Received: 11 January 2021 | Revised: 12 August 2021 | Accepted: 12 August 2021

DOI: 10.1002/asi.24571

RESEARCH ARTICLE

JASIST WILEY

Open research data repositories: Practices, norms, and metadata for sharing images

Karin Hansson | **Anna Dahlgren**

Department of Culture and Aesthetics, Stockholm University, Stockholm, Sweden

Correspondence
Karin Hansson, Department of Culture and Aesthetics, Stockholm University, SE 106 91 Stockholm, Sweden.
Email: khansson@dsv.su.se

Funding information
Swedish Research Council, Grant/Award Number: 2018-06057

Abstract

Open research data repositories are promoted as one of the cornerstones in the open research paradigm, promoting collaboration, interoperability, and large-scale sharing and reuse. There is, however, a lack of research investigating what these sharing platforms actually share and a more critical interface analysis of the norms and practices embedded in this datafication of academic practice is needed. This article takes image data sharing in the humanities as a case study for investigating the possibilities and constraints in 5 open research data repositories. By analyzing the visual and textual content of the interface along with the technical means for metadata, the study shows how the platforms are differentiated in terms of signifiers of research paradigms, but that beneath the rhetoric of the interface, they are designed in a similar way, which does not correspond well with the image researchers' need for detailed metadata. Combined with the problem of copyright limitations, these data-sharing tools are simply not sophisticated enough when it comes to sharing and reusing images. The result also corresponds with previous research showing that these tools are used not so much for sharing research data, but more for promoting researcher personas.

1 | INTRODUCTION

Metadata and the archiving practices that produce it are becoming increasingly important as a means to navigate the rapidly growing volume of images online. Unlike textual sources, where, for example, OCR-reading, topic-modeling, and name-entity recognition have made searching and finding digitized text material much easier, visual sources in the broadest sense are still very much dependent on manually created textual metadata as they are not in themselves constituted by textual information.

In the context of academia, open data repositories where researchers can directly enrich images with metadata can potentially contribute to the searchability of images, situating them in a wider research context. Increasing datafication of research practices and their infrastructure also creates new incentives to enable several researchers to share metadata and data (Kidwell et al., 2016; Nosek et al., 2015; Roche et al., 2014). Idle resources such as researchers' image collections locked away on personal computers or offices, which have until now been difficult to commodify, will no longer be wasted (Open Research Data Task Force, 2018). This "open research" paradigm is expressed in different ways in research strategies, recommendations, and data management plans (Harrower et al., 2020; Poole & Garwood, 2020; Wuttke, 2019). Everyone seems to agree that:

> The days of keeping our research results to ourselves are over. (Carlos Moedas, Commissioner for Research, European Commission, 2016, p. 35)

Also, scholars in the humanities, whose areas of investigation are not usually categorized in terms of "data," are now expected to share their research material in ways that make it reusable and scalable (Earley-Spadoni, 2017; Given & Willson, 2018; Trace & Karadkar, 2017; Wuttke, 2019). Central for this open research paradigm are open research data sharing platforms, where the researcher not only publishes the results but, potentially, both the research processes and data collections (European Commission, 2016).

Digital data repositories have already had a major impact in scientific research areas that have traditionally been collaborative, and where data can be more easily quantified and reused (European Commission, 2016). Technology and digital practices also play a significant role in research conducted by scholars within the humanities, and the area of digital humanities promotes the extended use of quantitative research methods (Given & Willson, 2018). These sharing opportunities serve two purposes. First, they simplify the examination and review of the scholarly research material (transparency). Second, they enable the reuse of already collected material for new inquiries (efficiency). However, when it comes to sharing through open data sharing repositories, there is still much work to be done.

Latour (1987) points out that with the right metadata and ontologies, research data can function as "mobile" objects, and thus successfully move between different production contexts while still providing sufficient evidence. The reuse of data thus requires not only the data itself but advanced knowledge of how that data were produced. Research data are always produced for a specific purpose, like all data carriers of history and local cultural norms (Pasquetto et al., 2019). Consequently, tensions arise between the possibility of reusing data in new contexts and the risk of losing important information about the data's context when it is moved from one context to another (Borgman, 2015; Leonelli, 2015; Loukissas, 2019).

Therefore, data are rarely reused, and Pasquetto et al. (2019) show that reuse rarely occurs even in projects that collaborate and use the same methods. Reuse is avoided because it requires a great understanding of context and the transfer of information is very time-consuming. Not only does the researcher need the knowledge to interpret the data and understand what it contains, but also good metadata/paradata, knowledge of the models, theories, tools, techniques, and other contextual circumstances associated with its creation (Borgman et al., 2015). Thus, the actual transactional costs of reusing data often outweigh the benefits.

