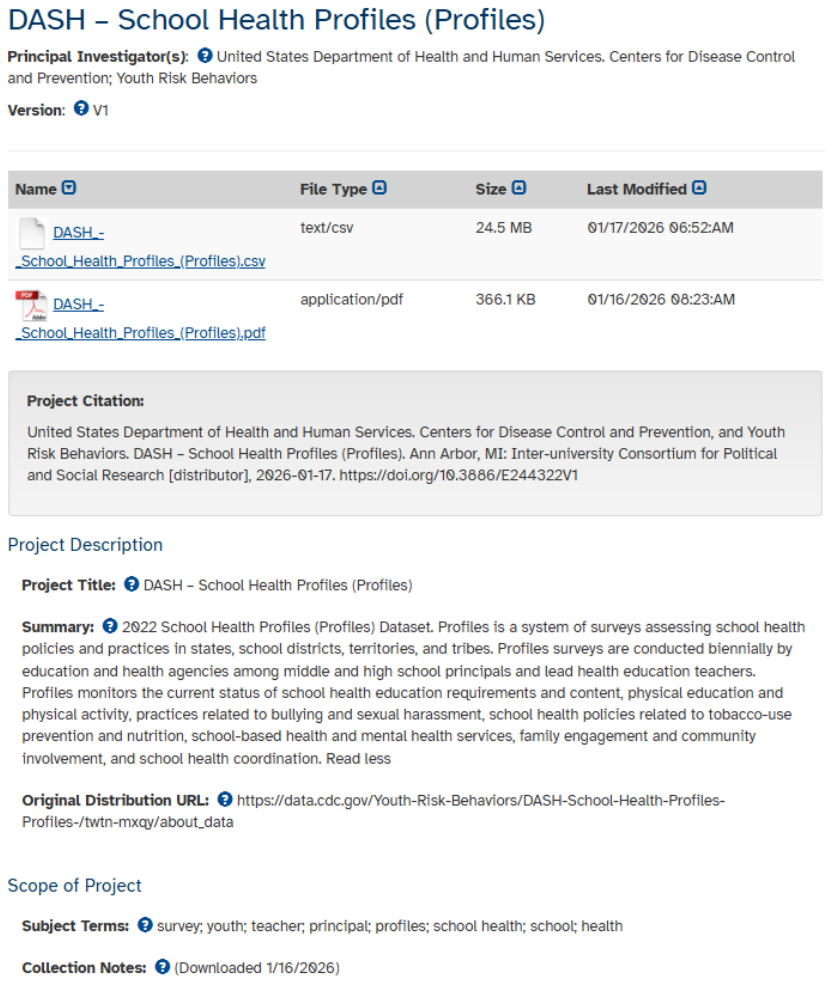
Project Description

Alternate Title
NWS

Summary
The National Wellbeing Survey (NWS) is an annual population-based cross-sectional survey of adults aged 18 to 64 in the United States first collected in 2021. Survey topics include psychosocial wellbeing, social relationships and support, participation in social activities, physical health, mental health, health behaviors, health care use, COVID-19, incarceration history, employment quality and experiences, socioeconomic measures, political orientation, and demographic measures.

Citation
Monnat, Shannon M., Gutin, Iliya, Grove, Joshua, Montez, Jennifer Karas, Wolf, Douglas A., and Wiemers, Emily E. National Wellbeing Survey, United States, 2024. Inter-university Consortium for Political and Social Research [distributor], 2026-01-06. https://doi.org/10.3886/ICPSR39493.v1

A curated ICPSR study



DASH – School Health Profiles (Profiles)

Principal Investigator(s): United States Department of Health and Human Services. Centers for Disease Control and Prevention; Youth Risk Behaviors

Version: V1

Name	File Type	Size	Last Modified
DASH_-_School_Health_Profiles_(Profiles).csv	text/csv	24.5 MB	01/17/2026 06:52:AM
DASH_-_School_Health_Profiles_(Profiles).pdf	application/pdf	366.1 KB	01/16/2026 08:23:AM

