RESEARCH ARTICLE

JASIST WILEY

Essential work, invisible workers: The role of digital curation in COVID-19 Open Science

Irene V. Pasquetto[1] | Amina A. Abdu[2] | Natascha Chtena[3]

[1]College of Information Studies, University of Maryland, College Park, Maryland, USA

[2]School of Information, University of Michigan, Ann Arbor, Michigan, USA

[3]Publishing Program, Simon Fraser University, Burnaby, British Columbia, Canada

Correspondence
Irene V. Pasquetto, College of Information, University of Maryland, College Park, MD 20742, USA.
Email: irenevp@umd.edu

Abstract

In this paper, we examine the role digital curation practices and practitioners played in facilitating open science (OS) initiatives amid the COVID-19 pandemic. In Summer 2023, we conducted a content analysis of available information regarding 50 OS initiatives that emerged—or substantially shifted their focus—between 2020 and 2022 to address COVID-19 related challenges. Despite growing recognition of the value of digital curation for the organization, dissemination, and preservation of scientific knowledge, our study reveals that digital curatorial work often remains invisible in pandemic OS initiatives. In particular, we find that, even among those initiatives that greatly invested in digital curation work, digital curation is seldom mentioned in mission statements, and little is known about the rationales behind curatorial choices and the individuals responsible for the implementation of curatorial strategies. Given the important yet persistent invisibility of digital curatorial work, we propose a shift in how we conceptualize digital curation from a practice that merely "adds value" to research outputs to a practice of knowledge production. We conclude with reflections on how iSchools can lead in professionalizing the field and offer suggestions for initial steps in that direction.

1 | INTRODUCTION

In this paper, we discuss the prevalence, transparency, and visibility of digital curation practices in open science (OS) initiatives during the COVID-19 pandemic. By exploring to what extent, how, and by whom research outputs like data, software, and publications were curated for reuse and public consumption, we characterize the role that digital curation practices and practitioners played in executing such initiatives.

Digital curation practices—such as assessing, preserving, and documenting research information and tools—enable research outputs to travel between contexts and be reused by a variety of communities (Borgman, 2015;

Hemphill et al., 2022; Leonelli, 2020). Additionally, curatorial practices actively contribute to increased scientific reliability and reproducibility. By preserving, documenting, and communicating the sociotechnical settings in which research outputs emerge, curatorial practices inform reusers about the potential of research outputs to contribute to knowledge building and scientific progress, as well as about the intrinsic limitation of said outputs (Leonelli, 2019b; Leonelli et al., 2021). Thus, the reliability of analytical results and subsequent interpretations will be largely dependent on the quality and accuracy of these curatorial activities and processes.

In an early pandemic opinion paper, Shankar et al. (2021) reflected on some of the challenges and

opportunities for digital curation in light of the COVID-19 pandemic. The authors pointed at how digital curation can promote public trust in institutions, but also noted the widespread need for formal academic education and credit for curatorial interventions, and for a coordinated engagement around data ethics and security. Drawing on Shankar et al.'s commentary, we empirically researched how OS initiatives engaged with such opportunities and to what extent curatorial work was characterized by the challenges identified by Shankar et al. While Shankar et al. mostly focused on digital curation of research data, we extended our analysis to include curatorial efforts intended at preserving, documenting, and disseminating research publications and software.

In Summer 2023, we conducted a content analysis of available information regarding 50 OS initiatives that emerged—or substantially shifted their focus—between 2020 and 2022 to address COVID-19 related challenges. Our findings suggest that the reach and success of the projects was often proportional to the amount of curatorial activities. In other words, the most visible and most trusted OS initiatives were also the ones most heavily invested in digital curation. Such initiatives were also the ones with the most robust funding and organizational support. Yet, besides some notable exceptions, curatorial practices were rarely included in the mission statements of the initiatives, and, overall, little is known about the individuals and teams responsible for conducting such practices. Also, even among highly visible, well-funded projects, only a handful of projects reported on the rationales behind curatorial decisions.

We propose that the root of this paradoxical state of affairs—digital curation being at once essential and invisible—might reside in the fact that, in OS initiatives, digital curation work is often overshadowed by other forms of intellectual work, both at the rhetorical and epistemological level. At the rhetorical level, OS initiatives tend to promote goals of enabling research speed, effective communication, and consensus, and, at the epistemological level, initiatives tend to highlight achievements in data analysis, synthesis, and data visualization. In particular, the epistemic focus on data analysis instead of curation might be partially due to the fact that digital curation as a field has not been effective in clarifying—and arguing for—its relevance in enabling processes of knowledge production, and, more often, tends instead to present itself as a sub-discipline of data science. An alternative approach—we propose—would be shifting from a conceptualization of digital curation as a practice that "adds value" to research outputs, to one of digital curation as a practice of knowledge production. We close the paper with a reflection on how iSchools should and could lead the way towards a professionalization of the field, and give some suggestions for first, small steps.

2 | RELATED WORK

2.1 | The role of digital curation in Open Science

Like all forms of information (Gitelman, 2013; Loukissas, 2019), research outputs are shaped and directed by local epistemic cultures and context-specific norms (Cetina, 1999). When research outputs are properly preserved and documented, they become potentially reusable across a variety of research settings in unforeseen ways, and thus amenable to generating novel knowledge (Leonelli, 2020; Pasquetto et al., 2017, 2019). Broadly defined, digital curation consists in a set of practices that make research outputs interpretable and reusable across different research settings (Hemphill et al., 2022). Many definitions of digital curation exist. For example, in 2015, a working group of digital curation stakeholders defined digital curation as "the active and ongoing management and enhancement of digital assets for current and future reuse" (National Research Council, 2015). In practice, the term digital curation is used to refer to any stage of content preparation preceding analysis, including selection, appraisal, and categorization of content; data wrangling, cleaning, formatting; metadata schemas, authoritative vocabulary, and ontologies development; and preservation and sustainability policy development (Digital Curation Centre, n.d.). By documenting and communicating locality of scientific outputs, this set of practices ensures that science outputs are not only properly interpreted, but also trusted across different epistemic cultures (Lee & Stvilia, 2017).

OS projects enable processes of information exchange and integration by making scientific information and its close associates (software, etc.) available at no cost, as quickly as possible (Woelfle et al., 2011). Thus, digital curation activities are a critical part of OS projects, as, without curation, available resources can end up being unused or, worse, misused (Hastings, 2021; Thomer et al., 2022; Yakel, 2007). In certain instances, curation of open research artifacts can shape the research agenda of an entire discipline. Delfanti (2016), for example, showed that, in order to be recognized as legitimate and productive members of their community, physicists must abide by the curatorial practices, and in particular the classification schema, structured by arXiv.org—a prominent preprint server in physics. Digital curation can also serve as a means to limit or prevent misuses and

misinterpretations of open research data and publications (Ćurković et al., 2021). During the COVID-19 pandemic, for example, preprint digital archives designed new curatorial interventions aimed at avoiding misinterpretation of COVID-19 preprints and their associated data (Yan, 2020). These interventions included adding filtering mechanisms to the publication process, as well as creating cautionary labels for the preprints' download pages (Kwon, 2020).

The importance of curatorial practices in OS is generally recognized among funding agencies, governments, and, increasingly, OS researchers and practitioners. For example, best practices for curating research data for reuse are being adopted worldwide (Wilkinson et al., 2016), and discussions about how to develop appropriate digital curatorial practices in ever-changing scientific settings are prominent at OS conferences such as FORCE 11 and the annual meeting of the Research Data Alliance. Still, when it comes to allocating resources for OS projects, and research in general, typically the efforts necessary to make research outputs trustworthy and reusable (i.e., curation) cannot count on sufficient support, both in terms of financial resources and workforce availability (National Research Council, 2015). In particular, a lack of financial support for digital curation threatens both the funding of necessary training programs in digital curation and the ability to staff and maintain OS projects. Thus, OS projects often rely on unpaid, voluntary labor in order to fulfill curatorial tasks (Darch, 2014). Darch (2014) notes that establishing the credibility of volunteer-produced scientific products presents an important challenge to the funding and use of citizen science projects. He points to crediting methods for volunteer curators as a tactic for enhancing credibility and trust. Yet, credit for curatorial works remains a persistent problem in OS projects. Authors in the field of digital curation point at the lack of visibility and credit for curatorial work as key drivers of the field of practice's limited ability to attract needed resources. In particular, researchers have stressed the importance of making visible both curatorial activities and the craft involved in order to support the success and continued development of curatorial infrastructures (Plantin, 2019; Thomer et al., 2022).