The standardization of methods and techniques in research is, according to Borgman et al. (2015), generally very poor, and it is often up to each individual researcher to manage the data. This means that data will tend to be managed locally and in ways that best suit the individual researcher. These problems are not specific to the humanities but are widespread around the reuse of research data where there is insufficient contextual information about how the data was generated. This means that even scientific data are rarely reused as intended, rather they are used primarily as a reference (Borgman et al., 2015).

It can be shown that much of the historically oriented humanities center around the reuse of data, which is why other research traditions may be interested in these practices. Unlike science where data are largely generated, many scholars within the humanities reuse sources that are collected and managed by others and categorized and curated as part of larger research infrastructure, typically museums, libraries, and public and private archives, not only intended for research.

But several obstacles prevent the sharing of researchers' data in these contexts. For example, most archives used by researchers in the cultural heritage sector, like archives and museum collections, do not allow researchers to share their data in the same archival context that their original material derives from. This reality is also why copyright issues often present problems for both accessing and sharing data. This is especially true with regard to images, as they cannot be cited partially like text. Art history scholars, for example, have been shown to keep large personal collections of digital images, including their own images and downloaded online materials, but there are obstacles to sharing these materials widely online owing to copyright issues (Long & Schonfeld, 2014).

Several studies have been carried out to understand these obstacles and promote more widespread use of digital tools for open research in the humanities. These studies look, foremost, at the researchers' attitudes to open research. A survey in the UK by Van den Eynden et al. (2016), which included 583 senior researchers, identified a problem around researchers' attitudes to open research. It showed that researchers in the humanities are open to sharing but have very little experience of digital data sharing, and also lack motivation. The humanities researchers in the survey also had problems with the normative definition of "data," not seeing that their research generates or relies on data, but rather on materials, narratives, and observations. As pointed out by Lisa Gitelman (2013), the very concept of data is immersed in notions of objectivity, which interacts poorly with the humanities' focus of interpretation and critical analysis.

More practical problems were identified in a design study to support the sharing of data in the humanities. Here Trace and Karadkar (2017) list some commonalities

in the research process which they identify as obstacles to fitting into available open research tools: humanities scholarships are highly individualistic in nature; the research is characterized by a process in which researchers draw from theory, source, and text while introducing new perspectives and reflections; the researcher is also part of the method where interpretation and reflection are at the core (Long & Schonfeld, 2014).

In conclusion, previous research shows that qualified metadata is central to the reuse and sharing of data and explains why data are rarely reused. This is especially true of the humanities—an area where there is room for improvement when it comes to more systematic data sharing and reuse. On the one hand, researchers' knowledge or attitude to datafication is seen as part of the problem, indicating that this group needs to be addressed differently from researchers in other disciplines. On the other hand, legal and technical obstacles are shown to hinder the increased sharing and reuse of image research data. In addition, there is a lack of detailed knowledge about what research data repositories entail, especially when it comes to sharing images as data.

In this study, which is part of a larger research project on the development of digital sharing practices in the visual cultural heritage, we focus on the data sharing opportunities for researchers in the humanities, having images as a central part of their research material. From this perspective, we compare the possibilities and constraints in some larger universal open research data repositories that commonly support the sharing of image data.

The article is organized as follows: The next section provides an overview of recent research on open research data repositories and academic social networking sites (ASNS) and specifies our research questions. The third section describes our data and methods. The fourth section presents the results, and in the final section, we discuss the results in detail and suggest some recommendations for the development of image data repositories for the humanities.

2 | OPEN RESEARCH DATA REPOSITORIES

The long-term preservation and availability of research in different forms has historically been a major concern for the research community. Digitization and digitalization can strengthen traditional scholarly communication, but might also enable a reconfiguring of the norms, technologies, and institutions (Plantin et al., 2018). This shows why digitized data repositories are interesting to study with regard to both the expressed norms and the underlying systems.

There is an abundance of initiatives and services offering mechanisms for the identification and archiving of data, as well as for sharing them and making them citable. In August 2020, the Registry of Research Data Repositories (re3data.org) listed 2,554 repositories, of which 276 were categorized as fitting for the humanities (re3data, n.d.). Most of them are local or subject-specific, covering a particular area of research or data, such as "Images of Medieval Art and Architecture," serving a global community of medieval scholars (ref). But an increasing number of data repositories are "generalist" and open for a wide range of researchers. There is also a demand among researchers for more universal repositories and a more systematic sharing of all aspects of research (Assante et al., 2016). Infrastructures for sharing data are expensive to construct and maintain, and most research projects are small and time-limited, which does not allow for the long-term preservation of shared infrastructure (Borgman et al., 2015). Research projects within the humanities are especially limited in this way and tend to be carried out by individuals or small teams of researchers. The resulting scientific papers are usually single-authored or involve, at most, a handful of co-authors. The standardization of research material, methods, and language (concepts) is minimal and responsibility for data management falls to the researcher. Therefore, more universal data repositories are important for the long-term preservation and reuse of data in these settings.