Project Citation:
United States Department of Health and Human Services. Centers for Disease Control and Prevention, and Youth Risk Behaviors. DASH – School Health Profiles (Profiles). Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2026-01-17. https://doi.org/10.3886/E244322V1

Project Description

Project Title: DASH – School Health Profiles (Profiles)

Summary: 2022 School Health Profiles (Profiles) Dataset. Profiles is a system of surveys assessing school health policies and practices in states, school districts, territories, and tribes. Profiles surveys are conducted biennially by education and health agencies among middle and high school principals and lead health education teachers. Profiles monitors the current status of school health education requirements and content, physical education and physical activity, practices related to bullying and sexual harassment, school health policies related to tobacco-use prevention and nutrition, school-based health and mental health services, family engagement and community involvement, and school health coordination. Read less

Original Distribution URL: https://data.cdc.gov/Youth-Risk-Behaviors/DASH-School-Health-Profiles-Profiles-/twtn-mxqy/about_data

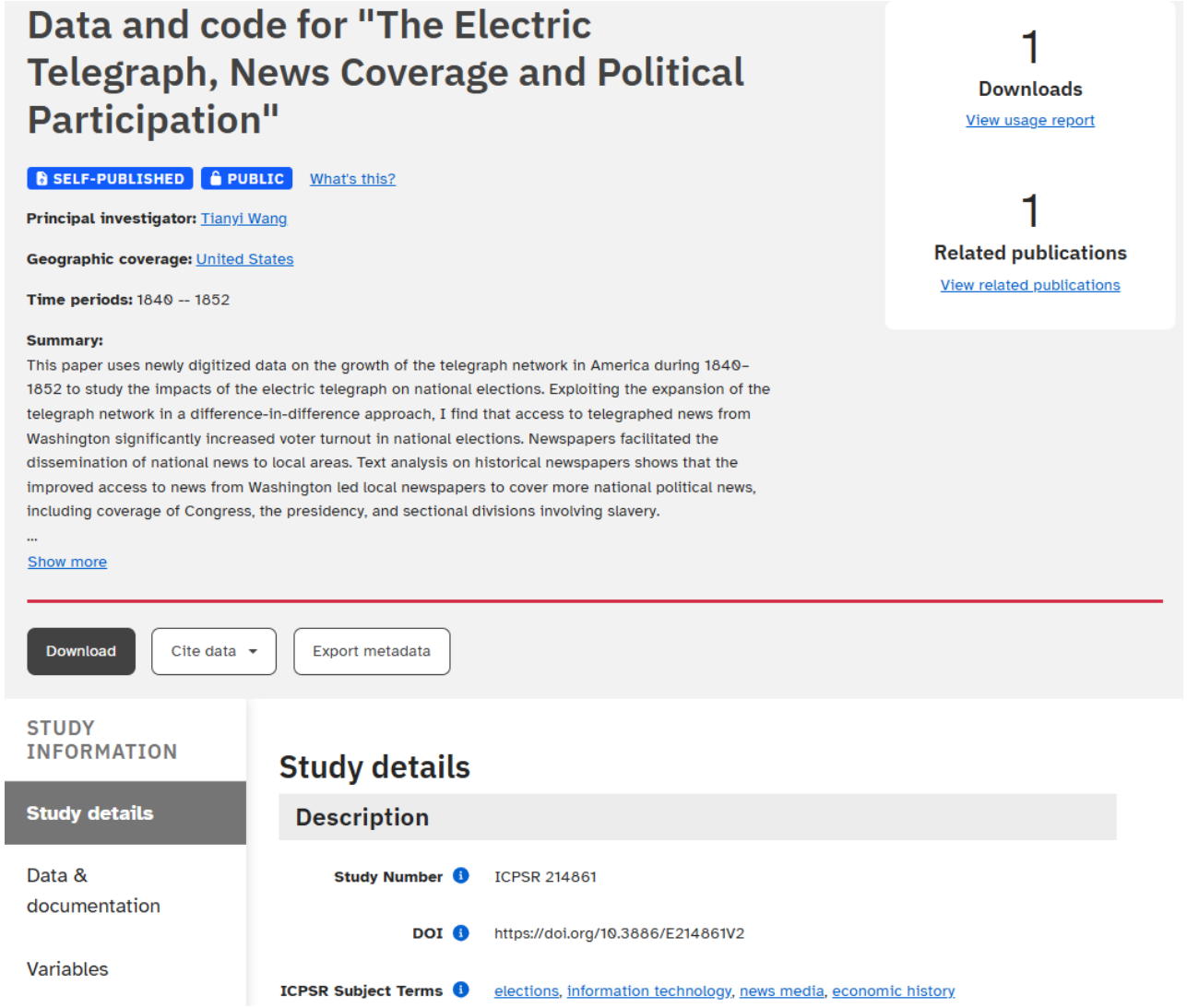
Scope of Project

Subject Terms: survey; youth; teacher; principal; profiles; school health; school; health

Collection Notes: (Downloaded 1/16/2026)

A self-published ICPSR study

Future State: One Set of Metadata Fields for All Studies



A screenshot of an ICPSR study that has been migrated to the new platform and makes use of the new metadata enhancements. All curated and self-published studies will share this layout and have the same metadata fields post-migration.

Enhancement #2: More Support for Persistent Identifiers (PIDs)

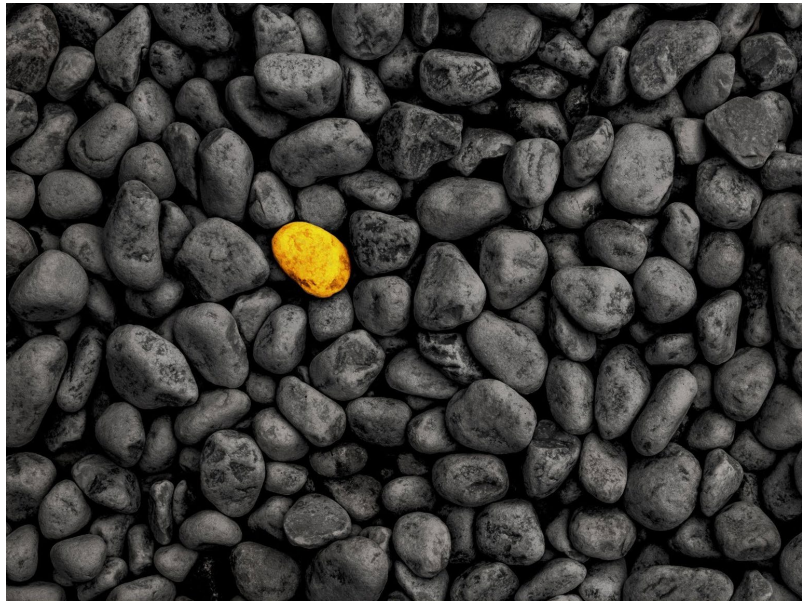
PIDs and Why They Matter

Definition

- Unique, long-lasting references to a digital object
- Example: Digital Object Identifiers (DOIs) for journal articles, datasets, and preprints