2.2 | Open Science during the COVID-19 pandemic

The COVID-19 pandemic provided opportunities for OS to accelerate scientific discovery, enable equitable participation globally, and enhance the public understanding of science (Alemneh et al., 2020; Hastings, 2021; Liu et al., 2022; Verovšek & Gorišek, 2023; Weisenberg, 2023). Thus, the pandemic has resulted in a significant increase in OS products, including open access (OA) publications, open data, and open software projects (Tse et al., 2020). Notable examples range from data repositories and analysis tools like Nextstrain,[1] a scientific collaboration between researchers to facilitate the use of pathogen genome data, to data visualizations like The Atlantic's Covid Tracker and preprint review initiatives like *Rapid Reviews Infectious Diseases*[2] (formerly known as *Rapid Reviews/COVID-19*), an OA overlay journal published by MIT Press in collaboration with researchers at UC Berkeley. Existing work shows that OS efforts have been far from uniform. OS projects demonstrated a variety of goals including assessing the spread of COVID-19, informing the public, predicting the future, and supporting decision-making. Many COVID-19 dashboards, one of the most common pandemic OS products, were built primarily for informative purposes (e.g., the Covid-19 Data Explorer[3] or the Novel Coronavirus (COVID-19) Infection Map[4]), while a smaller percentage attempted to support decision-making (Khodaveisi et al., 2023). The Institute For Health Metrics and Evaluation's Covid-19 Projections[5] dashboard, for example, was intended to "help hospitals and policymakers plan how to allocate resources." As noted by Ivanković et al. (2021), "clear links between current trends and past policy decisions and individual behavior" (p. 13) can help facilitate decision-making around specific actions, but have often been limited or missing from OS projects (Barbazza et al., 2021; Bos et al., 2021). Projects also varied in their intended audience: while many OS projects during the pandemic were intended for a general audience, others were designed with researchers, policymakers, or healthcare professionals in mind (Khodaveisi et al., 2023). We observed this in our own data, finding for example, that the WHO's Covid-19 Research Database[6] was created specifically with researchers in mind, while other projects like the Johns Hopkins COVID-19 Dashboard[7] targeted a wider range of stakeholders, including the general public. Ivanković et al. (2021) observed that awareness of the audience and their needs is a key part of creating actionable OS products. Challenges in the implementation of OS during the pandemic included reduced research quality, overburdened publishing infrastructures, language barriers (Homolak et al., 2020; Kraemer et al., 2021), lack of awareness of OA resources (Matonkar & Dhuri, 2021), and perceived legal restrictions around demographic data collection (Ivanković et al., 2021). Besançon et al. (2021) argued that many challenges stemmed from an incomplete

implementation of OS principles, with many researchers falling short of fully embracing crucial tools like pre-registration, data-sharing, code-sharing, and open review. The authors point to scientific waste (e.g., due to poorly designed studies or duplicate efforts), conflicts of interest and lack of quality control during fast-tracked review processes, distrust of results, retractions, and misuses of preprints in science communication.

In light of such opportunities and challenges, Shankar et al. (2021) highlighted the importance of adopting proper digital curatorial practices for the successful implementation of OS during the pandemic. In their early pandemic paper, the authors noted that digital curation, when implemented as a collective action, could fill research infrastructure gaps revealed by the pandemic. Shankar et al. argue that curation can also enhance science communication, and they emphasize the effects that curatorial control and transparency about curatorial decisions can have on public trust in science. They also highlight that curatorial activities can help with ensuring data protection and privacy. Because of the important role curators play in the production, communication, and ethics of OS, Shankar et al. argue that the work of curation should be credited and properly funded. They also suggest that iSchools should adopt educational programs in digital curation to properly prepare students to deal with the challenges they identified.

In this paper, we draw on Shankar et al.'s observations about the value of digital curation during the pandemic to empirically investigate its actual role in enhancing, or even enabling, COVID-19 related OS initiatives. We explore to what extent, how, and by whom research outputs like data, software, and publications were curated for reuse and public consumption. We center our research on four aspects of digital curation identified by Shankar et al. prevalence, visibility, transparency, and credit. Finally, we extend their arguments about data curation to other forms of digital curation including research publications, research code, research tools, and data aggregators.

Our guiding questions are:

1. To what extent were digital curation practices prevalent and visible in open science initiatives during the COVID-19 pandemic?
2. How transparent were the motivations and rationales driving curatorial practices and processes in these initiatives? And how were these documented and communicated to users?
3. How, in what ways, and to what extent was credit provided for curatorial work conducted in and for these initiatives?

3 | METHODOLOGY

3.1 | Data collection and analysis

In order to understand the prevalence, visibility, transparency, and credit attributed to digital curation practices in OS initiatives during the COVID-19 pandemic, we conducted a content analysis of available information regarding 50 OS initiatives that emerged—or substantially shifted their focus—between 2020 and 2022 to address COVID-19 related challenges. We collected most of the information about such initiatives on their official websites, partners' websites, media outlets, and related academic publications.

Using keyword search on Google Search and Google Scholar, we analyzed search results about open science initiatives emerging from the first two pages of search results, for each search. Specifically, we searched for initiatives that included OS, open data, or open software related keywords—*open science*, *crowdsourced (or crowd-sourced) science*, *open access*, *preprint server*, *preprint platform*, *open peer review*, *rapid peer review*, *open data*, *dashboard*, *content aggregator*, *database*, *dataset*, *data repository*, *data archive*, *data sharing*, *open software*, *free software*, *software repository*, *software archive*—in combination with COVID-19 related keywords—*covid-19*, *covid*, *SARS-CoV-2*, *coronavirus*, *pandemic*. When we reached 50 initiatives, we noticed that finding new initiatives was becoming particularly laborious as few new initiatives were emerging and those that were limited in both scope and visibility. As we reached saturation, we stopped collecting initiatives.

Because the sample was derived from common search terms typically closely aligned with the mission of reputable open science initiatives, we believe that we gathered the most popular initiatives. Even if some were missed, we believe we captured the majority of popular initiatives based on the repeated appearance of initiatives across search results. It is highly unlikely that significant initiatives were excluded, and if any were, they would represent a small fraction, indicating only a minimal bias in our selection.

Types of initiatives included: data-related initiatives (such as data repositories, dashboards, and visualization tools), publication-related initiatives (such as content aggregators, publication repositories, and peer-review initiatives), and other initiatives (such as software tools and best practice recommendations). The search was conducted in English and only English language initiatives were included. Please consult our data collection's spreadsheet to see the initiatives that we selected and analyzed. Please consult the codebook for the definition of each variable studied.

TABLE 1 An overview of the initiatives and key variables in our sample by initiative type, with examples for illustrative purposes.

Initiative type	Total	New	Active	Curation mentioned	Example
Dashboards and data visualizations	17	15	9	2	Covid Tracking Project (The Atlantic)
Data repositories	14	12	7	4	Covid-19 Nursing Home Data (CMS)
Publication aggregators, collections, repositories, and search tools	13	13	10	6	LitCovid (NCBI)
Open software and applications	4	3	4	0	MicrobeTrace (CDC)
Peer-review initiatives	3	2	2	0	Rapid Reviews Covid-19 (MIT and UC Berkeley)
Other	2	2	1	0	Guidelines for Data Sharing (Research Data Alliance

Note: The full data set is available on GitHub. Note that the total sums to more than 50 because some initiatives are counted across two different initiative types.