In this overview, we have looked especially at comparisons of these more universal repositories—what we call open research data repository (ORDR). In addition, we scrutinize research on ASNS, as these sites tend to overlap and increasingly include data repositories among their services, whereas the data repositories develop more social network functionality and related metrics.

ORDR and ASNS have been thoroughly researched from different perspectives. Assante et al. (2016) have analyzed six research data repositories that focus on available metadata at the dataset level covering aspects of publishing. Other research topics include usability evaluations that compare the availably of functions (Bhardwaj, 2017; Vasquez & Caicedo Bastidas, 2015), and quality control in self-publishing (Sicilia et al., 2017). A well-researched theme includes the users' motivation and incentives. A large-scale international survey by Figshare stated that the reason for publishing research data is, foremost, the potential for increased impact and visibility of the research (24%); second, its public benefit (20%). Other motivations include: getting credit (9%), funder requirement (8%), publisher requirement (8%), and trust in the person requesting the data (7%) (Hahnel et al., 2017). Similarly, despite their framing as tools for networking and sharing, findings indicate that ASNS are used mostly for information consumption, and less so for sharing, and just occasionally for peer interaction (Meishar-Tal & Pieterse, 2017). One important use is for promoting an online persona (Barbour & Marshall, 2012). Research also indicates that there appear differences between research

disciplines in their platform preferences (Ortega, 2015) and their preferred data format (Thelwall & Kousha, 2016).

Another topic of research examines the different metric systems that not only measure publications and citations but sometimes also user activity within the tool (Ovadia, 2014). Reading activity monitored in platforms such as Mendeley can, for example, provide good early impact evidence as this metrics appear before citations and correlate with citations in the long term (Thelwall, 2019). Sharing research data also becomes part of this metric when data, such as publications, are given unique DOIs and can be cited. Analyses of research data in the Data Citation Index (DCI) have shown that while most research data remain uncited, since 2008 there has been a growing trend in citing datasets (Kraker et al., 2015).

Critical voices suggest that these additional metrics can be seen as part of the increasing instrumentalization of the research process where the researcher's status is translated into a more formal and quantitative economic system (Busch, 2017; Hall, 2016). Today digital intermediaries have also become important as a virtual infrastructure for independent researchers—what Hall (2016) calls the "Uberification of the University"—as less secure working conditions with flexible, temporary, project-based jobs create a gig economy in academia, enabled by digital intermediaries that act as algorithmic gatekeepers in control of the metric system.

However, there is little research investigating what these platforms for data sharing actually share. Also, we have not found studies that analyze the available metadata for the data in any detail, even though a lack of relevant metadata is frequently cited as the reason why open data are not reused (Pasquetto et al., 2019). There is a need for more critical interface studies of norms and practices embedded in these systems, in this instrumentalization and upscaling of research practice. In response to this research need, this article compares research practices in the humanities focusing on sharing and reusing images, with the functionality of some available interfaces.

The following research questions drive this study:

- What are the research norms and practices expressed in the datafication and the interface esthetics?
- What types of research outputs/materials are categorized and measured in the interfaces, and which are not?

3 | DATA AND METHODS

In this study, we analyze five data sharing interfaces to see how the data need described above of humanities scholars can be met. A useful concept for describing our focus in the analysis is what Gaver calls *technology affordances*, meaning that to be affordable, the properties of the artifact the user acts upon must be available in a form that makes sense and is relevant to the user (Gaver, 1991). This way of looking at technology considers the combination of the user's capabilities and the possibilities of the artifact.

We are also influenced by Hutchby's (2001) affordance theory framework, recognizing both the constraining and enabling materiality of artifacts. This framework allows us to look at technical affordance broadly, in between constructivism and determinism. As pointed out by David Lakoff and Mark Johnson in the, by now, classic *Metaphors We Live By*, the words and concepts we use shape not only our thinking but also our perception of ourselves and others, not least how we act (Lakoff, 1980). That is, we assume that language and technology are important both on a macro level, as a carrier of ideologies and norms that construct meaning, affecting how we perceive affordance, and on a micro-level, as something that determines action. In this interface content analysis, we therefore focus on both the norms expressed in these interfaces and the type of practices that are encouraged, as well as on the "datafication" that constrains possible action—that is, how researchers' potential activities and materials can actually be transformed into quantifiable units through the interface.

Our initial pilot study of the online interfaces pointed to a gap between the more individualistic research approaches in the humanities and the ideas of the research process expressed in the interfaces. Therefore, we looked especially at how the researcher's identity is framed in the interface through textual and esthetic discourses, what can be called the *signifiers of research paradigms*. The pilot study also showed that in cases where copyright is not a problem, the sharing of the visual data is often prevented by the interface, owing to the lack of basic functionality, such as the ability to create metadata for the images. At best, there are opportunities to enrich image files with a comment, but not with more structured data of the type used by humanities-oriented image researchers. In this study, we therefore examine how interfaces supporting the sharing of visual research data match the information needed (the need for different types of metadata), by researchers in the humanities as referred to above: what we call the *technical means of metadata production* available to the user.

