

# Gaps in actionable scholarly publishing meta(data)

Megan O'Donnell  
Head of Research Data Services  
Iowa State University, University Library

# Disclaimer: I'm not a metadata librarian



Head of Research Data Services  
part time subject librarian

- I manage an institutional data repository.
- I curate and publish data sets.
- I also advise, consult, and teach.
- Publication metadata is data.

# Libraries are:

- Mechanisms of discovery and knowledge making.
- Heavy consumers AND producers of (meta)data.
- Information and data management experts.
- Maintainers of official records.
- And publishers.

But not most scholarly publishing records.

# Typical metadata

## Institutional Repositories

- Original works (ETD, tech reports, etc.)
  - Without assigned DOIs they're not *part of the scholarly linked data environment.*



## Green OA

- Can link back to the official publication but this generates *multiple sources of valuable metrics.*

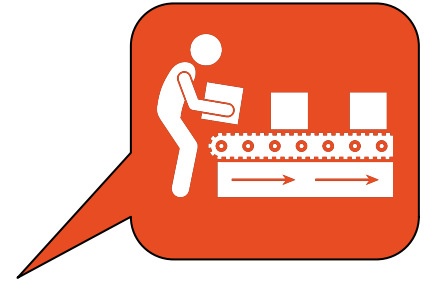


Icons by [Gan Khoo Lay](#) from Noun Project

# Typical metadata

## People IDs (ORCID):

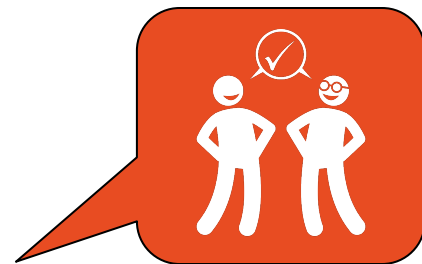
- Typically, opt-in but with uneven adoption.
- Still not a part of university systems.
  - Not well integrated with **HR systems**.
  - Not “baked in” to most **library systems**.
- Highly reliant on authors to create, maintain, and **link**.
  - *Active and retroactive record updates from authors and libraries are possible.*



# Emerging metadata

## Resources and Centers (HPC, imaging, etc.):

- Track use and justify funding for expensive resources and expertise.
  - Who did we help? What research did we enable?
- *Standards are needed.*
  - Attribution as authors? By citation?
  - *Libraries can manage and update records for local resources.*





# Emerging metadata

## Billing Verification for R&P Agreements

- Wrong institutions listed
  - College / Lab / Center  $\neq$  University
- Email address errors.
  - Library is never contacted.
  - Email addresses belong to child-orgs.
- Query ROR for related organizations?
  - Not granular enough?
  - Doesn't track web domains.

 <https://ror.org/041m9xr71>

### Ames Laboratory

#### FACILITY

#### LOCATION

Ames (GeoNames ID 4846834)

United States

#### WEBSITE

<https://www.ameslab.gov/>

#### RELATIONSHIPS

Parent Organization(s)

United States Department of Energy

Related Organization(s)

Iowa State University

# Problematic metadata

## Funding and awards

- *Standards are needed.*
  - Award ID numbers
  - Discovery and verification
  - Attribution guidelines and metadata.
- Support for different award types:
  - Cooperative agreements, contracts, etc.



*Ames Laboratory operates under a DOE contract and has project numbers.*



*CSAFE is funded through a cooperative agreement with NIST and four institutions. We have been given three different numbers over four-years.*



# Incomplete metadata

## Data, data sets, datasets, and databases

- Lack of metadata coordination between publishers and repositories.
  - ORCIDs\*, funding,\* article publication date,\*...
  - Publications are not updated with links to data
  - data *is* updated with links to publications.
- *You cannot query which publications:*
  - Have data or claim to have data.\*



\* Relates to August 25, 2022 [Ensuring Free, Immediate, and Equitable Access to Federally Funded Research](#) OSTP memo and [NSM-33](#).

Icons by [Gan Khoon Lay](#) from Noun Project

# Poor metadata & discovery chaos

TABLE 1. Crude and age-adjusted prevalence of self-reported physician-diagnosed diabetes among Asians and Native Hawaiians or other Pacific Islanders (NHPis) aged ≥18 years, in 32 states, the District of Columbia, and Guam\* — Behavioral Risk Factor Surveillance System, United States, 2011–2014

State/Territory	Race	No. of respondents	Crude prevalence % (95% CI)	Age-adjusted prevalence* % (95% CI)
Alaska	Asian	313	7.1 (4.5–11.1)	9.0 (5.6–14.3)
Arizona	Asian	108	3.4 (2.0–5.8)	4.9 (2.9–8.0)
California	NHPI	224	14.9 (9.7–22.3)	19.1 <sup>§</sup> (13.5–26.4)
	Asian	3,102	8.9 (7.7–10.3)	9.7 (8.4–11.1)
Colorado	Asian	589	5.2 (3.5–7.8)	7.1 (4.9–10.1)
Connecticut	Asian	647	4.8 (3.3–7.0)	8.0 (5.6–11.3)
District of Columbia	Asian	351	2.8 (1.6–4.9)	5.1 (3.0–8.5)
Florida	Asian	661	7.6 (5.5–10.5)	10.2 (7.5–13.7)
Georgia	Asian	371	4.7 (2.8–7.6)	9.5 (6.0–14.5)
Hawaii	NHPI	930	10.7 (8.4–13.6)	13.9 <sup>§</sup> (11.1–17.3)
	Asian	9,503	10.9 (10.1–11.8)	8.8 (8.0–9.6)
Illinois	Asian	520	5.7 (3.8–8.4)	9.0 (6.2–12.9)
Indiana	Asian	318	6.3 (3.9–10.2)	10.5 (6.5–16.6)
Iowa	Asian	230	5.2 (3.2–8.5)	11.4 (6.9–18.2)
Kansas	Asian	685	4.2 (3.0–6.0)	8.3 (6.0–11.5)
Maryland	Asian	965	7.1 (5.4–9.2)	9.1 (7.1–11.6)
Massachusetts	Asian	628	5.5 (4.1–7.3)	10.7 (7.9–14.3)
Michigan	Asian	470	4.9 (3.4–7.1)	9.1 (6.3–13.0)
Minnesota	Asian	892	3.8 (2.6–5.7)	6.2 (3.7–10.1)
Nebraska	Asian	455	4.3 (2.6–6.9)	6.9 (4.4–10.8)
Nevada	Asian	485	12.2 (8.8–16.8)	13.4 (10.0–17.7)
New Jersey	Asian	231	7.8 (6.4–9.5)	9.9 (8.1–11.9)
New Mexico	Asian	250	7.9 (5.0–12.3)	9.9 (6.5–14.7)

*Ceci n'est pas un ensemble de données.*

An image of a data table accompanied by the French phrase: "this is not a data set."  
(this is play off Rene Magritte's "The Treachery of Images," 1929).

- **Item type** is underdeveloped for non-traditional items.

Data set =

- Tabular data (yes)
  - File sets (often)
  - Tables (no)
  - Graphs (no)
  - PDFs (no)
- This is not useful metadata.



# Summary of gaps and next steps:

- More standards and standardization.
- Metadata verification (who and how).
- Post-publication updates and enrichment.
- Communication between data and text publishers.
- Goals and ethics for scholarly publication *metadata* (it is data after all...).