Benefits

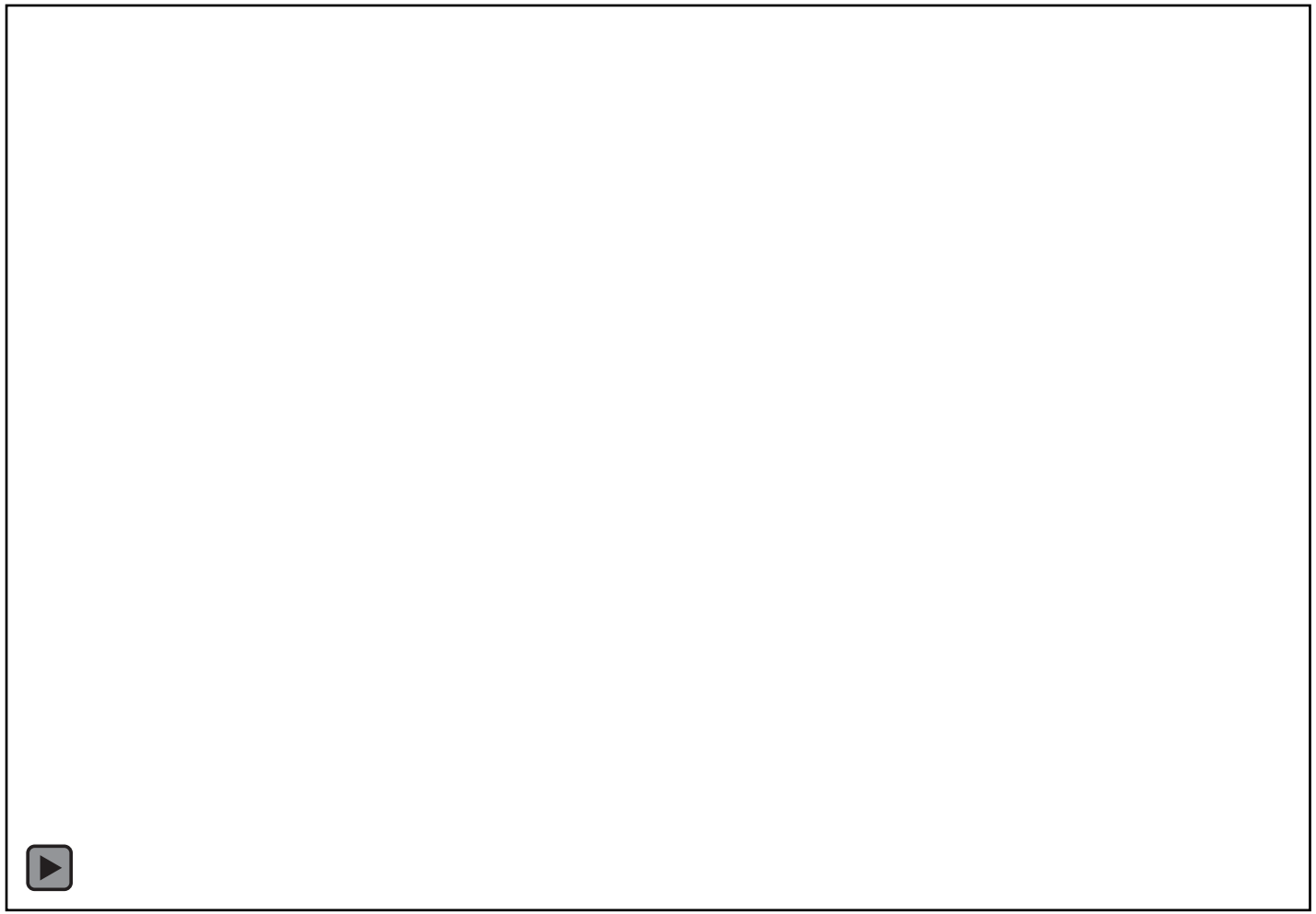
- Uniquely identify people and organizations in research
  - Which John M. Smith?
- Solves problems with variant name forms
  - University of California, Los Angeles vs. University of California-Los Angeles
- Support interoperability across platforms
  - Saves researchers time and limits duplication across platforms