Once we selected the initiatives, we developed a codebook that contained a series of variables to be identified in said initiatives. The codebook was organized around our themes of interest surrounding digital curation in OS: prevalence, visibility, transparency, and credit. We collected 29 variables for each initiative, for a total of 1450 collected and analyzed variables. The second author collected and coded all the initiatives (Table 1). The coding author met weekly with the rest of the team to discuss interpretation of findings and ensure homogeneity in the coding process.

Key coded variables that we collected for each one of the 50 initiatives included:

- The type of initiative
- The type of leadership
- Launch and end dates
- The goals and challenges involved
- Information about data sources
- Information about editorial and curatorial decisions
- Access to code and underlying data
- Access to tools/materials/software
- Credit type
- Credit visibility
- Contributors
- Privacy
- Curation type mentioned
- Contributions mentioned
- Associated DOI
- Number of Citations (Web of Science)
- Number of Citations (Google Scholar)
- Media Mentions (MIT Media Cloud)

3.2 | Limitations

We conducted a content analysis of available online information in order to identify to what extent OS initiatives provided information over how, by whom, and when digital curation was practiced. In doing so, we used OS initiatives' choices over availability of online information as a proxy of how each initiative valued and conceptualized digital curation as a professional practice. However, this methodology presents limitations. Sometimes, found information was not complete or unclear. Also, choices about how to present digital curation work on official websites can be only partially reflective of the actual ways in which OS practitioners value digital curation. For example, we discuss how lack of transparency about rationales behind digital curatorial practices can be related to a research culture within OS that undervalues curation as at the epistemological level. However, it might be that initiatives had other reasons not to unpack and present such choices.

Also, available online content does not provide us with complete information about actual processes of curation and the distribution of such processes among the initiatives and their partners. For most initiatives, we could only partially reconstruct the data curation life cycle: many initiatives utilize data from multiple sources, especially if the initiatives are pulling data from major health agencies (CDC, WHO), a bulk of curation work could have happened there, which may contribute to ambiguities about how to represent these activities.

We partially mitigated this limitation by, when in doubt, contacting the initiative's leadership to inquire about the meaning of the information we found online. Additionally, our close, personal knowledge of the field helped us to articulate an interpretation of the findings that is grounded in actual digital curation practices in academia and in OS alike. Two of the authors have extended experience conducting longitudinal, ethnographic work on OS initiatives, and one author has been conducting research on digital curation practices for over 10 years.

Another potential limitation is that the study analyzed OS initiatives that took place at a particular

moment in time, the COVID-19 pandemic, where the need to release science outputs as fast as possible might have altered the priorities of such initiatives. In other words, it is possible that, in a non-emergency situation, the same initiatives would have made different choices in terms of how to curate research outputs.

However, our findings and research design mitigate such a limitation in multiple ways. First, in relation to our research question on the prevalence of digital curation practices, we found that most of the initiatives greatly invested in curation, suggesting that even in an emergency situation digital curation is pivotal to the success and credibility of OS.

In addition, this limitation does not apply to our research questions related to the transparency and visibility of digital curation practices. We did not look at such initiatives during their developmental stage or launch stage or during the peak of the pandemic, where it would have made sense to invest in speed rather than in providing rationales over curatorial choices (transparency) and credit for curators (visibility). Instead, we collected available information on digital curation from their websites in summer of 2023, when most of the pressure to collect and share outputs as fast as possible had faded out.

Overall, we consider the focus on the COVID-19 pandemic as a strength, not only as a limitation, of our paper. We chose to look at COVID-19 for specific reasons. While the COVID-19 pandemic is a specific case, it reveals certain dynamics that characterize OS more broadly. Its specific features reflect larger issues that the OS community cares about, such as the need to make policy decisions and the need for accelerated science, and exposes broader problems that were already lurking in OS infrastructures, such as rushed reviews and publication of OS outputs and difficulties at curating and distributing content for multiple audiences at once.

Finally, another challenge we faced was the changing nature of these online OS initiatives. In particular, between when we began data collection and the end of our analysis, a number of initiatives became inactive, including some websites which disappeared altogether leading to broken links. In these cases, we relied on screenshots we captured at the time of our initial data collection along with archived versions of the initiative websites on the Wayback Machine. Occasionally, entire datasets were removed from websites: for example, Open-Aire removed their public list of data sources. We anticipate that the initiative webpages we analyzed will continue to change or be taken down in the future, an issue which speaks to the lack of sustainability of many pandemic OS projects. While it is expected that many of these initiatives would stop collecting and curating new data once the emergency faded out as these projects may

have outlived their purpose (Donovan, 2023), what was more surprising was that, for some initiatives, entire datasets became unavailable, or some initiatives disappeared from the Internet altogether.

4 | FINDINGS

4.1 | Prevalence and visibility of digital curation in Open Science projects

During the pandemic many OS initiatives emerged, often centered around data or publications. The most common initiatives were dashboards and data visualization tools—focused on communicating scientific information such as the number and location of cases, as well as data repositories and archives—centered on facilitating access to scientific data. Content aggregators and publication browsing tools, which provide a centralized location for storing and retrieving relevant papers, were also common. A variety of actors were involved in leading these initiatives. Universities, individual scientists, governments, nonprofits, publishers, and private corporations all led in the creation and curation of OS projects during the pandemic. In addition to these varied forms of leadership, participants in OS initiatives ranged from traditional researchers to trained volunteers and the general public who were able to contribute via email feedback, user submissions, GitHub, and direct project involvement.

The OS initiatives we analyzed shared many of the same goals. Notably, many initiatives emphasized increasing the speed of scientific progress. For example, GISAID's[8] goals included "enabling rapid and open access to epidemic and pandemic virus data" while Google's COVID-19 Search Trends Symptoms Dataset[9] aimed to "provide an earlier and more accurate indication of the reemergence of the virus in different parts of the country." In addition to speed, many initiatives cited an intent to reach beyond traditional scientific audiences in order to communicate with other stakeholders like government officials, healthcare professionals, and the public. For example, the Research Data Alliance's guidelines[10] targeted "policymakers and funders," while the COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University directed their efforts at "researchers, public health authorities, and the general public." Finally, a number of initiatives identified synthesis and consensus among their goals. HealthMap,[11] a dashboard created by Boston Children's Hospital, endeavored to "achieve a unified and comprehensive view of the current global state of infectious diseases." Technical aspects highlighted in mission

statements included enabling data analysis, making content accessible, and tracking content.

All OS projects practiced digital curation (albeit to different extents). However, curatorial practices were rarely mentioned among the goals or in the initiatives' mission statements. Most often, curation was a key, yet, for the most part invisible and challenging part of the projects. For example, rapid peer-review initiatives born or significantly expanded during the pandemic relied on curatorial work in order to make their publications easily findable. Articles approved for publication by rapid peer-review initiatives were not merely listed on the initiatives' websites, but organized and tagged by category to allow researchers to search through and find them easily. Organizing papers into categories to facilitate discovery requires thorough knowledge about multiple, and sometimes overlapping, research domains, and a detailed understanding of researchers' information seeking needs. Such curatorial work was as essential for the success of these initiatives as the rapid review process itself, as publications containing critical new findings about COVID-19 are of no use if researchers cannot find them. Yet, in most rapid review projects, little information was made available about curatorial activities, with the exception of the MIT-led *Rapid Reviews COVID-19 (RR:C19)*, which provided detailed information about editorial rationales, including rationales related to categorizing papers into disciplines.

