

Metadata: The accelerant we need

Kathryn A. Kaiser^a, John Chodacki^b, Ted Habermann^c, Jennifer Kemp^d, Laura Paglione^{e,*}, Michelle Urberg^f and T. Scott Plutchak^g

^a*Assistant Professor, Department of Health Behavior, School of Public Health, 227J Ryals Public Health Building, University of Alabama at Birmingham, Birmingham, AL, USA*

^b*California Digital Library, Oakland, CA, USA*

^c*Metadata Game Changers, Boulder, CO, USA*

^d*Crossref, Lynnfield, MA, USA*

^e*Spherical Cow Group on behalf of Metadata 2020, Oxford, UK*

^f*ProQuest, Seattle, WA, USA*

^g*University of Alabama (retired), Birmingham, AL, USA*

Abstract. Large-scale pandemic events have sent scientific communities scrambling to gather and analyze data to provide governments and policy makers with information to inform decisions and policies needed when imperfect information is all that may be available. Historical records from the 1918 influenza pandemic reflect how little improvement has been made in how government and policy responses are formed when large scale threats occur, such as the COVID-19 pandemic. This commentary discusses three examples of how metadata improvements are being, or may be made, to facilitate gathering and assessment of data to better understand complex and dynamic situations. In particular, metadata strategies can be applied in advance, on the fly or even after events to integrate and enrich perspectives that aid in creating balanced actions to minimize impacts with lowered risk of unintended consequences. Metadata can enhance scope, speed and clarity with which scholarly communities can curate their outputs for optimal discovery and reuse. Conclusions are framed within the Metadata 2020 working group activities that lay a foundation for advancement of scholarly communications to better serve all communities.

Keywords: Metadata, public health emergencies, environmental disasters, academic silos, public policy, COVID-19

1. Introduction

In the wake of the novel coronavirus (SARS-CoV-2) infecting people all over the world, researchers have turned their attention to the myriad issues that must be addressed in order to understand and rein-in the pandemic. Journal editors, publishers, and librarians have responded by working to make the results of this research as widely available as possible. The Wellcome Trust, for example, organized a statement signed by numerous funders, research institutions, and publishers, urging that all parties “ensure that research findings and data relevant to this outbreak are shared rapidly and openly to inform the public health response and help save lives” [1]. On April 15, the NIH announced the release of the COVID-19 Portfolio Tool [2], one of many resources that have been developed to assist data- and text-mining of the COVID-19 Research Dataset (CORD-19), a corpus of over forty-five thousand research articles relevant

* Corresponding author: Laura Paglione. Tel.: +1 301 500 2139; E-mail: lpaglione@metadata2020.org.

to the pandemic [3]. This outpouring of effort and energy is gratifying, but significant barriers to making effective use of this mass of material remain.

Critical questions need to be answered in order to establish effective policies and guidelines: Are there drugs that are effective in treating COVID-19? Do the antibodies created in an infected person provide effective immunity and for how long? What is the rate of asymptomatic infections? And many more. The data and information that might answer these questions may be buried in the COVID-19 collection, but the efforts to efficiently make use of it are hampered by the lack of sufficiently rich, connected, reusable, and open metadata that describes the structure and content of those journal articles, preprints, and datasets.

Metadata provide a crucial piece of the infrastructure that binds the scholarly enterprise together. They act as the accelerant that enables people, and increasingly machines, to connect the data and information contained in discrete articles and datasets and generate new, actionable knowledge. While the lack of sufficiently robust metadata slows discovery by scholars and researchers during the best of times, it is in times of crisis that the need for improving metadata becomes most apparent. Ten years ago, understanding the impact of the Deepwater Horizon disaster in the Gulf of Mexico required huge amounts of marine, atmospheric, and biologic data. Sorting through the data to effectively pinpoint environmental effects and potential health effects on the populations along the coast was slowed by the lack of appropriate metadata.

In times of crisis, it is not enough to more effectively analyze papers and datasets in order to identify the scientific facts that can be revealed; those facts have to be put to use in developing policy and guidelines for action. Policy makers need access to accurate and up-to-date information across various disciplines in the humanities and social sciences in order to balance the public health, economic, and social upheavals that crises such as these can bring.