New PIDs at ICPSR

- **Open Researcher and Contributor ID (ORCID)** for people involved in research
  - Individual principal investigators
- **Research Organization Registry (ROR) IDs** for organizations involved in research
  - Investigator affiliations
  - Principal investigators
  - Funding agencies
- Support for grant IDs (less widely adopted)
  - Uniform resource identifiers (URIs)
  - Grant numbers

Adding PIDs to New Deposits



Screen capture of a deposit in ICPSR's new deposit environment. The user scrolls to the Principal Investigators field and clicks on Add PI Information. The user then enters a PI's name in the Search for a Person field. The PI's ORCID, given name, family name, email address, and affiliation are populated based on the content of the ORCID record. The user clicks Save, and ORCID and ROR information are visible in the deposit.

Retrofitting Studies with ORCID and ROR IDs

PIDs are being added to studies as they move to the new platform.

- ROR: wide coverage, easy to search and match
- ORCID: inconsistent coverage, harder to search and match

97%

Investigator Affiliations Matched to ROR

91%

Funding Agencies Matched to ROR

81%

Principal Investigators Matched to ORCID

Laying the Foundation: The Future of PIDs

- **Current** focus:
  - Make it possible to add PIDs to new studies
  - Populate PIDs in existing studies
- **Future** opportunities:
  - Enhance search to make better use of underlying PIDs (e.g., search for people by PID, not name)
  - Enable researchers to auto-populate ICPSR datasets to ORCID



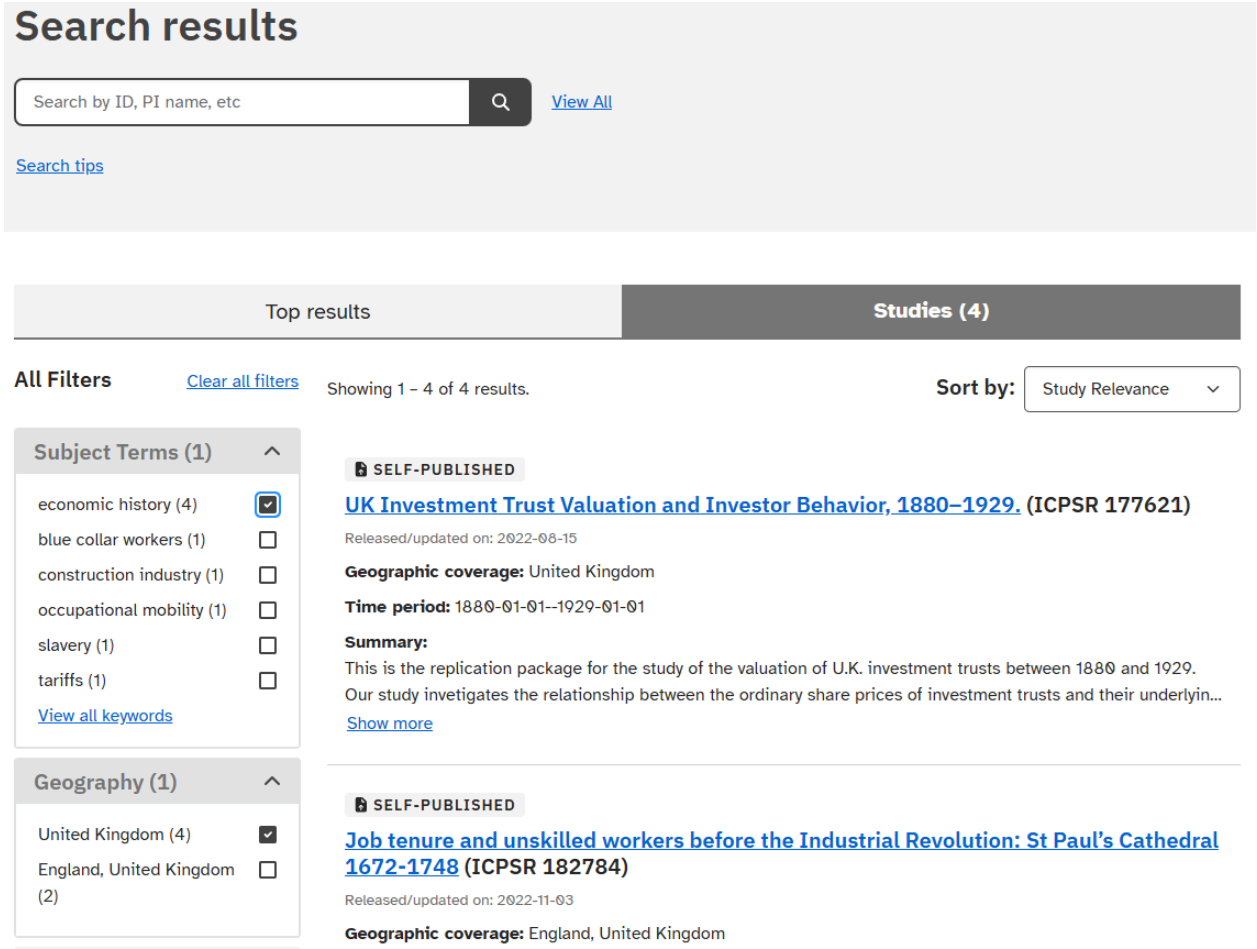
Enhancement #3: Introducing New Required Fields