The interface study was conducted from March to August 2020 and provides an in-depth analysis of five open data repositories in the English-speaking sphere, designed for "the long tail" of individual or smaller groups of researchers to share results across a diversity of research fields of science: ResearchGate Data, Humanities Commons CORE deposit, Harvard Dataverse, Figshare, and Zenodo (Table 1).

TABLE 1 List of open data repositories; name and URL, about, starting year, organization, country, and metadata standard

Name (URL)	About	Year	Organization	Country	Metadata
ResearchGate Data (https://www.researchgate.net/about)	ResearchGate Data is an integrated part of ResearchGate, one of the largest ASNS with over 20 million users according to their website in April 2021. The purpose of ResearchGate is to provide networking opportunities and tools to share publications and data with a "professional network for scientists and researchers." It is guided by a mission "to make research open to all." The interface encourages users to follow the activities of other users and engage in discussions with them	2008	Commercial, ResearchGate GmbH	Germany	DataCite Metadata Schema, Dublin Core
Commons Open Repository Exchange (CORE) (https://hcommons.org/core/)	CORE is part of Humanities Commons that was launched in 2016, to support a network for people researching and teaching in the humanities. In addition to the repository, the interface provides web publication opportunities, and teaching materials and tools. According to the website, in April 2021 there were 26,678 registered members, and almost 12,000 "Core deposits," comprising datasets and image files but also including articles and books	2016	Non-profit, the Modern Language Association	United States	–
Harvard Dataverse (https://dataverse.harvard.edu/dataverse/harvard)	Harvard Dataverse is an open-source web application being developed at Harvard's Institute for Quantitative Social Science (IQSS), together with a global network of collaborators and partners (Dataverse, n.d.). The aim is to share and preserve data from any research discipline and to explore and analyze research data. In April 2021, the website reported that there were 108,436 datasets and 993,701 files in the system	2006	Non-profit, Harvard University, Institute for Quantitative Social Sciences	United States	DataCite Metadata Schema, Dublin Core
Figshare (https://figshare.com)	Figshare's mission is to store, manage, and freely disseminate any kind of research output. According to the website, Figshare provides research data infrastructure for Wiley, SpringerNature, Taylor & Francis, PLOS, among others, and collaborates with a large network of academic institutions to enable open research. In April 2021, Figshare counted 1,364,584 datasets and 1,876,879 figures in their system	2011	Commercial. Funded by Digital Science (part of the publishing company Holtzbrinck)	United Kingdom, United States, and Romania	DataCite Metadata Schema, Dublin Core
Zenodo (https://zenodo.org)	Zenodo is an online repository for sharing publications and data. Launched in May 2013, as part of a European-wide research	2013	Non-profit, CERN, OpenAIRE, European Organization for	Switzerland	DataCite Metadata Schema, Dublin Core

(Continues)

TABLE 1 (Continued)

Name (URL)	About	Year	Organization	Country	Metadata
	infrastructure, the Zenodo repository was specifically designed to enable a universal data repository for individual researchers providing data sharing in a wide variety of formats across all fields of research (Sicilia et al., 2017; Wuttke, 2019). In April 2021, Zenodo counted 77,453 datasets in their system and 596,877 image files		Nuclear Research, European Commission Horizon 2020, and FP7 program		

Note: The information is compiled from depositories' websites, conversations with administrators, www.re3data.org, and named sources.

To enable a deeper understanding of these tools, we created active user identities where we explored the interactive as well as the social dimensions of the interfaces by engaging with them in various ways over 6 months, creating profiles, uploading data and publications, participating in groups, and so on. These interactions were observed and documented in screenshots and notes. The interdisciplinary nature of the research group, which consisted of researchers from art history and computer science, provided a useful setting for comparing different research practices and norms.

To discern patterns in the type of information deposited in these online platforms, we used three categories for types of data/metadata: format, content, and publishing. The format includes all details that describe the material features of the data such as file type, language, and version. Content includes all information concerning what the data are about, such as what or who is represented/depicted, where it originated, as well as its original creator and production/publication context. This is described through content descriptions in free text or keywords but also in the title or file name as these often designate the content of a dataset or file.

Finally, publishing includes information connected to the production of the dataset in the online repository. This includes the names of the scholar(s) who order or manage the images as well as information on the type and format of publication where they have been written about the deposit date and access rights.

3.1 | An introduction to the open data repositories in the study

The open data repositories selected for the study are open and free to use for individual researchers, directed to the humanities (among other areas), enabling long-term data sharing. The interface allows images as data and represents some of the more popular data deposits in the English-speaking world according to our overview of previous research and the deposits' own estimates of their number of users and published datasets. Mendeley Data and Dryad were, for example, excluded as they do not have images as a data category. We included Humanities Commons CORE for comparison, as it is designed with the humanities in mind, unlike the others that are directed to a broader group of researchers.