For content aggregators, curation represented both a core activity enabling the projects, and the most daunting task. To fulfill their missions, content aggregators had to fit new, incoming content (publications, data, etc.) into pre-existing organizing structures (metadata schemas, semantic relations, relational databases, etc.). Given the speed of research during the COVID-19 pandemic, such organizing structures also required constant modification and updating. Thus, content aggregators relied on curatorial work not only to properly describe and document new content in relevant forms, but also to modify and adapt those organizing structures so that aggregators could readily welcome new content. Yet, while we observed practices of curation unfolding on aggregators' websites, we found little mention and description of such practices in their mission statements, project goals, or about pages. The NIH-led iSearch COVID-19 portfolio[12] was the only aggregator that identified "curation" as a key goal of the project. Instead, aggregators most often listed curatorial activities—such as categorizing, cleaning, and organizing digital content—among their key challenges. For example, LitCovid[13]—a curated literature hub developed by the NIH—lists "triaging papers into relevant categories as research evolves" as a key challenge of the project, and CORD-19[14]—a content

aggregator by the US-based Allen Institute—lists "handling data from multiple sources, cleaning metadata, and providing machine readable text" as challenges. This finding speaks to the fact that curatorial work is not a one-time activity, but an iterative process upon which the durability of these infrastructures depends. Information about who was responsible for curatorial work and how decisions are made was also generally scarce, with the exception of OpenAIRE COVID-19 Gateway,[15] an aggregator led by the European Commission that shared a spreadsheet that listed each dataset that the organization assessed, and information about whether and why each dataset was or was not included in the aggregator.

In distributed, decentralized collaborative projects—such as #coronavirussyllabus,[16] COVID-19 Social Science Research Tracker,[17] and WikiProject COVID-19[18]—curatorial work took the form of source selection, appraisal, description, and organization. These projects aimed at tracking and collecting existing content and resources on COVID-19 with the goal of making the research and educational community at large aware of such resources. To maximize chances of creating a comprehensive resource, they relied on participants from multiple disciplines and communities to contribute with their own content. In distributed, collaborative projects, decisions have to be made about who should be invited to contribute and how, how to judge validity, accuracy, and relevance of contributed content, and how to describe and organize such content to enable discovery through searching and browsing. Thus, curatorial activities were once again central to the fulfillment of these projects' missions, even though, once again, we found no mention of curation as an important component of these projects, with the exception of Wikipedia whose standards for source selection and appraisal are formalized and publicly shared (Wikipedia, n.d.).

While content aggregators bring together distributed content in one place to facilitate search and browsing by internet users, dashboards provide data-driven syntheses of a given phenomenon. Typically, the key feature of dashboards is to return easily interpretable visual overviews of the represented phenomenon. In a dashboard, digital content from multiple sources is transformed into a common format so that it can be visualized using centralized data visualization tools. The typical COVID-19 dashboard displayed not one but a set of multiple data visualizations that narrate different aspects of the pandemic. Due to the changing scientific understanding of COVID-19, significant international attention, shifting data collection practices, and high volume of data, pandemic dashboards had to be frequently updated with incoming data at a much faster rate than dashboards typically require. Many of the initiatives we looked at

encountered data inconsistencies arising due to instability of methods for data collection over time, particularly in periods of rapidly changing protocols. The shift towards at-home testing, for example, impacted testing and case number data. Dashboards also faced privacy challenges, as the data underlying these visualizations were obtained via different methodologies and relied on different case detection, testing strategies, and reporting practices. Data wrangling, cleaning, and reformatting enabled dashboards to compensate—at least partially—for this lack of standardization for research methodologies, formats, and ethics protocols. Multiple dashboards addressed the need for these curatorial activities and the challenges of implementing them on a daily basis, such as the WHO Coronavirus (COVID-19) Dashboard[19] and the COVID-19 Surveillance Dashboard.[20]

Open software products responding to the COVID-19 pandemic were much less common than OS products related to publications or data. One exception was the Pangolin tool,[21] a software initiative used to assign genome sequences to their most likely lineage. The Pangolin tool demonstrates that curatorial work can play an essential role in open software products: software initiatives like Pangolin not only involve the selection of tools and data to build upon but also description to ensure that these tools are used within the appropriate context. For example, the Pangolin tool is accompanied by extensive documentation of use cases and dependencies as well as implementation instructions for users and information about the software's outputs and interpretation of those outputs. Both description of the software and underlying decisions behind appraisal were relatively visible—for example, the Pangolin documentation references certain modeling choices as well as information about when various models are appropriate, and the metrics they used to assess the models (accuracy and training time). However, software products require maintenance and updating, which itself relies on curatorial labor. The Pangolin team, for example, relies on a "a team of experts and volunteers from around the world who will work to maintain these lineage designations alongside crowdsourced input through GitHub requests" (O'Toole et al., 2021, p. 8).

4.2 | Transparency, trust, and credit in and for digital curation practices

Overall, curatorial work was present in OS projects in multiple forms, including appraisal and selection of existing content, organizing content into relevant categories and structures, creating and maintaining structured metadata and metadata schemas, cleaning and refining datasets in preparation for analysis. However, such curatorial activities were for the most part invisible as they were rarely mentioned in projects' mission statements, or explicated elsewhere. When curation was mentioned, we found an emphasis on appraisal over other kinds of curation. In most projects, curatorial activities were framed as challenges. Curatorial work was made visible almost exclusively when a project's goals explicitly included curation—a term used most commonly to mean selection, appraisal and categorization—for example in Elsevier's "Novel Coronavirus Information Center" project,[22] whose mission was to publish curated collections of COVID-19 publications.

Even when curatorial work was mentioned, there was little information about the rationales behind curatorial decisions, such as why certain content was included and certain content was not, or how organizing categories came to be. Lack of transparency over curation decisions was prevalent especially among industry-led projects such as in Google COVID-19 Open Data Repository,[23] Google COVID-19 Community Mobility Reports,[24] and Bing COVID-19 Tracker.[25] Nevertheless, we found some notable exceptions. The creators of the Atlantic's Covid Tracking Project[26] provided detailed information about how data were selected and assessed, what phenomena the collected data did and did not represent, how the released data related to other datasets existing elsewhere, and on how decisions about not automating the collection of certain datasets were made. Nextstrain—an open source platform born out of a collaboration within academia that visualizes "real-time snapshots of evolving pathogen populations"—also documented and explained rationales behind several curatorial choices. For example, in a dedicated page titled "Glossary," the initiative listed and defined what key terminology central to the initiative meant in the specific context of Nextstrain. The team created similar pages for data formats and data files, and also shared detailed guidelines for software and platform users on how to make sense of visualized data and on how to contribute their own data to Nextstrain. In addition to this, they provided contact information for the teams that were directly responsible for curating and maintaining specific datasets. UShER[27]—a Genome Browser project that visually depicts the evolutionary relationships among COVID-19 genome sequences—engaged in a similar transparency effort, as did the team at the US federal government's Centers for Medicare and Medicaid Services that developed the COVID-19 Nursing Home Data[28] data repository. Others used formal publication as a medium to outline curatorial choices and challenges, as in the case of a Johns Hopkins team that published an article on the lessons learned from their data collection and visualization initiative.

4.3 | Labor and credit in and for digital curation practices

However, even among projects that were transparent about their rationales beyond curatorial choices, scarce information was available about the individuals or teams responsible for taking and implementing curatorial decisions. Only 6 initiatives acknowledged individuals who performed curatorial work. For example, Google's Covid-19 Open Data Repository indicates that their data are "gathered automatically as well as from volunteers and contributors," but does not specifically credit any individuals involved in data curation activities. It seems worthy to point out that, while it is common for large companies and corporations not to acknowledge individual staff contributions in large-scale projects, it is recognized as good practice in citizen science to give credit to volunteer contributors. Similarly the CORD-19 dataset page includes in their acknowledgments the institutions that participated in the creation of the dataset, but not the individuals within those organizations doing curatorial work.

Overall, individual-level credit was rare for all labor categories analyzed, including leadership and data science (40 initiatives did not have any individual-level credit attribution for specific roles). When individuals were credited, for example, on a team page or in the acknowledgments of a publication, this credit was typically not tied to specific activities but to general work on the project. Some projects, like the above mentioned Covid Tracking Project and Nextstrain, however, listed and credited each individual who participated in the initiative on their team page. Yet, roles were not singled out and, for this reason, it remains hard to infer whether curators were included in the team page, and, if so, how prevalent curators were as opposed to other experts (e.g., data analysts, designers). A number of initiatives cited labor as an obstacle to maintaining their operations, and, in particular, database-provider EBSCO highlighted the need for librarians and information professionals to maintain EBSCO's COVID-19 Resources[29] content aggregator. In total 19 of the 50 projects we surveyed (38%) were no longer active as of summer 2023. Lack of stable forms of data curation work might have impacted the longevity of open science projects initiated during the pandemic.