Required Fields for New Studies

- Previous required fields:
  - **Title** → *what* the data are about (short description)
  - **Summary** → *what* the data are about (longer description)
  - **Principal Investigator(s)** → *who* collected the data
- New additional required fields:
  - **Time Period(s)** → *when* the data pertain to
  - **Geographic Coverage** → *where* the data pertain to
  - **Subject Terms** → *what* the data are about (using controlled vocabulary terms)
- No longer optional in the ICPSR deposit form

Why Require These Fields?

- Make it easier for users to find data
- Surface data based on particular times, places, and topics of interest
- Support filtering in search results



A screenshot of a search in the Journal of Economic History archive, filtered for studies on economic history in the United Kingdom.

Enhancement #4: Expanded Use of Controlled Vocabularies

What's a Controlled Vocabulary?

Definition

- A limited, organized list of words or phrases used to describe information resources

Examples

- Library of Congress Subject Headings
- International Classification of Disease (ICD) codes
- Music or movie genres in streaming platforms

Benefits

- Facilitate searching and browsing across studies
- Bring together studies based on topic or methodology
- Make use of controlled vocabularies widely used in social science research (DDI Controlled Vocabularies)

New Controlled Vocabularies at ICPSR

Subjects

- Medical Subject Headings (MeSH)
- Journal of Economic Literature (JEL) classification codes
- ICPSR subject thesaurus (now required)

Methodology

- Time methods *(longitudinal, time series)*
- Sampling procedures *(probability, complete)*
- Units of analysis *(individuals, households, geographic units)*
- Collection modes *(interview, physical measurements)*
- Smallest geographic unit *(county, census tract)*

Sources and Formats

- Data source types *(medical records, biological samples)*
- General data format *(text, geospatial)*

Populating New Controlled Vocabulary Fields

- Some mapped from existing metadata fields with different values
  - Collection Modes
  - Time Methods
  - General Data Formats
  - Data Source Types
- Some reviewed and retrofitted based on previous values
  - Geographic Unit (free text field) → Smallest Geographic Unit
  - Units of Observation (CV maintained by ICPSR) → Analysis Units
  - Subject Terms (non-standard terms) → ICPSR Subject Terms (authorized terms)
- All available for new studies going forward

Enhancement #5: Geographic Coverage and Nationally Representative Studies

Geographic Coverage: Before

- Based on ICPSR-specific controlled vocabulary
  - Labor-intensive to maintain
  - Not comprehensive
- Place names stored as strings
  - No connection between different levels of geography
  - Limits ability to search and filter



Screenshot of a study focusing on the Geographic Coverage field. The field contains the terms "Detroit," "United States," and "Michigan" with no clear relationship to one another.