Since 2017, the Metadata 2020 initiative has been elevating the discourse to highlight the universal challenge our communities face in improving metadata quality. A cross-community collaborative effort encompassing the work of a wide variety of stakeholders, the initiative recently published recommendations for Principles and Best Practices [4]. Metadata need to be “Compatible, Complete, Credible, and Curated” Crises like Deepwater Horizon and COVID-19 bring the need for efficient access to trustworthy, usable knowledge and data into sharp focus.

In 2011, the writer Nicholas Carr wrote about the difference between “situational overload” and “ambient overload”. The former is the traditional needle in the haystack problem - trying to find the right piece of information buried in all the noise. But now we are all facing ambient overload - so much information of interest that we feel overwhelmed trying to make sense of it. “Ambient overload doesn’t involve needles in haystacks. It involves haystack-sized piles of needles” [5]. Better metadata can help.

2. The role of metadata in retrieving and analyzing the scientific literature

The COVID-19 pandemic shines a light on the inefficiency in how scientists and publishers of scientific literature describe and curate research articles. Under normal conditions, when a group of scientists wants to synthesize and evaluate the quality and completeness of a collection of studies on a topic or issue, they perform what is known as a systematic review. Using the standard metadata for describing articles (descriptive titles, abstracts, subject headings, keywords), they attempt to identify all of the relevant literature. This methodology has been applied most widely in (and came out of) the biomedical community since the 1990s, but now other disciplines (e.g., engineering, business, etc.) are using the same techniques to consolidate and evaluate a body of research. There are guidelines for rigorously conducting and reporting these systematic reviews, which may also include a meta-analysis, which applies statistical

techniques to merge numerical outcomes (e.g., average change in body weight after a weight loss treatment) into an overall estimate. This approach aids in predicting outcomes in similar populations and averages out minor variations that may be due to sample size or treatment dose. The results inform treatment guidelines and policies in the biomedical arena, but can inform anybody wishing to take research results in aggregate to determine what research is still needed, or if perhaps no further research is needed to address a particular question.

Applying systematic review techniques to the articles in the CORD-19 dataset could help to ensure that clinicians, public health workers, and policy makers have the best information upon which to make critical clinical, research, and policy decisions. However, it takes too long to complete these reviews without extensive resources. One study reported an average of over five researchers and more than sixty-six weeks from the time a systematic review was planned and registered to when it was published [6]. The reason for this delay is rooted in how articles are indexed in various databases. Standard metadata provide insufficient information about the article content, and little context to aid researchers in creating search syntax that is comprehensive without yielding considerable waste. Thus, the results of a search can generate thousands of potentially useful articles that must be sifted through, often with few or none actually being applicable to a highly-structured research question. Data from several studies shows this “search precision” (the number of actually relevant articles in the set of articles retrieved) to average less than three percent.

Attempts to apply very sophisticated semantic search tools and artificial intelligence to address this problem of low search precision have had little success. Search precision can be raised to double digit levels, but concerns remain about the high degree of false positives (irrelevant articles) and false negatives (relevant articles that should have been retrieved but were not). When there are relatively few articles on a subject, and they may be published in any one of thousands of journals, even very thorough searches may miss important studies. There are presently hundreds of millions of research objects one may wish to search through, scattered across numerous databases and repositories, but no comprehensive curation and metadata strategy has been developed to speed their findability [7].

No expert can keep up with the scope or speed of outputs in the global research environment in many domains. We need to index the scientific record in a completely searchable relational database, with components that include persistent identifiers and links to supporting data repositories. For biomedicine, structured metadata could be a machine- and human-readable object that provides context in the PICOS (population, intervention, comparison, outcome(s), and study design) framework. This five-point framework is currently used to manually select the studies for systematic reviews, but presently too much of the needed information is unsearchable or unavailable in content platforms or abstracting and indexing services. Further, researchers in low resource settings may not be able to access and use the information in articles behind paywalls.