To identify the repositories, we used the following filters when searching re3data.org: providing a DOI, supporting images as content type, for the humanities. Then we checked that these were also: open access, for open upload, free for individual researchers, international, for general use, and online (having no software constraints).

4 | RESULTS

4.1 | Visual and textual signifiers of research paradigms

Norms about different research practices are expressed in different ways in the interfaces: first, directly as mission statements in self-presentation, where the aims of the repositories on their "About" pages promote transparency, sharing, and open research. ResearchGate declares for example that: "We're guided by our mission to connect the world of science and make research open to all" (ResearchGate, n.d.). Similarly, Humanities Commons is "focused on providing a space to discuss, share, and store cutting-edge research and innovative pedagogy" (Humanities Commons, n.d.). The Zenodo "About" page declares that they are "Passionate about Open Science!" Thus, core values in research are about openness and sharing, promising a platform that expands and opens up the research process, connecting resources and people.

However, other aspects of academia, such as being a place for criticality and reflection, are not touched upon at all.

Research in the humanities is most often assessed in terms of quality rather than quantity, which is why its value is typically perceived as difficult to translate into simple metrics, even though such attempts are being made (Ochsner et al., 2016). Therefore, we assume a blunt focus on metrics signifies a research paradigm that is remote from the humanities. Along this line, Harvard Dataverse provides a pragmatic presentation of their *raison d'être*, declaring that the repository: "Increases the visibility of your research. Gets you credit for your data through citations. Satisfies data sharing requirements from funders and journals." Also, Figshare chooses to provide a more bibliometric argument for using their repository, to "get more citations for all of the outputs of your academic research," and their home page tagline is "credit for all your research" (Figshare, n.d.). Thus, the mission statements in these interfaces promote, on one hand, an ideal about research data as something that should be open, shared, and reused, and, on the other hand, the reality of the research world where research metrics function as a currency where all types of research output can be scored and become gamified.

Other norms about research are expressed in the words and concepts used to describe the content or features of the sites. ResearchGate describes the research environment as a "lab" (Figure 1), a concept that signals some form of chemical or technical invention rather than a critical analysis of cultural artifacts. Similarly, on Harvard Dataverse, datasets and data are organized in "Dataverses" that can be shared repositories for a group of researchers. Figshare provides this functionality in what is called "projects."

In comparison, in Zenodo the data repositories do not have multiple users, but it is suggested that you can share the data with a "community," either by applying to join one or by creating one yourself, thus emphasizing the performativity of research worlds. Humanities Commons CORE deposits are shareable in a similar way, where "deposits" (datasets) can be shared by up to five groups that the user can be a member of. The restriction in numbers indicates a norm where fewer qualitative relations are preferred rather than the ability to broadcast the research in a large network.

So, in ResearchGate, Dataverse, and Figshare collaborative research is directly supported by providing spaces for data sharing, indicating a type of research where groups of researchers are directly involved in the research process, and publications have lists of authors. In Zenodo and Humanities Commons, the research norm is not primarily about collaboration within research projects but more about sharing results within a selected network. While there are differences in how the research process is supported in terms of collaboration, all the interfaces are clearly positioned within an academic setting where the focus is on publication.

In the interface for datasets and data, publishing metadata such as the publication context is central, rather than metadata describing the content of the dataset/data. When metadata for describing images is available, it is relevant only in a publication setting but not as research material. The categories that exist, such as plot, drawing, photo, or diagram, only disclose what kind of visual format or medium the image has, whether it is iconographic (photo, drawing) or a schematic depiction of something (plot, diagram) but not what it is about (Figure 2).

