

DATA CURATION NETWORK

Beyond compliance: Curation as essential open science infrastructure

Wind Cowles, PhD

Associate Dean for Data,
Research, and Teaching
Princeton University

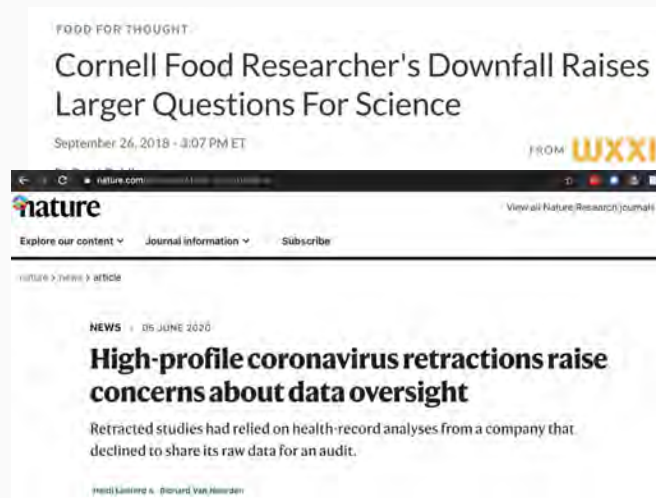
Mikala Narlock

Director, Data Curation Network
University of Minnesota

Presented October 19, 2023: “Developing New Approaches to Promote Equitable and Inclusive Implementation of Open Scholarship Policies.” Hosted by The National Academies of Sciences, Engineering, and Medicine’s Roundtable on Aligning Incentives for Open Scholarship

Open Data vs. Misinformation

- 1221 research articles with authors affiliated with institutions in the United States were retracted between 2013 - 2023 for reasons that included data (Retraction Watch DB)
- Sharing data openly is only part of solution
 - *If shared data appears unorganized or poorly documented, that doesn't lead to increased confidence against misinformation*
- Lessons from COVID



Compliance

- “I have to share my data, code, etc.”



Open
Equitable
Accessible

What is data curation?

“Data curation is the active and ongoing management of data through its lifecycle and interest and usefulness to scholarship, science, and education; curation activities enable data discovery and retrieval, maintain quality, add value, and provide for re-use over time.”

- University of Illinois' Graduate School of Library and Information Science

What is data curation?

“Data curation is akin to work performed by an art or museum curator. Through the curation process, data are organized, described, cleaned, enhanced, and preserved for public use, much like the work done on paintings or rare books to make the works accessible to the public now and in the future.”

- ICPSR

Levels of Curation

Distributed as deposited

Level 0

Record Level Curation

perform brief metadata checks for increased **findability (F)**



Level 1

File Level Curation

review files arrangement and perform file format conversions for increased **accessibility (A)**



Level 2

Documentation Level Curation

review documentation and request/add missing information for increased **reusability (R)**



Level 3

Data Level Curation

open files and review data contents and may annotate /edit the data for accuracy or **interoperability (I)**



Level 4



Local *and* Distributed Work

- Institutional support structures for research data and computation
- Key institutional ingredients: Staff, infrastructure, and policy
- Training and advice for individual researchers is crucial
- Professional data curation, in a community of practice that shares expertise in specific kinds of data, allows for long-term value through the ability to reuse the data, and continuity of services through professionalization

Incentivizing Open Data

- Open data access and sharing are difficult to incentivize
 - Time
 - Trust
 - Perceived value of the work
- External pressures can only go so far
 - Funders & Journals
 - Research Communities

Incentivizing Open Data

- Support an ecosystem of institutional, domain-general, and specialist repositories
 - Data curation and metadata standards create better, more reusable data
- Place value on data as a co-equal output to traditional publications
 - Citation as reward: e.g., Make Data Count

Fund curation (not just storage)

Data are published (not just shared)

Contact us!

Wind Cowles

windcowles@princeton.edu

Mikala Narlock

mnarlock@umn.edu