Comparing different strategies for identifying scientific literature to study a question illustrates some of the problems. In a scoping review article [8] to summarize “...clinical, epidemiological, laboratory, and chest imaging data related to the SARS-CoV-2 infection”, the authors selected sixty articles that met the criteria, from a starting pool of two thousand four hundred and sixty-five articles retrieved from five major databases. A comparison of the overlap between these articles that were included and the CORD-19 dataset for the same time period (articles published in 2019 and 2020 up to a search date of January 1, 2019 to February 24, 2020) shows three articles in common. The approach used to generate CORD-19 and the traditional systematic review method yield vastly different lists of articles and data sources. While existing systems have their place, metadata enrichment is needed to speed targeted assessment/inventory of a body of literature. Making lots of information available does not make it immediately more usable if the volume is large and difficult to navigate. A new paradigm - where articles are treated as datasets

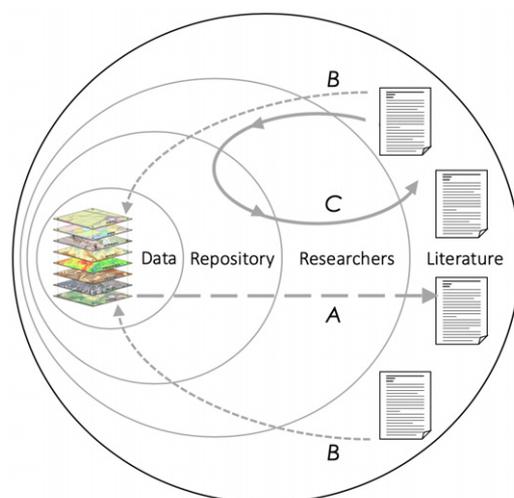


Fig. 1. Data flows in Deepwater Horizon and Covid-19.

or as part of a larger dataset still forming - is needed to complement the old one in order to make and keep the scientific record useful and alive in the future. If study authors wrote and presented their work with rich metadata that made it accessible and reusable, we can make scientific knowledge a relational and searchable database to fully preserve the investments made.

3. The role of metadata in retrieving and analyzing datasets

The relationship between metadata and research data is similar to the well-known relationships between chickens and eggs. It is never really clear which comes first. The nature of the relationship varies with the type of data. The Deepwater Horizon (DH) oil spill in the Gulf of Mexico starting in April 2010, with nearly five million barrels of oil spilled, is considered one of the worst-ever environmental disasters. In this case, many background environmental datasets were available before the spill, but were not easily discoverable or accessible because of non-standard or incomplete metadata. Efforts were focused on improving data discovery and integration into analytical tools. In this case the “where” and “when” of the data were critical for discovery and integration, so Geographic Information Systems (GIS) and internet mapping played an important role in data discovery.

Many of the datasets needed to fully understand the impact of DH and to make policy and regulatory recommendations for the future had been collected and curated by federal, state, and local government agencies. These provided a foundation for researchers across many sectors. In these cases, the data came first, and repositories play a critical role in data curation and metadata creation to support discovery and access. This data flow is illustrated by arrow A in Fig. 1. Data are collected, go into a repository for curation where they are discovered, accessed, used, and analyzed by researchers who publish results in papers. In these cases, minimal metadata requirements support discovery and links to more detailed metadata are provided to help researchers use and understand the data. When researchers use these data in papers, they are expected to cite the original data sources using persistent identifiers (curves B in Fig. 1). The papers thus build a layer of detailed documentation and interpretation around the data, facilitating deeper

understanding, trust, and reuse. When the metadata is insufficient or inconsistent, this building process risks being incomplete and potentially faulty.

In the COVID-19 case the papers and reports *are the data* and the metadata are generally those associated with the publication of the paper rather than metadata that describes the data reported in the papers. In Fig. 1, this path is labeled C. The papers go into repositories, generally associated with publishers, and researchers retrieve and analyze the papers in order to create new results which go back into the same cycle.

While researchers can read the papers, they must rely on citations and the likely rare links in those papers (B in Fig. 1), in order to discover, access, and use the original data. If they can discover and access the original data, they would likely need more detailed metadata, e.g. parameter names and descriptions, processing history, data quality information, etc., to understand and trust the data and, of course, they would need to be able to access and read that metadata and associated data regardless of the format in which they are provided.