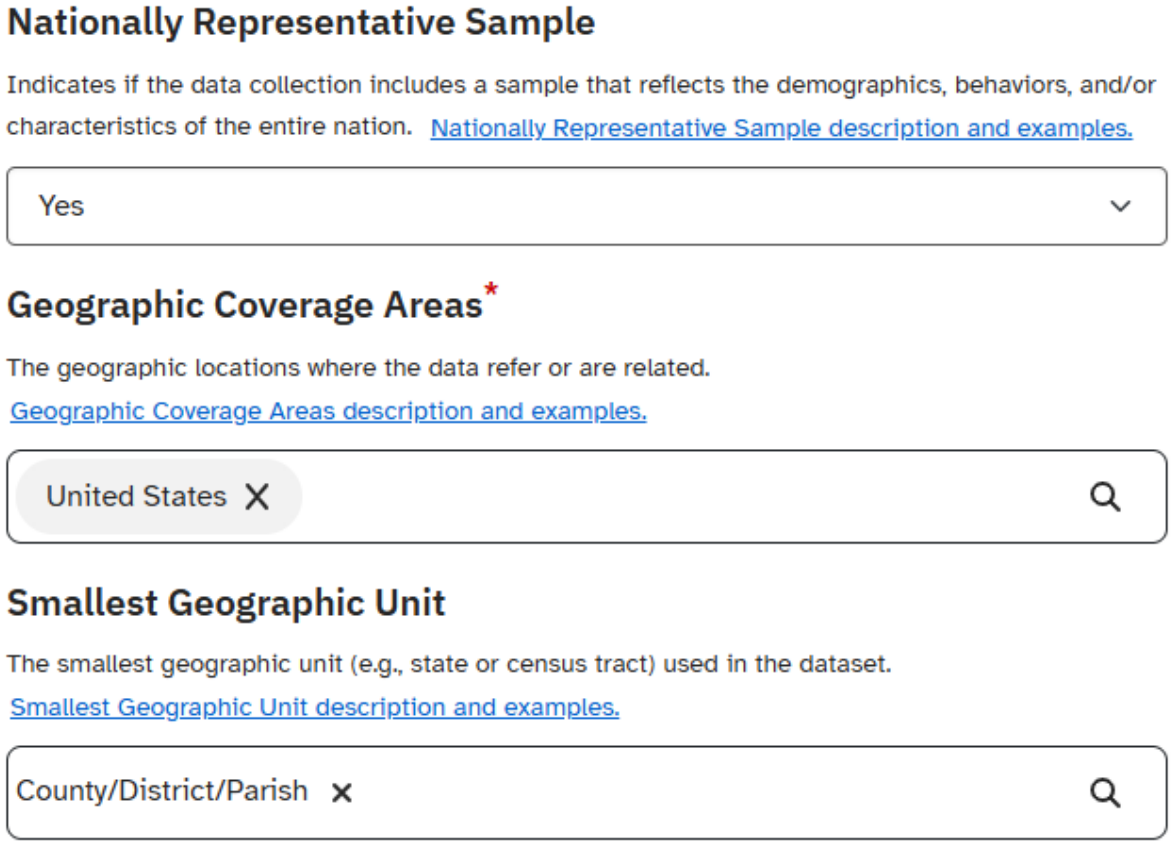
Geographic Coverage: After

- Standardized via GeoNames.org
- Stored with full geographic hierarchy
- Each place tagged by type (city, state, country)
- Clearer, more accurate geographic precision
- Much more comprehensive and don’t need to maintain locally
- Lays the foundation to do more with geography in the future



