

Commonplace

Data as Community

Dylan Ruediger

Published on: Aug 29, 2022

URL: <https://commonplace.knowledgefutures.org/pub/c621zmfi>

License: [Creative Commons Attribution 4.0 International License \(CC-BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)

Over the past several decades, federal agencies such as the NSF and NIH have made substantial investments in building an infrastructure to support open sharing of research data among researchers. These investments have funded the creation of a host of platforms and tools, among the most important of which are the decentralized network of domain and generalist repositories available to researchers. The technical infrastructure necessary for widespread deposit and reuse of data is now relatively mature.

Researchers' behaviors, however, have lagged behind. In many fields data sharing is still more of an ideal than a reality. Indeed, recent results from a national [survey](#) of faculty conducted by Ithaka S+R indicate that while two out of three faculty in scientific and medical fields believe that depositing research data in public repositories is important, just 40 percent of scientists and half of faculty in medical fields report that they often or occasionally share research data.

The gap between researcher's sentiments and practices is often attributed to academic incentive structures that do not reward researchers for the considerable labor involved in organizing and preparing data for FAIR compliant sharing. To the extent this is true — and it is clearly an important factor — there are two obvious ways to increase data sharing practices: developing incentive structures that reward data sharing and implementing robust mechanisms to enforce compliance.

Since 2019, Ithaka S+R has also been tracking a third incentive through [case studies](#) of what we have called “data communities.” **Data communities are fluid networks of scientists who voluntarily exchange and reuse data across disciplinary boundaries to advance shared or complementary research goals.** Examples of data communities include: FlyBase, the Cambridge Structural Database, and ZooArchNet. Typically organized around a domain repository, data communities promote data sharing among researchers with mutual interests in research questions to topics. Understanding the dynamics behind successful data communities can shed light on how to encourage the creation of norms and practices that focus on a common good as much as on rewarding individuals.

Earlier this year (2022), Ithaka S+R hosted a multi-day incubation workshop, funded by the National Science Foundation and organized in partnership with the Data Curation network designed to explore the concept of data communities in more detail. The workshop included fourteen research teams from a wide range of disciplines who were either establishing a domain repository or seeking to expand the user-base and functionality of an existing one. A core goal of the workshop was to explore whether the concept of a “data community,” originally conceived as a descriptive term for a specific type of data sharing initiative, could be used as a philosophy of action to assist researchers in their efforts to build or sustain domain repositories.

Our [findings](#) from the workshop suggest that **there is power in thinking about repositories as a gathering place for communities, not just as an archive of datasets.** Framing conversations around the idea of “data communities” encouraged participants to center the social, cultural, and interpersonal relationships that support

data sharing and consider challenges relating to growth through the lens of human relationships and community building.

The value of this shift in focus is most obvious in relation to one of the core challenges many domain repositories face: enticing new users to submit and reuse data. Many of the research teams that participated in our workshop were grappling with the question of how to expand beyond an initial group of core users, often by improving functionality of their repositories to attract new depositors, or enhancing metadata to improve discoverability to encourage reuse. One lively topic of discussion on how to leverage one the unique advantages of domain repositories, namely their capacity to bring together researchers with overlapping research agendas. Generalist repositories, by definition, will have a larger pool of potential users, but domain repositories are well positioned to identify and cultivate a specific audience, particularly if they conceive of their work as community building. Researchers identify deeply with their disciplines and with the research problems they dedicate their careers to solving: they are most readily motivated to share researchers with whom they share interests, and with whose research they identify. **Focusing on the communal nature of the repository, and articulating its value to members of a research community can be an important part of an outreach strategy.**

Less obviously, foregrounding the social infrastructure for data sharing can also inform approaches to metadata and ontologies, two issues that present formidable challenges to both new domain repositories and established repositories looking to expand. Metadata and ontologies are often treated primarily as descriptive tools that facilitate data discoverability. Throughout the workshop, it became clear that metadata and ontologies also embody community norms and shared — if contestable — frameworks of understanding. **By emphasizing metadata as a social process rather than a descriptive output, as a call for engagement rather than a technical standard, several research teams in our workshop were able to conceptualize new ways to use the process of articulating metadata standards as a form of community building, a way of engaging researchers in conversations that fosters their investment in sustaining a repository.**

At least one data community in our cohort had experimented with ways to use community building as a framework for creating **metadata for hydrological samples** that would be useful to members of different disciplines. This process involved hosting disciplinary workshops that collaborated to develop metadata and then worked to harmonize their results. As a member of that data community noted, creating discipline-specific standards resulted in bottom-up and consensual metadata that emerged organically from conversations with **existing and potential users, who now had a stake in the results that they would not have had if they were simply asked to comply with a pre-existing standard.** This example led to fruitful brainstorming about how to engage users, through conversation or even through gamification, badging, or metadata building events structured along the lines of hackathons.

The idea that social change is a necessary part of promoting open science is well known, but the social infrastructure for data sharing has often taken a backseat to efforts to build the technical infrastructure

necessary to support it. Academic incentive structures and the culture of researchers in many disciplines continue to hinder widespread sharing of research data. Understanding why researchers choose to voluntarily share data will not solve these problems. However, our work suggests that they can contribute to cultural change by creating community-generated incentives for data sharing rooted in shared interest.

Bio: Dylan Ruediger is a senior analyst at Ithaka S+R.