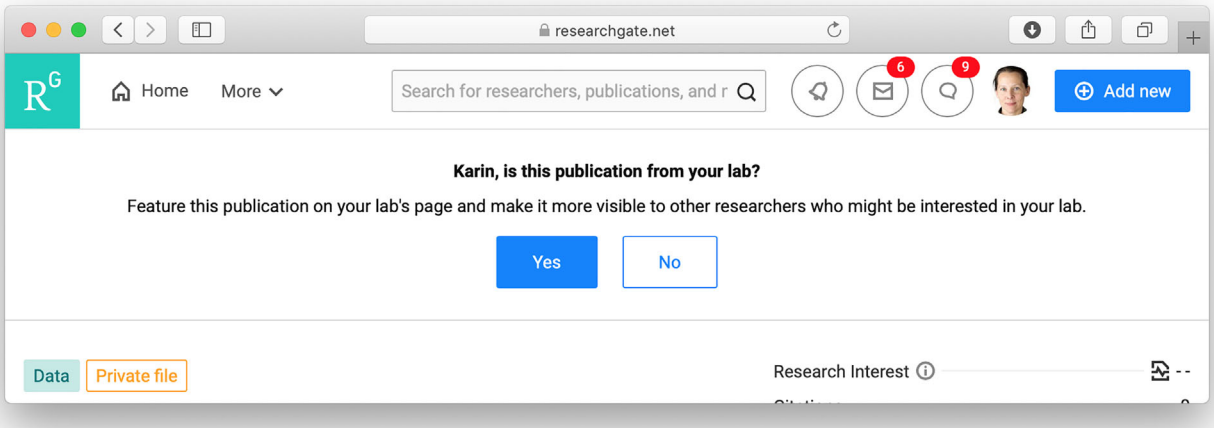

FIGURE 1 ResearchGate interface automatic suggestion when publishing data [Color figure can be viewed at wileyonlinelibrary.com]

All five interfaces collect and display available metrics about the user, such as the number of published items, but the metrics are only displayed for other users in the interfaces that have a social network functionality, like Figshare, ResearchGate, and the Humanities Commons CORE. These three interfaces allow the user to personalize their profiles with self-presentations in text and sometimes through images. But even though these three interfaces all contain similar user data, it is not so much the functionality, but whether and how the metrics are displayed that positions the interface within different research spheres. ResearchGate, for example, provides visualizations of the user metrics, where the research output, in terms of numbers, is emphasized through diagrams (Figure 3). Humanities Commons, on the other hand, provides a profile page that can be personalized esthetically where the researcher's institutional context and background are highlighted, while publication data are hidden at the end of the page (Figure 4). Harvard Dataverse does not display users' profile pages publicly for other users, but organizations can create a customized page for groups of researchers, departments, and faculty to share data, thus emphasizing the collaborative process at the institutional level rather than from the individual user's perspective. This is also possible in ResearchGate.

When looking at the available metadata that describe images, it is apparent that it is very limited and clearly linked to publishing in a research context. In Humanities Commons CORE, images are called "visual records," while on other sites all research material is called "data," This clearly frames the site within an interpretative humanities research paradigm that opposes the idea of data as an objective fact, and instead as something created. Thus the very use or non-use of the word "data" in these sites indicates a particular view of the object of study in research.

Humanities Commons is an initiative to create a specific place for researchers within the humanities to share their research. Here the way the research and researcher are addressed clearly indicates the repository's positioning with regard to the humanities, but also that it does not separate research data (the researched material) from scholarly output. The data repository does not differentiate between secondary research data, such as journal articles and other scholarly output in words and images, and primary research data, such as the texts, images, or films on which the secondary research data are based. This indicates an idea about the process involved in humanities' research being something that cannot be separated in terms of data and result. They have also chosen to call their datasets "Deposits" to indicate an investment in a shared asset like a currency that enables a value exchange, rather than something that is being stored. The metadata for the type of "core" that can be inserted here is significantly more diverse than in the other interfaces, from poetry to performance, to images, to books—a total of about 40 different categories including datasets. However, the core/data are treated as one file, so for uploading multiple images they need to be archived one by one in their own deposit, or in one compressed file.

Although there is great diversity in Humanities Commons CORE in terms of types of data, the forms of publishing that are described are clearly academic here as well; the focus, in the end, is not on describing the content of the data or core, but on placing it within an academic publishing context. This applies to all five interfaces in this study. Most significantly, the individual datafiles often cannot be defined separately, other than in a very simplistic way such as a file name. The structure of the various systems assumes that the actual "data" in the project is similar and can be described on a general level. The interfaces are thus differentiated for diverse

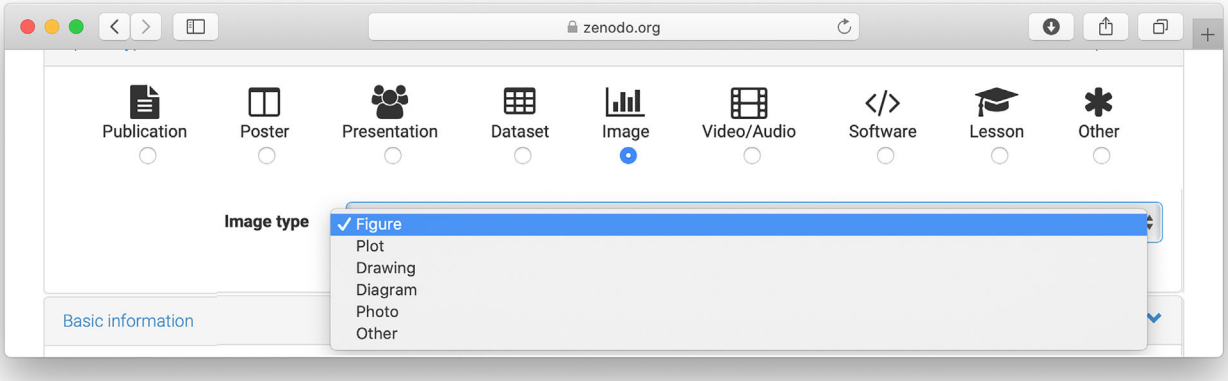

FIGURE 2 Zenodo's definitions of publication types and image types [Color figure can be viewed at wileyonlinelibrary.com]