4.4 | Projects popularity and degree of data reuse

Based on the data that we collected on media mentions and data citations, we also observed that the initiatives that more heavily invested in curatorial practices—and

made these explicit on their websites—were generally also the ones that were frequently covered in the news and whose content ended up being reused in many derivative projects (see, e.g., PANGOLIN; Nextstrain; and COVID-19 Dashboard by Johns Hopkins). Two notable exceptions emerged: (1) Google COVID-19 Community Mobility Reports received many media mentions even though curation was scarce and not visible, which can be however explained by the general popularity of the search engine and related company; and (2) The Atlantic Covid Tracking Project received relatively few media mentions, but this can be explained by the fact that the project started in early 2020 and ended in 2021, while we were only able to collect media mentions via Media Cloud data beginning in 2022. Also, because the Atlantic Covid Tracking Project does not have a DOI, we could not collect data on actual reuse of the initiative for research purposes. However, the Atlantic Covid Tracking Project's website notes that the initiative was cited in more than 1000 research articles and over 7700 news articles.

5 | DISCUSSION

During the pandemic, numerous Open Science (OS) initiatives emerged, mainly focused on sharing data and publications. These initiatives, such as dashboards, data repositories, and content aggregators, were led by universities, governments, and corporations, aiming to increase the speed of scientific progress and reach beyond traditional audiences. Digital curation played a critical role in OS by ensuring research outputs were preserved, organized, and reusable across various contexts. Curation involved practices like content selection, data cleaning, and metadata development. It ensured the proper interpretation and trustworthiness of research, enabling its reuse across different research cultures.

However, our work confirmed Shankar's et al. (2021) concerns about challenges like inconsistent curation practices, lack of transparency, and limited credit for curatorial work. In the discussion, we question to what extent digital curation is a recognized profession or a practice. While well-curated OS projects often gain visibility and trust, it is unclear if this success is due to the quality of curation or existing institutional support. Despite the vital role curatorial work plays in managing, organizing, and making research outputs reusable, digital curation is rarely credited at an individual or team level. The field lacks professional recognition, with digital curation often seen as a subset of data science rather than a distinct profession. This undervaluing of curation, both in visibility and funding, hampers its development and

recognition. Thus, echoing Shankar's et al., we also urge for greater professionalization of digital curation, including formal education programs, clear career paths, and sustainable funding. In addition, we suggest that digital curation is essential for knowledge production, as it ensures the legitimacy and reusability of scientific data. Highlighting the importance of making curatorial practices visible, we call for using the transparency of curatorial work as a criterion for evaluating OS projects, particularly as OS becomes more prominent in science communication. Proper recognition of digital curation is key to improving the quality and trustworthiness of OS initiatives.

5.1 | Digital curation as a practice or a profession?

Because our work is based on a content analysis on available online information, we cannot infer whether a causal relation exists between adopting best practices for digital curation and conducting successful, high visibility OS projects—as the data that we collected and analyzed suggests. Nor can we infer, if there is a causal relationship, its direction. For example, it might be the case that the practices of curating content and documenting curation increased a perceived sense of professionalism among the public towards these projects, and, as a consequence, these projects gained in trust and visibility. But it could also be that these projects were able to implement proper curatorial practices because they originated as new branches of existing, well-funded and resourceful infrastructures where curation was already valued, and, for this reason, they could rely on available labor, skills, and cultural sensibility to practice curation. In the latter scenario, the success of the projects would be primarily linked to the fact that these initiatives were already part of visible and well-estimated institutions, which also happened to value digital curation. In either scenario, these initiatives deeply valued curatorial practices and embraced them with pride. These projects promoted their digital curatorial efforts on the main pages of their websites and curation was presented with a similar degree of visibility used to present analytical results, if not higher. This finding speaks to previous research that showed how curating digital assets is not enough for an infrastructure to be trusted, such curatorial work has also to be made highly visible and be internally valued (Lin et al., 2020; Thomer et al., 2022; Yakel et al., 2013).

One aspect that characterizes all projects we looked at, including the most successful and properly curated projects, was that initiatives rarely provided team-level credit to those performing curatorial activities, and none singled out the individuals within these organizations responsible for digital curation work. Also, only 10 of the 50 projects we analyzed referred to digital curation practices as such. It should be noted that some of these initiatives involve government agencies and government data sources, in which individuals are probably inherently less likely to be credited on a website, which in turn helps explain why individual-level credit attributions for specific roles were scarce across all labor types. However, even small-scale initiatives did not acknowledge teams and individuals. These findings might be partially rooted in the fact that there is little talk of digital curation as a profession in itself outside of science policy and digital curation research settings. As we have seen, practices of digital curation—such as information appraisal, documentation, and cleaning—are widely adopted in OS initiatives, and their necessity for producing legitimate science is recognized. What seems to be missing is the awareness that these skills can be—and often are—learned, taught, and practiced as part of the *profession* of digital curation, in the same way that statistical analysis and data visualization are learned, taught, and practiced as skills of the profession of data science. Instead of digital curation professionals being systematically recognized (and hired) as such, the general trend seems to be that other disciplines—such as the fairness in machine learning community—develop and adopt digital curation practices without referring to these practices as "digital curation," nor linking and citing the long LIS tradition of digital curation (Bender & Friedman, 2018; Gebru et al., 2021; Mitchell et al., 2019; Paullada et al., 2021; Peng et al., 2021).

Over the past 10 years or so, digital curation researchers and stakeholders have argued for the need of increasing professionalization of the field (Cushing & Shankar, 2019; Kim et al., 2013; National Research Council, 2015). Meanwhile, digital curation continues to be practiced, but it has not yet been fully professionalized (Kouper, 2016). For example, the U.S. Bureau of Labor Statistics does not currently recognize digital curation in the Standard Occupational Classification System, posing a challenge for understanding where digital curation is in demand and how the digital curation curation workforce should be trained (National Research Council, 2015). iSchools can lead the way towards standardizing digital curation as a profession, first and foremost by engaging in extensive networking efforts aimed at creating clear employment paths for their graduates, and by reaching outside the boundaries of iSchools to offer certificates and internships to prospective domain experts. Adequate, sustainable funding for digital curators are also needed. Along with a request for a data plan, researchers should also be allowed to request funding for digital curation.

This should be seen as necessary as receiving funding for OA publishing. Recently, some progress has been made on this issue. In her role as the director of the White House Office of Science and Technology Policy (OSTP), Alondra Nelson, addressed publicly the importance of making science outputs reusable via ad hoc curatorial practices, and not merely freely available (Anderson & Wulf, 2022; Gill & Nelson, 2022; Nelson et al., 2022). In the 2022 memorandum on *Ensuring Free, Immediate, and Equitable Access to Federally Funded Research*, OSTP also committed to allowing researchers to include costs associated with complying with public access policies into their open access and data management proposals (Nelson, 2022).

5.2 | Show me the context: Digital curation as knowledge production

Another factor that might play a role in the lack of visibility of digital curation practices is that the importance of digital curation for knowledge production processes is often underestimated. Such underestimation can be found, for example, in many definitions of digital curation that characterize digital curation as "adding value" to existing digital content, and especially to research data (National Research Council, 2015; Poole, 2016). These definitions rely on an economic conceptualization of research data as assets, and successfully and rightfully make the case that digital curation increases the economic value of data as it turns data into reusable, fungible objects (Leonelli, 2019a). There is no doubt that digital curation is good for competitiveness, innovation, and scientific advancement (Poole, 2017). However, a fairer definition of digital curation would recognize not only its added value to digital assets, but also the fact that, without digital curation, the very legitimacy of scientific findings is at risk, especially in an open scientific setting. When undergoing public scrutiny, research findings are recognized as pieces of evidence depending on the extent to which the underlying data and processes are properly curated. There is no legitimacy of scientific research, if data (and related software) cannot be properly interpreted, if data cease to exist due to lack of preservation, or if data exist somewhere but cannot be found. In other words, digital curation processes enable open science to fulfill its foundational epistemological proposition (i.e., that data accessibility and reusability lead to greater scientific legitimacy).