While the nature of the resources, i.e. data or journal articles, is significantly different between DH and COVID-19, in both crises there is a strong and immediate focus on tools needed for discovery of relevant resources for purposes that may not have existed at the time the data were generated. In the DH case, these tools were typically GIS based (internet mapping or desktop) and examples still persist (e.g. the NOAA Gulf of Mexico Data Atlas [9], the Gulf of Mexico Research Initiative [10], or the NOAA Resource Damage Assessment and Restoration Data Integration Visualization Exploration and Reporting [11] tools). In the COVID-19 case [3], they are mostly text search and discovery tools powered, in some cases, by machine learning and artificial intelligence. These tools work because the resources are generally in one format (Portable Document Format, PDF [12]). A decade ago, during the DH disaster, online mapping was in its infancy and many of the tools provided were pushing the online mapping envelope. Today, the tools being used for COVID-19 data are pushing the natural language processing and machine learning envelope. In the pandemic case, semantic tools are being promoted without detailed knowledge of their immediate utility as compared to other approaches or in comparison to how article-level metadata would speed new developments. In both cases, critical data and metadata needs drive data discovery and access capabilities forward, building capabilities, resources and resiliency for the next crisis.

4. The role of metadata in developing policy

Gaps in metadata for article-level content or for data sets have, as the two previous examples have discussed, impeded the speed of new and unique research or the thorough examination of data sets chronicling a key moment of crisis. These critical research and data needs have also brought to light the importance of metadata for social and community policy making and the tools needed to make social science and humanities research open and available. Metadata may not directly drive policy or social decision making, but they are part of a pipeline from the inception of research to the broad use of data and published work. Individual pieces can make significant changes in what research is pursued and what results are concluded from collected data.

Metadata begin to reveal their influence the moment a researcher enters terms in a database, discovery platform, or catalog. They make it possible for more accurate research results to be found more quickly, especially where made openly available and in conjunction with open content. However, the interdisciplinary open COVID-19 collections produced by aggregators who generally focus their collections on

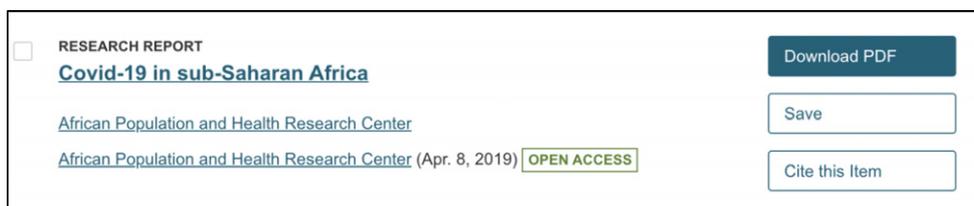


Fig. 2. Paper displaying incorrect publication date of Apr. 8, 2019 (Accessed May 9, 2020).

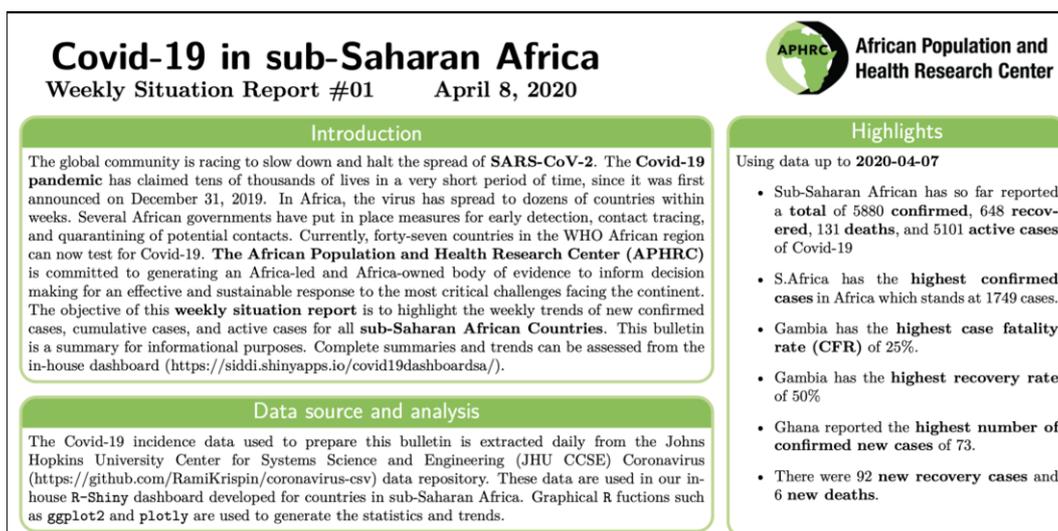


Fig. 3. Same paper as in Fig. 2 displaying correct publication date of Apr. 8, 2020 (Accessed May 9, 2020).

social science and humanities research do not fundamentally change the metadata landscape. In Fig. 2, notice the year of publication listed is 2019 rather than 2020. On the publication itself the year is correctly identified as 2020 (see Fig. 3). This error can potentially trip up URL resolvers that use open URL construction to build a link resolver from article metadata indexed in a platform. For a researcher, the article could potentially be suppressed from a list of resources available in JSTOR if they look for research published inside of a particular date range. The speed at which these collections were released is to be lauded, but as this small error shows, one small piece of errant metadata can contribute to data misrepresentation.