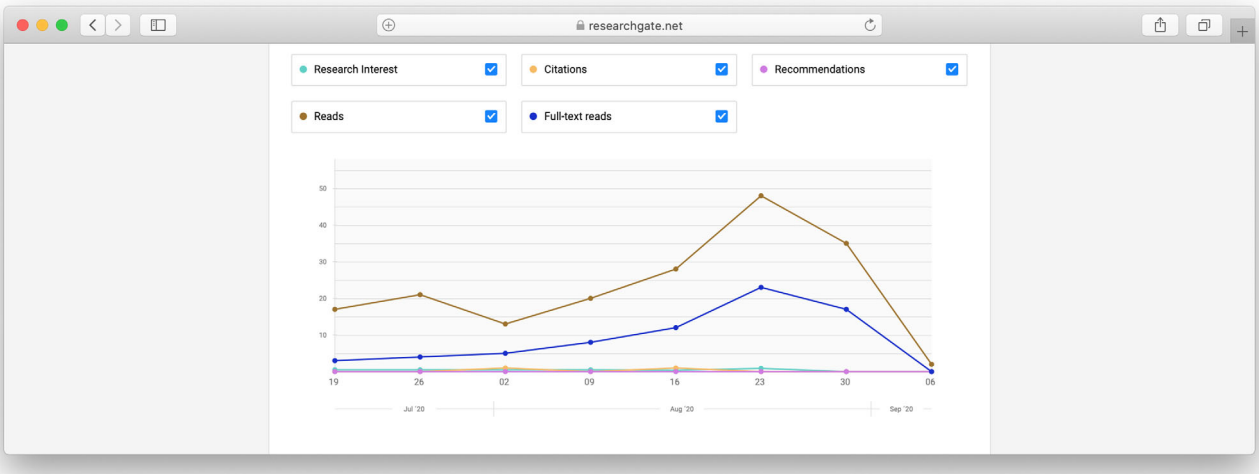

FIGURE 3 ResearchGate's visualization of user metrics [Color figure can be viewed at wileyonlinelibrary.com]

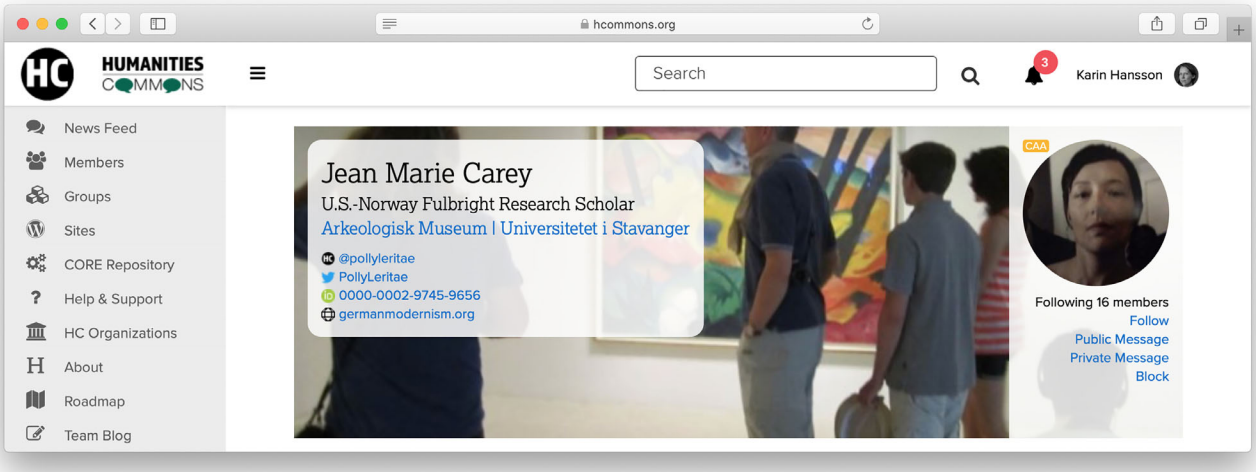

FIGURE 4 Humanities Commons user profile's top of page (published with permission from Jean Marie Carey) [Color figure can be viewed at wileyonlinelibrary.com]

user groups in terms of signifiers of research paradigms, from more collaborative research practices to those that are more individualistic, and with different epistemologies. But under the esthetics and rhetoric of the interface surface, they are designed in a similar way.

4.2 | Technical means of metadata production

When categorizing the metadata that is available in the interfaces into three main types, it is obvious that identifying the copyright owner and pointing out the technical, legal, managing aspects of the data related to this is the primary concern, rather than the content or the format of the images themselves (See Figure 5).