Leonelli goes as far as proposing that "for any object to be identified and recognized as datum, it needs to be portable" (Leonelli, 2020, p. 6). According to Leonelli, every piece of data, even the most simple one, needs to move from one epistemic context to another epistemic context in order to function as datum, even within the same laboratory (from an instrument reading into a database, for example). Thus, while certain aspects of digital curation might not be always essential to the actual practice of research (e.g., long-term preservation of data and software), there are other aspects of digital curation that seem to be foundational to the very production of knowledge (e.g., data cleaning and organization).

Definitions of digital curation should do justice to digital curation's essential role in the knowledge production and validation process, as much as data collection and analysis are. This shift in the narrative about what constitutes digital curation should be paired with a shift in the narrative about what matters most when it comes to legitimizing knowledge, from "show me the data" to "show me the data and the context." In a global digital world in which anyone can claim to be producing legitimate knowledge and being an expert, being able to explicate, preserve, and most importantly communicate settings of knowledge production will increasingly be a requirement, or even a synonym, for trust and quality.

5.3 | Teaching digital curation as a profession in itself

Recognizing the value of digital curation as a practice of knowledge production might also require drawing clear boundaries between practices of digital curation and practices of data science. In a survey of 65 iSchool curricula, Ortiz-Repiso et al. (2018) note that few programs provided focused curricula in the area of digital curation, particularly compared to data analytics. While some iSchools (e.g., the University of North Carolina) offer specialized master's degrees in digital curation, digital curation is often presented in higher education curricula in LIS and data science as a subordinate component of data science. For example, the University of Illinois's bachelors program in Information Science and Data Science requires a single course on Data Management, Curation, and Reproducibility as part of their data science core courses. From an educational point of view, promoting and teaching digital curation as a component of data science has its benefits. By being exposed to digital curation, prospective data scientists develop an awareness and sensitivity towards the role that context, information quality, and information sustainability play in knowledge production. Similarly, by being exposed to data science, prospective digital curators learn necessary technical and methodological skills that are increasingly essential for digital curation practices. Also, because data science has in general more appeal than digital curation, by

associating digital curation with data science, LIS researchers can attract funding opportunities that they would not otherwise have access to.

However, this codependent—and to some extent servant—relationship that digital curation holds in relation to data science might present its overlooked costs. First, by leaving the spotlight to analytical efforts, digital curation practitioners lose in visibility, influence, and political capacity, which they would need to argue for the centrality of digital curation practices in enabling knowledge production practices, and the resources needed to make this happen. Second, it confuses institutions and funders and, especially, employers on what set of skills, exactly, are needed to practice digital curation, and about what training is needed to teach such skills.

Instead, when digital curation is taught as a profession in itself, its value for knowledge production processes is at once clarified and made visible. For example, Murillo and Yoon (2021) developed a curriculum for teaching digital curators how to assist communities working with "community data," where community data is defined as "data that describe the local context and are used for community decision-making." As noted by the scholars, with increased data availability and utilization, community organizations face data curation challenges without a curation expert within their organizations. Specifically, Murillo and Yoon's work teach digital curators how to manage complex datasets generated by a variety of sources, which can include open government data and school data, but also data from private sectors and other local organizations working for community or social development. This line of work is particularly important in public health, where communities find themselves in need of support to keep data meaningful and accessible, manage and preserve data for long-term use, and appraise data for the fitness of use. Thus, Murillo and Yoon's curriculum identifies, teaches and explains the value of digital curation, while, at the same time, presenting and promoting digital curation as a profession in itself.

5.4 | Valuing digital curation as a core practice of OS

Finally, we propose that a way to encourage the recognition of digital curation as a knowledge production practice within OS initiatives could be to use the "degree of visibility of curatorial practices" as a criterion to evaluate OS projects. Desired and expected positive outcomes for OS projects typically include increased quality of research and increased efficiency of research (e.g., by reducing redundancies in research), accelerated progress and impacts (e.g., public health improvements), increased equity and diversity in research, and increased trust and accountability in the research process (Ali-Khan et al., 2018).

The degree of visibility and clarity about the rationales behind curatorial choices can then function as a proxy to evaluate the extent to which OS initiatives prioritize and promote trust and accountability in the research process. Incentivizing OS projects to invest in curation and to be transparent about their curatorial choices will be increasingly important because OS seems to be slowly moving towards occupying a more prominent and explicit function in the science communication ecosystem.

To some extent, OS has always been concerned with enabling better collaboration and cooperation among science stakeholders and their publics (Fecher & Friesike, 2014; MacGregor et al., 2014; Mirowski, 2018). However, before the pandemic, the focus on science communication and public engagement was neither explicit nor occupied a central role in OS, which was more traditionally focused on increasing access and reusability of research outputs (McKiernan et al., 2016; Nosek et al., 2015). Many of the OS projects that emerged during the COVID-19 pandemic intended to produce and disseminate content that could reach beyond traditional scientific audiences, from policymakers to the general public. Many projects also intentionally aimed to help scientists reach consensus over the legitimacy of ongoing research. Thus, enabling better communication processes—from experts to the public, but also within specialized audiences—emerged as a central concern for COVID-19 OS projects.

6 | CONCLUSIONS

The goal of this paper is to inform the information science community about the nature, relevance, and prevalence of digital curatorial activities in pandemic open science. Digital curation work was essential for enabling OS efforts during the COVID-19 pandemic. We found that, generally, the most popular and well-funded OS initiatives were also those that were most heavily invested in digital curation. Yet, even the projects that invested the most in curation and had the most resources, were not transparent about rationales used for curation and about who performed curatorial work. Within such initiatives, digital curation practices did not simply "added value" to digital assets, but made digital assets legitimate (i.e., publicly recognizable as valid and trustworthy) in the first place. Digital curation, in other words, was critical for OS initiatives fulfilling their very

mission of enabling greater transparency and participation in science. Thus, curation remains a largely invisible, uncredited form of "essential labor."

ORCID

Irene V. Pasquetto https://orcid.org/0000-0002-2790-0629

ENDNOTES

[1] https://nextstrain.org/.

[2] https://rrid.mitpress.mit.edu/.

[3] https://ourworldindata.org/explorers/coronavirus-data-explorer.

[4] https://hgis.uw.edu/virus/.

[5] https://covid19.healthdata.org/.

[6] https://covid19researchdatabase.org/.

[7] https://coronavirus.jhu.edu/map.html.

[8] https://gisaid.org/.

[9] https://pair-code.github.io/covid19_symptom_dataset/.

[10] https://www.rd-alliance.org/group/rda-covid-19-rda-covid19-omics-rda-covid-19-epidemiology-rda-covid19-clinical-rda-covid19-2.

[11] https://www.healthmap.org/covid-19/.

[12] https://icite.od.nih.gov/covid19/search/.

[13] https://www.ncbi.nlm.nih.gov/research/coronavirus/.

[14] https://www.kaggle.com/datasets/allen-institute-for-ai/CORD-19-research-challenge.

[15] https://www.openaire.eu/openaire-covid-19-gateway.

[16] https://covid19research.ssrc.org/wp-content/uploads/2020/10/Coronavirus-syllabus_20201019.pdf.

[17] https://github.com/natematias/covid-19-social-science-research.

[18] https://www.wikidata.org/wiki/Wikidata:WikiProject_COVID-19.

[19] https://covid19.who.int/.

[20] https://nssac.github.io/covid-19/dashboard/#faq.

[21] https://cov-lineages.org/resources/pangolin.html.

[22] https://www.elsevier.com/connect/coronavirus-information-center.

[23] https://health.google.com/covid-19/open-data/.

[24] https://www.google.com/covid19/mobility/.

[25] The Bing Covid-19 Tracker is no longer available online. An archived version is available at https://web.archive.org/web/20230407081115/https://www.bing.com/covid/local/unitedstates.

[26] https://covidtracking.com/.

[27] https://genome.ucsc.edu/covid19.html.

[28] https://data.cms.gov/covid-19/covid-19-nursing-home-data.