Presuming that the research process is not hindered by errors inside a platform and that a researcher has been thorough with search strategies, discovery is also affected by large-scale silos that persist not only among disciplines, but also across content providers. Fragmented infrastructure keeps content repositories (repositories or journals) from being linked to other content repositories or to keepers of metadata central for discovery at large-scale (e.g. Abstracting & Indexing services, library catalogs, funding metadata, researcher metadata) [13]. For humanities and social science scholars, this means that they will not discover published research relevant to their topic without combing through multiple platforms. Lack of linkage across content aggregators and discovery metadata is further compounded by paywalls and the need for university credentials.

Large-scale cross-disciplinary collection aggregators such as ProQuest, EBSCO, JSTOR, and Project-Muse help scholars discover content outside of disciplinary boundaries, but it still takes scholars who are especially curious to pursue authoritative research outside of their own specialties. One particularly intrepid paper by Emily Oster identifies possible connections between economic conditions and “violent scapegoating”, in which she draws plausible connections between the increase in witch trials between 1520 and 1770 and the decrease in temperature (the so called “little ice age”) and/or slow economic growth at the same time [14]. The research gathered for this study is drawn from historical, economic, and meteorological data, for which none of the resources would have been found in the same abstracting and indexing database. Note that even in this case, this article was found not through an abstracting and indexing service, but through previous exposure to Oster’s recent books *Expecting Better* [15] and *Cribsheet* [16]. In other words, repository or database design combined with metadata chosen for indexing often does not allow for significant serendipitous cross-discipline discovery; it takes a researcher who knows where to look in an ever-expanding universe.

When research is done well and thoroughly the results can change the way humans behave individually and as a society. The COVID-19 pandemic has given rise to retrospective examinations of previous moments of large-scale illness. Among the previous illnesses currently being discussed are the H1N1 flu, SARS, MERS and the 1918 “Spanish” flu. Of these, the 1918 flu and its social effects bears the greatest similarities to the situation in which we now find ourselves (at least for the Western world). For humanities and social science scholars asked to weigh in on the historical record for the 1918 flu in order to help drive current policy or to reopen the dialogue about this event as it pertains to COVID-19, robust metadata about contemporary newspaper reporting or scholarship are essential to execute the discovery, research, and publication cycle. While metadata may not always be able to provide full context within history, good metadata can provide links to resources that can.

During the first and second quarters of 2020, anecdotal stories have been shared in the news about the effects of social distancing and reopening society too soon. St. Louis, MO’s successful social distancing strategies in 1918 have been cited as an example to follow in this current pandemic [17]. Meanwhile, Richmond Virginia successfully began to “flatten the curve” during the same period, only to reopen the city too soon (see video recording by Ed Ayers [18]). The data fueling both of these stories is based in part on newspaper reporting from 1918 and population data about deaths, but it takes a good researcher to bring these points together and great metadata to bring these topics to the surface in a database search. And, as the German Ethics Council has recently shown, government-organized panels can leverage the expertise of ethicists, religious scholars, historians and others to weigh in on the scope and duration of the social distancing mandated by the government to provide important perspectives and predictions [19].

5. What can metadata do?

Times of crisis stress test how research gets done. Crises can highlight deficiencies in the metadata supply chain and provide a sense of urgency to address them, to facilitate faster research and to make the formulation and implementation of policy better informed. The examples included here share some common themes that illustrate the power and importance of metadata and why its quality is so important. Unfortunately, efforts borne of crises are too often quick fixes or workarounds, providing only temporary access or short-lived collaborations that dissolve when the emergency is over. Other events show a post-crisis pattern of settling back into legacy approaches and insufficient workflows. After all, at any given point in time, the community has only existing metadata to rely on and even successful improvements

most commonly apply only to newly created metadata, rather than enhancing the existing record. Without sustained efforts to improve the full corpus of metadata, we can expect the pattern to repeat itself with the inevitable next crisis. It is said that one should never let a crisis go to waste. A recent article in the *Chicago Booth Review* outlines “Four ways to ensure innovation continues after the crisis” [20]. We should apply principles like these to our thinking about metadata, for example, challenging orthodoxies, forcing constraints, working outside of organizations, and moving quickly, perfecting later.