Nationally Representative Studies

- Sample designed to represent the demographics and characteristics of an entire nation
- Increases confidence that findings are valid for the whole population
- Information currently found only in free text metadata fields or documentation
- New yes/no field to clearly designate studies as nationally representative
- Users looking for nationally representative data will be able to filter for it directly, without guessing from documentation
- Enables future searching and filtering functionality



Takeaways: Making the Most of ICPSR's Enhanced Metadata

Grow Your ORCID

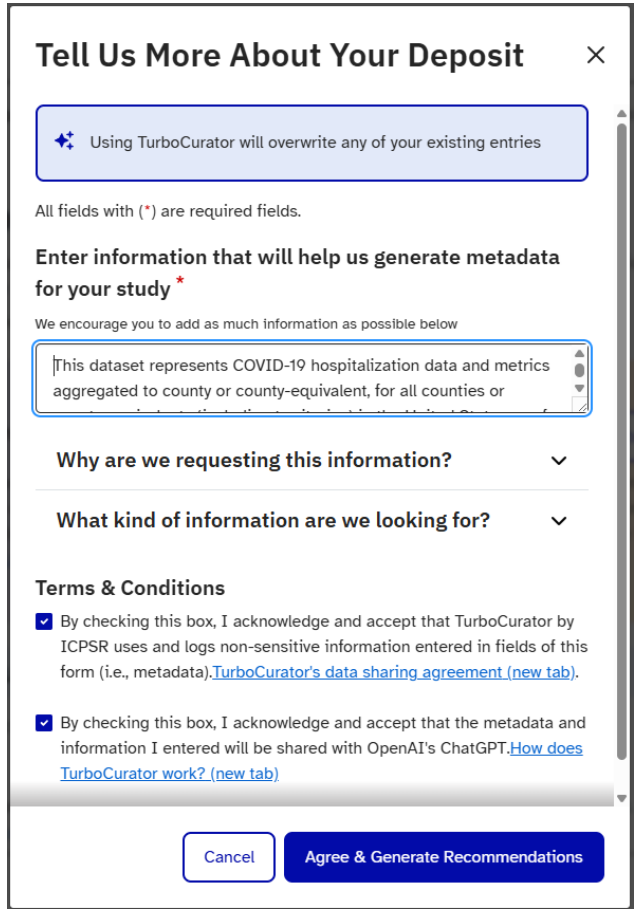
- ☑ Create an ORCID at orcid.org
  - Free
  - Open to any researcher
  - Customizable
  - As of 25 January, required in applications and reports to the National Institutes of Health (NIH)
- ☑ Optimize your ORCID
  - Add or update affiliations (including ROR)
  - Allow update by trusted organizations like Datacite
  - Advice: info.orcid.org/researchers
- ☑ Enter ORCID in new ICPSR deposits

Applies to:

- Depositors
- Co-investigators
- Users you support

Be Prepared for New Required Fields

- Reminder: geographic coverage, time period, and ICPSR subject terms are required, not optional, in the new platform
- Be ready to provide these details for new deposits
- Tip: TurboCurator can suggest ICPSR subject terms



A screenshot of ICPSR's TurboCurator tool with a study description provided.

Explore ICPSR's Metadata Enhancements

- Browse repositories that use the new platform
  - Journal of Economic History Data Repository: icpsr.umich.edu/sites/jeh
  - Panel Study of Income Dynamics Repository: icpsr.umich.edu/sites/psid
  - National Neighborhood Data Archive (self-published studies only): icpsr.umich.edu/sites/nanda
  - More coming soon
- For an in-depth look at new metadata fields, check out these studies:
  - JEH: Reform, Rails, and Rice: Political Railroads and Local Development in Thailand
  - PSID: PSID data extracts for: The effect of child support on fathers' labor supply
  - NaNDA: Grocery and Food Stores by Census Tract and ZCTA

Questions? Need Advice? Contact ICPSR



WEBSITE
www.icpsr.umich.edu



EMAIL
icpsr-help@umich.edu



PHONE
734-647-2200