[29] https://covid-19.ebscomedical.com/.

REFERENCES

Alemneh, D. G., Hawamdeh, S., Chang, H. C., Rorissa, A., Assefa, S., & Helge, K. (2020). Open access in the age of a pandemic. *Proceedings of the Association for Information Science and Technology*, *57*(1), e295. https://doi.org/10.1002/pra2.295

Ali-Khan, S. E., Jean, A., MacDonald, E., & Gold, E. R. (2018). Defining success in open science [version 2; peer review: 2 approved]. *MNI Open Research*, *2*, 2. https://doi.org/10.12688/mniopenres.12780.2

Anderson, R., & Wulf, K. (2022). New light on the new OSTP memo: An interview with Dr. Alondra Nelson. *The Scholarly Kitchen*. Retrieved from https://scholarlykitchen.sspnet.org/2022/10/11/new-light-on-the-new-ostp-memo-an-interview-with-dr-alondra-nelson/

Barbazza, E., Ivanković, D., Wang, S., Gilmore, K. J., Poldrugovac, M., Willmington, C., Larrain, N., Bos, V., Allin, S., Klazinga, N., & Kringos, D. (2021). Exploring changes to the actionability of COVID-19 dashboards over the course of 2020 in the Canadian context: Descriptive assessment and expert appraisal study. *Journal of Medical Internet Research*, *23*(8), e30200.

Bender, E. M., & Friedman, B. (2018). Data statements for natural language processing: Toward mitigating system bias and enabling better science. *Transactions of the Association for Computational Linguistics*, *6*, 587–604. https://doi.org/10.1162/tacl_a_00041

Besançon, L., Peiffer-Smadja, N., Segalas, C., Jiang, H., Masuzzo, P., Smout, C., Billy, E., Deforet, M., & Leyrat, C. (2021). Open science saves lives: Lessons from the COVID-19 pandemic. *BMC Medical Research Methodology*, *21*(1), 1–18. https://doi.org/10.1186/s12874-021-01304-y

Borgman, C. L. (2015). *Big data, little data, No data: Scholarship in the networked world*. MIT Press.

Bos, V. L., Jansen, T., Klazinga, N. S., & Kringos, D. S. (2021). Development and actionability of the Dutch COVID-19 dashboard: Descriptive assessment and expert appraisal study. *JMIR Public Health and Surveillance*, *7*(10), e31161.

Cetina, K. K. (1999). *Epistemic cultures: How the sciences make knowledge*. Harvard University Press.

Ćurković, M., Košec, A., Bedeković, M. R., & Bedeković, V. (2021). Epistemic responsibilities in the COVID-19 pandemic: Is a digital infosphere a friend or a foe? *Journal of Biomedical Informatics*, *115*, 103709. https://doi.org/10.1016/j.jbi.2021.103709

Cushing, A. L., & Shankar, K. (2019). Digital curation on a small Island: A study of professional education and training needs in Ireland. *Archives and Records*, *40*(2), 146–163. https://doi.org/10.1080/23257962.2018.1425135

Darch, P. (2014). Managing the public to manage data: Citizen science and astronomy. *International Journal of Digital Curation*, *9*(1), 25–40. https://doi.org/10.2218/ijdc.v9i1.298

Delfanti, A. (2016). Beams of particles and papers: How digital preprint archives shape authorship and credit. *Social Studies of Science*, *46*(4), 629–645. https://doi.org/10.1177/0306312716659373

Digital Curation Centre. (n.d.). *What is digital curation?* DCC. Retrieved from https://www.dcc.ac.uk/about/digital-curation

Donovan, D. (2023). Johns Hopkins winds down pioneering pandemic data tracking. *The Hub*, October 30, 2023. Retrieved from https://hub.jhu.edu/2023/02/10/coronavirus-resource-center-ending-tracking/

Fecher, B., & Friesike, S. (2014). Open science: One term, five schools of thought. In *Opening science: The evolving guide on*

how the Internet is changing research, collaboration and scholarly publishing (pp. 17–47). Springer.

Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumé, H., III, & Crawford, K. (2021). Datasheets for datasets. *Communications of the ACM*, *64*(12), 86–92. https://doi.org/10.1145/3458723

Gill, S. S., & Nelson, A. (2022). *Research and policy opportunities for a shared agenda* [Recorded conversation]. Presented at Knight Foundation informed conference 2022: Conversations on democracy in the digital age.

Gitelman, L. (2013). *Raw data is an oxymoron*. MIT Press.

Hastings, J. (2021). Scientific ontologies, digital curation and the learning knowledge ecosystem. *Qurator*, *9*.

Hemphill, L., Pienta, A., Lafia, S., Akmon, D., & Bleckley, D. A. (2022). How do properties of data, their curation, and their funding relate to reuse? *Journal of the Association for Information Science and Technology*, *73*(10), 1432–1444. https://doi.org/10.1002/asi.24646

Homolak, J., Kodvanj, I., & Virag, D. (2020). Preliminary analysis of COVID-19 academic information patterns: A call for open science in the times of closed borders. *Scientometrics*, *124*, 2687–2701. https://doi.org/10.1007/s11192-020-03587-2

Ivanković, D., Barbazza, E., Bos, V., Brito Fernandes, Ó., Jamieson Gilmore, K., Jansen, T., Kara, P., Larrain, N., Lu, S., Meza-Torres, B., Mulyanto, J., Poldrugovac, M., Rotar, A., Wang, S., Willmington, C., Yang, Y., Yelgezekova, Z., Allin, S., Klazinga, N., & Kringos, D. (2021). Features constituting actionable COVID-19 dashboards: Descriptive assessment and expert appraisal of 158 public web-based COVID-19 dashboards. *Journal of Medical Internet Research*, *23*(2), e25682. https://doi.org/10.2196/25682

Khodaveisi, T., Dehdarirad, H., Bouraghi, H., Mohammadpour, A., Sajadi, F., & Hosseiniravandi, M. (2023). Characteristics and specifications of dashboards developed for the COVID-19 pandemic: A scoping review. *Journal of Public Health*, *32*, 1–22. https://doi.org/10.1007/s10389-023-01838-z

Kim, J., Warga, E., & Moen, W. E. (2013). Competencies required for digital curation: An analysis of job advertisements. *International Journal of Digital Curation*, *8*(1), 66–83. https://doi.org/10.2218/ijdc.v8i1.242

Kouper, I. (2016). Professional participation in digital curation. *Library & Information Science Research*, *38*(3), 212–223. https://doi.org/10.1016/j.lisr.2016.08.009

Kraemer, M. U., Scarpino, S. V., Marivate, V., Gutierrez, B., Xu, B., Lee, G., Hawkins, J. B., Rivers, C., Pigott, D. M., Katz, R., & Brownstein, J. S. (2021). Data curation during a pandemic and lessons learned from COVID-19. *Nature Computational Science*, *1*(1), 9–10. https://doi.org/10.1038/s43588-020-00015-6

Kwon, D. (2020). How swamped preprint servers are blocking bad coronavirus research. *Nature*, *581*(7807), 130–131. https://doi.org/10.1038/d41586-020-01394-6

Lee, D. J., & Stvilia, B. (2017). Practices of research data curation in institutional repositories: A qualitative view from repository staff. *PLoS One*, *12*(3), e0173987. https://doi.org/10.1371/journal.pone.0173987

Leonelli, S. (2019a). Data—From objects to assets. *Nature*, *574*(7778), 317–320. https://doi.org/10.1038/d41586-019-03062-w

Leonelli, S. (2019b). Data governance is key to interpretation: Reconceptualizing data in data science. *Harvard Data Science Review*, *1*(1). https://doi.org/10.1162/99608f92.17405bb6

Leonelli, S. (2020). Learning from data journeys. In S. Leonelli & N. Tempini (Eds.), *Data journeys in the sciences* (pp. 1–24). Springer. https://doi.org/10.1007/978-3-030-37177-7_1

Leonelli, S., Lovell, R., Wheeler, B. W., Fleming, L., & Williams, H. (2021). From FAIR data to fair data use: Methodological data fairness in health-related social media research. *Big Data & Society*, *8*(1). https://doi.org/10.1177/20539517211010310