Metadata 2020’s aims all along have been to address two overarching goals. First and foremost, to create awareness of the significant challenges related to insufficient and incomplete metadata and the important benefits that can be achieved through improvement. And second, to advocate for action through the creation of tools and resources for creating richer, connected, reusable and open metadata. Bringing the wider community together for the in-depth conversations necessary to retool for the longer-term has been central to the focus of Metadata 2020’s efforts. It was clear when the initiative started in 2017, based on interviews with a cross-section of the community, that the case for change is fundamental, will be ongoing, and must be coupled with resources in order to achieve positive, collective improvements that meet diverse community needs, while recognizing the diversity of community approaches. In this final year of the initiative, it is worth reviewing those resources here. A literature review [21,22] nicely summarizes the roles and challenges within each community and may be considered in the context of a great variety of existing best practices [23] that were collected near the start of the initiative, along with use cases gathered over the past few years. Cross-community personas [24] were developed from interviews, surveys and community group discussions to help illustrate and validate the resources provided, particularly the Metadata 2020 Principles [4] and Best Practices [25]. During the 3 years of the Metadata 2020 initiative, the lessons learned have been integrated into the outputs that have been vetted by the community through direct participation in their development and feedback on draft versions.

This work necessarily builds upon and supports a number of existing efforts, so it is important to recognize the achievement of like-minded initiatives, perhaps most notably FAIR [7]. Great strides have been made in the recognition of metadata as a strategic yet underserved area of scholarly communication [26]. But, perhaps because of the great diversity of the larger research community and because the focus of open scholarship is understandably on content, a central, unifying advocate for open, enriched metadata has been lacking.

6. What is being done now and what can be done for the future?

Metadata serve as a way to query research literature. Even when metadata are insufficient, they enable tools such as the aggregator and bibliographic databases mentioned above to serve as much-needed services. PubMed has yet to be overtaken by the likes of Google and spends considerable resources, including human curation, to maximize the utility of its metadata. Considering the volume of metadata in relation to full text, it is clear that metadata can make language translations much easier.

Assessing whether existing tools and approaches might suffice in the next crisis is really a judgment on whether the current metadata landscape supports 21st century research. The answer, clearly, is ‘no’. Metadata 2020 has advocated for the entire community to recognize that making improvements today is an investment in resiliency for the next crisis and a vastly-improved research ecosystem. We have seen what commitment and investment have done for the Human Genome Project, providing a springboard for many new discoveries and cures. In the case of metadata, much of what is needed is already in place, with far less investment required and a payoff shared across a wider-scope of disciplines.

7. If metadata are the accelerant, people are its engine

It is a central tenet of this effort that a critical mass of advocacy and awareness is necessary to bring about systemic changes that improve the efficiency of research and confidence in the systems supporting it. Friction is perhaps most noticeable in a crisis, but it is no less of a drag on finding solutions to problems under normal circumstances. Given the inherently technical nature of metadata and the scope of its reach, it is necessary to outline how everyone in each community can participate, not just practitioners and experts.

Metadata 2020 describes four personas involved in the metadata landscape: Creators, providing descriptive information (metadata) about research and scholarly outputs; Custodians, storing and maintaining this descriptive information and making it available for consumers; Curators, classifying, normalizing, and standardizing this descriptive information to increase its value as a resource; and Consumers, knowingly *or unknowingly* using the descriptive information to find, discover, connect, and cite research and scholarly outputs. These personas exist across all metadata-involved organizations in scholarly communication and can make near-term, positive changes with few additional resources. For example, Creators, such as publishers, are in a good position to do outreach with authors on the importance of metadata to enhance discoverability of their work. They can also collaborate with Consumers on feedback loops for reporting metadata errors and enhancements. Custodians of metadata are custodians of the scholarly record and so are well-positioned to expose the work of Creators. Librarians, for example, may partner with Creators on author outreach. Curators are perhaps uniquely positioned to tailor metadata for community-specific needs, such as in subject databases and data repositories. The Metadata 2020 Principles and Best Practices serve as a framework for everyone to take action to improve metadata.

8. Conclusion

At this time of global crisis, the need to enhance the critical infrastructure provided by metadata is clear. It requires many hands. FAIR and the Metadata 2020 Principles and Best Practices need to become less aspirational, more common practice. Many communities play a part - researchers, publishers, librarians, data publishers repository managers, services providers and funders. In order for metadata to be an accelerant to bettering society, we need to recognize how everyone moves across the personas as we engage in our daily work. Recognizing how people in other communities also move through these personas will help us to reach out and work together to address the tremendous challenges that we face. Looking to these changes as longer-term goals rather than taking action today would mean accepting the pattern of metadata crisis management, rather than changing the status quo. One of the lessons to be learned from the COVID-19 pandemic is the need to start working now to prepare for the next crisis. Improving how we handle metadata is a critical part of that preparation.

Acknowledgements

The authors would like to recognize the hundreds of global contributors to Metadata 2020. The insights offered here were shaped through discussions, writing, and sharing by members of this distinguished group. We also recognize the March 2020 NISO Plus conference that catalyzed the formalization of these ideas in the form of a paper. We also are thankful to Crossref, both for initiating the creation of the Metadata 2020 project and for supporting its work.

References

- [1] D. Carr, Sharing research data and findings relevant to the novel coronavirus (COVID-19) outbreak, [Internet], Wellcome Trust, 2020, Available from: <https://wellcome.ac.uk/coronavirus-covid-19/open-data>, last accessed June 6, 2020.
- [2] G. Santangelo, New NIH Resource to Analyze COVID-19 Literature: The COVID-19 Portfolio Tool, [Internet]. National Institutes of Health Office of Extramural Research, Extramural Nexus, 2020, Available from: <https://nexus.od.nih.gov/all/2020/04/15/new-nih-resource-to-analyze-covid-19-literature-the-covid-19-portfolio-tool/>, last accessed June 6, 2020.
- [3] About CORD-19, [Internet]. *Semantic Scholar*, 2020, Available from: <https://www.semanticscholar.org/cord19/about>, last accessed June 6, 2020.
- [4] K. Kaiser, J. Kemp, L. Paglione, H. Ratner, D. Schott and H. Williams, Methods & proposal for metadata guiding principles for scholarly communications, *Res Ideas Outcomes [Internet]*, Available from: <https://riojournal.com/article/53916/>, last accessed June 6, 2020.
- [5] N. Carr, Situational overload and ambient overload, [Internet]. *Rough Type* (Blog) 2011, Available from: <http://www.roughtype.com/?p=1464>, last accessed June 6, 2020.
- [6] R. Borah, A.W. Brown, P.L. Capers and K.A. Kaiser, Analysis of the time and workers needed to conduct systematic reviews of medical interventions using data from the PROSPERO registry, *BMJ Open [Internet]* 7(2) (2017), e012545, Available from: <http://bmjopen.bmj.com/lookup/doi/10.1136/bmjopen-2016-012545>, last accessed June 6, 2020.
- [7] M.D. Wilkinson, M. Dumontier, I.J. Aalbersberg, G. Appleton, M. Axton and A. Baak, The FAIR Guiding Principles for scientific data management and stewardship, *Sci Data [Internet]* 3(1) (2016), Available from: [10.1038/sdata.2016.18](https://doi.org/10.1038/sdata.2016.18), last accessed June 6, 2020.
- [8] I.J. Borges do Nascimento, N. Cacic, H.M. Abdulazeem, T.C. von Groote, U. Jayarajah, I. Weerasekara et al., Novel coronavirus infection (COVID-19) in humans: A scoping review and meta-analysis, *J Clin Med [Internet]* 9(4) (2020), 941, Available from: <https://www.mdpi.com/2077-0383/9/4/941>, last accessed June 6, 2020.
- [9] *Gulf of Mexico Data Atlas* [Internet], National Centers for Environmental Information, National Oceanic and Atmospheric Administration Station, 1985. Available from: <https://www.ncei.noaa.gov/maps/gulf-data-atlas/atlas.htm>, last accessed June 6, 2020.