Lin, D., Crabtree, J., Dillo, I., Downs, R. R., Edmunds, R., Giaretta, D., De Giusti, M., L'Hours, H., Hugo, W., Jenkyns, R., Khodiyar, V., Martone, M. E., Mokrane, M., Navale, V., Petters, J., Sierman, B., Sokolova, D. V., Stockhause, M., & Westbrook, J. (2020). The TRUST principles for digital repositories. *Scientific Data*, *7*(1), 144. https://doi.org/10.1038/s41597-020-0486-7

Liu, M., Bu, Y., Chen, C., Xu, J., Li, D., Leng, Y., Freeman, R. B., Meyer, E. T., Yoon, W., Sung, M., Jeong, M., Lee, J., Kang, J., Min, C., Song, M., Zhai, Y., & Ding, Y. (2022). Pandemics are catalysts of scientific novelty: Evidence from COVID-19. *Journal of the Association for Information Science and Technology*, *73*(8), 1065–1078. https://doi.org/10.1002/asi.24612

Loukissas, Y. A. (2019). *All data are local: Thinking critically in a data-driven society*. MIT Press.

MacGregor, J., Stranack, K., & Willinsky, J. (2014). The public knowledge project: Open source tools for open access to scholarly communication. In *Opening science: The evolving guide on how the internet is changing research, collaboration and scholarly publishing* (pp. 165–175). Springer.

Matonkar, P. V., & Dhuri, K. R. (2021). Open access and free resources on the Internet: Awareness and use during Covid-19 pandemic. *Library Philosophy and Practice*, *5153*, 1–15.

McKiernan, E. C., Bourne, P. E., Brown, C. T., Buck, S., Kenall, A., Lin, J., McDougall, D., Nosek, B. A., Ram, K., Soderberg, C. K., Spies, J. R., Thaney, K., Updegrove, A., Woo, K. H., & Yarkoni, T. (2016). How open science helps researchers succeed. *eLife*, *5*, e16800. https://doi.org/10.7554/eLife.16800

Mirowski, P. (2018). The future(s) of open science. *Social Studies of Science*, *48*(2), 171–203. https://doi.org/10.1177/0306312718772086

Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model cards for model reporting. In *Proceedings of the conference on fairness, accountability, and transparency* (pp. 220–229). ACM.

Murillo, A. P., & Yoon, A. (2021). A study of emerging trends in digital preservation literature: An analysis of journal articles presented in course syllabi. *Journal of Librarianship and Information Science*, *53*(4), 615–629.

National Research Council. (2015). *Preparing the workforce for digital curation*. The National Academies Press. https://doi.org/10.17226/18590

Nelson, A. (2022). Memorandum for the heads of executive departments and agencies: Ensuring free, immediate, and equitable access to federally funded research.

Nelson, A., Marcum, C., & Isler, J. (2022). Public access to advance equity. *Issues in Science and Technology*, *39*(1), 33–35.

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., Buck, S., Chambers, C. D., Chin, G., Christensen, G., Contestabile, M., Dafoe, A., Eich, E., Freese, J., Glennerster, R., Goroff, D., Green, D. P., Hesse, B., Humphreys, M., ... Yarkoni, T. (2015). Promoting an open research culture. *Science*, *348*(6242), 1422–1425. https://doi.org/10.1126/science.aab2374

Ortiz-Repiso, V., Greenberg, J., & Calzada-Prado, J. (2018). A cross-institutional analysis of data-related curricula in information science programmes: A focused look at the iSchools. *Journal of Information Science*, *44*(6), 768–784. https://doi.org/10.1177/0165551517748814

O'Toole, Á., Scher, E., Underwood, A., Jackson, B., Hill, V., McCrone, J. T., Colquhoun, R., Ruis, C., Abu-Dahab, K., Taylor, B., Yeats, C., du Plessis, L., Maloney, D., Medd, N., Attwood, S. W., Aanensen, D. M., Holmes, E. C., Pybus, O. G., & Rambaut, A. (2021). Assignment of epidemiological lineages in an emerging pandemic using the pangolin tool. *Virus Evolution*, *7*(2), veab064. https://doi.org/10.1093/ve/veab064

Pasquetto, I. V., Borgman, C. L., & Wofford, M. F. (2019). Uses and reuses of scientific data: The data creators' advantage. *Harvard Data Science Review*, *1*(2). https://doi.org/10.1162/99608f92.fc14bf2d

Pasquetto, I. V., Randles, B. M., & Borgman, C. L. (2017). On the reuse of scientific data. *Data Science Journal*, *16*, 8. https://doi.org/10.5334/dsj-2017-008

Paullada, A., Raji, I. D., Bender, E. M., Denton, E., & Hanna, A. (2021). Data and its (dis) contents: A survey of dataset development and use in machine learning research. *Patterns*, *2*(11), 100336. https://doi.org/10.1016/j.patter.2021.100336

Peng, K. L., Mathur, A., & Narayanan, A. (2021). *Mitigating dataset harms requires stewardship: Lessons from 1000 papers*. Paper presented at Thirty-fifth conference on neural information processing systems, datasets and benchmarks track.

Plantin, J. C. (2019). Data cleaners for pristine datasets: Visibility and invisibility of data processors in social science. *Science, Technology, & Human Values*, *44*(1), 52–73. https://doi.org/10.1177/0162243918781812

Poole, A. H. (2016). The conceptual landscape of digital curation. *Journal of Documentation*, *72*(5), 961–986. https://doi.org/10.1108/JD-10-2015-0123

Poole, A. H. (2017). "A greatly unexplored area": Digital curation and innovation in digital humanities. *Journal of the Association for Information Science and Technology*, *68*(7), 1772–1781. https://doi.org/10.1002/asi.23743

Shankar, K., Jeng, W., Thomer, A., Weber, N., & Yoon, A. (2021). Data curation as collective action during COVID-19. *Journal of the Association for Information Science and Technology*, *72*(3), 280–284. https://doi.org/10.1002/asi.24406

Thomer, A. K., Akmon, D., York, J. J., Tyler, A. R., Polasek, F., Lafia, S., Hemphill, L., & Yakel, E. (2022). The craft and coordination of data curation: Complicating workflow views of data science. *Proceedings of the ACM on Human-Computer Interaction*, *6*(CSCW2), 1–29.

Tse, E. G., Klug, D. M., & Todd, M. H. (2020). Open science approaches to COVID-19. *F1000Research*, *9*, 1043. https://doi.org/10.12688/f1000research.26084.1

Verovšek, P. J., & Gorišek, M. (2023). Experts in times of pandemic: Reframing the debate in the context of structural transformations of the public sphere. *Javnost—The Public*, *30*(4), 496–512. https://doi.org/10.1080/13183222.2023.2255821

Weisenberg, N. T. (2023). The right to science and the SDGs: What can we learn from COVID-19? *Frontiers in Sociology*, *8*, 1282721. https://doi.org/10.3389/fsoc.2023.1282721

Wikipedia. (n.d.). *Wikipedia: Reliable sources*. Retrieved from https://en.wikipedia.org/wiki/Wikipedia:Reliable_sources

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., ... Mons, B. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, *3*(1), 1–9. https://doi.org/10.1038/sdata.2016.18

Woelfle, M., Olliaro, P., & Todd, M. H. (2011). Open science is a research accelerator. *Nature Chemistry*, *3*(10), 745–748. https://doi.org/10.1038/nchem.1149

Yakel, E. (2007). Digital curation. *OCLC Systems & Services: International Digital Library Perspectives*, *23*(4), 335–340.

Yakel, E., Faniel, I. M., Kriesberg, A., & Yoon, A. (2013). Trust in digital repositories. *International Journal of Digital Curation*, *8*(1), 143–156. https://doi.org/10.2218/ijdc.v8i1.251

Yan, W. (2020). Coronavirus tests science's need for speed limits. *The New York Times*. Retrieved from https://www.nytimes.com/2020/04/14/science/coronavirus-disinformation.html