- [10] *Gulf of Mexico Research Initiative*, GRIIDC [Internet], Available from: <https://data.gulfresearchinitiative.org/pelagos-symphony/data-discovery>, last accessed June 6, 2020.
- [11] *Natural Resource Damage Assessment & Restoration Data & Visualization* [Internet], National Oceanic & Atmospheric Administration, NOAA, Available from: <https://www.diver.orr.noaa.gov/home>, last accessed June 6, 2020.
- [12] PDF. Three letters that continue to change the world [Internet], *Adobe* 2020, Available from: <https://acrobat.adobe.com/us/en/acrobat/about-adobe-pdf.html>, last accessed June 6, 2020.
- [13] A. Meadows, (NISO). Building a Sustainable Research Infrastructure, [Internet], *The Scholarly Kitchen*, 2020, Available from: <https://scholarlykitchen.sspnet.org/2020/04/30/building-a-sustainable-research-infrastructure/>, last accessed June 6, 2020.
- [14] E. Oster, Witchcraft, weather and economic growth in renaissance Europe, *J Econ Perspect [Internet]* 18(1) (2004), 215–228, Available from: <http://pubs.aeaweb.org/doi/10.1257/089533004773563502>, last accessed June 6, 2020.
- [15] E. Oster, *Expecting Better: Why the Conventional Pregnancy Wisdom Is Wrong—and What You Really Need to Know*, first ed. Penguin Books, 2014, p. 352.
- [16] E. Oster, *Cribsheet: A Data-driven Guide to Better, More Relaxed Parenting, From Birth to Preschool*, first ed. Penguin Books, 2020, p. 352.
- [17] R. Smith, Social measures may control pandemic flu better than drugs and vaccines, *BMJ [Internet]* 334(7608) (2007), 1341, Available from: <http://www.bmj.com/lookup/doi/10.1136/bmj.39255.606713.DB>, last accessed June 6, 2020.
- [18] E. Ayers, What happens when you reopen too soon? [Internet], *The Future of America's Past* (Video), VPM, PBS Documentaries, 2020, Available from: <https://futureofamericaspast.com/>, last accessed June 6, 2020.
- [19] D. Matthews, German humanities scholars enlisted to end coronavirus lockdown, *Times Higher Education [Internet]* (2020), Available from: <https://www.timeshighereducation.com/news/german-humanities-scholars-enlisted-end-coronavirus-lockdown>, last accessed June 6, 2020.
- [20] L. Lyman, Four ways to ensure innovation continues after the crisis, *Chicago Booth Review* (2020), Available from: <https://review.chicagobooth.edu/strategy/2020/article/four-ways-ensure-innovation-continues-after-crisis>, last accessed June 6, 2020.
- [21] W. Gregg, C. Erdmann, L. Paglione, J. Schneider and C. Dean, A literature review of scholarly communications metadata, *Res Ideas Outcomes [Internet]* (2019), 5, Available from: <https://riojournal.com/article/38698/>, last accessed June 6, 2020.

- [22] S. Daawson, C. Erdmann and W. Gregg, Metadata Article Collection, [Internet], *ScienceOpen*, Available from: <https://www.scienceopen.com/collection/Metadata>, last accessed June 6, 2020.
- [23] Links to Best Practices and Guidelines of the Scholarly Communications Community, [Internet], Metadata 2020, 2018, Available from: <http://www.metadata2020.org/resources/metadata-best-practices/>, last accessed June 6, 2020.
- [24] The Metadata 2020 Personas, [Internet]. Metadata 2020, Available from: <http://www.metadata2020.org/resources/metadata-personas/>, last accessed June 6, 2020.
- [25] The Metadata 2020 Practices [Internet], Metadata 2020, Available from: <http://www.metadata2020.org/resources/metadata-practices/>, last accessed June 6, 2020.
- [26] A. Stokes and S. Weston, *New Report on Digital Transformation in Publishing Reveals Sluggish Progress: Only 25 Percent of Publishers See Themselves as Ahead of Industry Peers*, Copyright Clearance Center [Internet], 2017, Available from: <http://www.copyright.com/new-report-digital-transformation-publishing-reveals-sluggish-progress-25-percent-publishers-see-ahead-industry-peers/>, last accessed June 6, 2020.