The metadata contained in the interfaces is primarily of the management metadata type, concerned with the location of the data within an archive system: whether the scholarly work, where it has been written up, is a book or journal, information about license, access rights, version, language, file type, DOI, and so on. The metadata related to the broader cultural context of historical images is minor in the types of metadata related to the image content. First, this category of metadata includes very few parameters (title, keywords, date) and second,

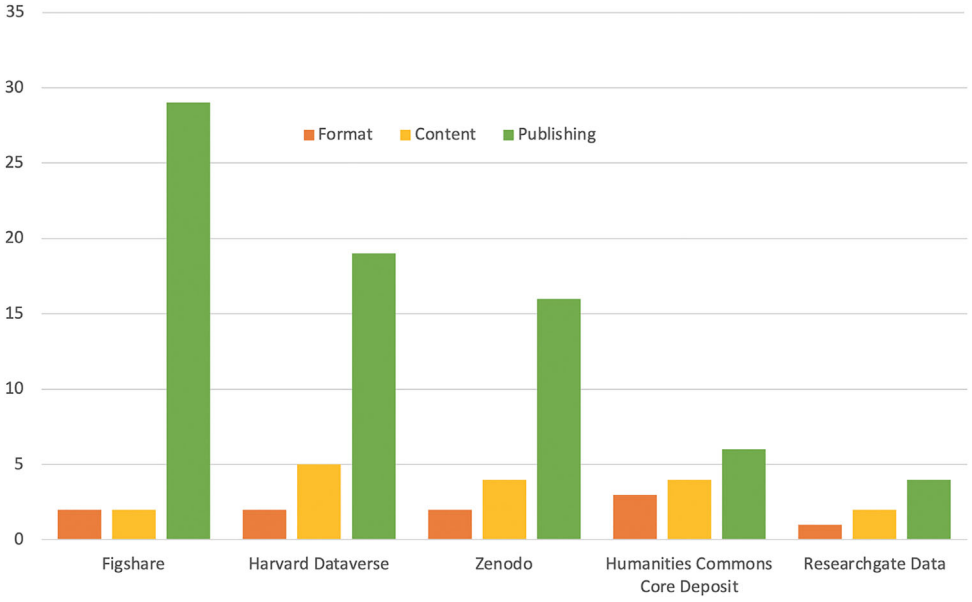

FIGURE 5 Number of metadata categories of the metadata types' format, content, and publishing, on dataset level in five open research data repositories [Color figure can be viewed at wileyonlinelibrary.com]

DATASET (collection level)

Metadata type	*Metadata category*	Figshare	H. Dataverse	Zenodo	Humanities C	Researchgate D
Format	Software		1			
	Language	1	1	1	1	
	File type				1	
	Item Type	1		1	1	1
Content	Title	1	1	1	1	1
	Subject area	1	1	1	1	
	Description		1	1	1	1
	Keywords		1	1	1	
	Time Period Covered		1			
Publishing	Author name	1	1	1	1	1
	Author id	1	1	1		
	Institution name	1	1	1		
	Institution id	1	1			
	Funding	1	1	1		
	Contributors	1		1	1	
	Origin of Sources	1	1			
	References	1	1	1		
	Embargo date	1	1	1	1	
	Related Material	1	1			
	Related Datasets	1	1			
	Other References	1	1			
	Data Sources	1	1			
	Characteristic of So	1	1			
	Documentation and	1	1			
	Publication type	1			1	
	Publication name	1		1		
	Conference name	1		1		
	Book/Report/Chapt	1		1		
	Thesis	1		1		
	Community/group	1		1	1	
	DOI	1	1	1		1
	Supplement to	1		1		1
	Licence	1		1	1	
	Access rights	1		1		
	Version	1		1		
	Series	1	1			
	Date of Collection		1			
	Publication date	1	1		1	1
	Depositor	1	1			
	Deposit date	1	1			

DATA (Item level)

Metadata type	*Metadata category*	Figshare	H. Dataverse	Zenodo	Humanities C	Researchgate D
Format	Image category			1		
	File path		1			
Content	File name		1			
	File description		1			
	Keywords/Tags		1			
Publishing	Provenance		1			
	Provenance description		1			

FIGURE 6 Comparing available metadata on the dataset level with metadata on the data level in five open research data repositories [Color figure can be viewed at wileyonlinelibrary.com]

they are completely devoid of any metadata in terms of the broader cultural-historical context of the image. This completely disregards the needs of image researchers, who primarily require information within a cultural-historical context, such as how and by whom an image has been produced, circulated, and used. Nor is there any information about the format of the original image, such as image medium or size, in these repositories—which is important information for many scholars in the humanities (see, for example, Chassanoff, 2018; Dahlgren & Hansson, 2021).

The emphasis or implied weight on sharing data is also reflected in the quantity of metadata available on the interfaces. ResearchGate provides, for example, a very simple interface without many metadata opportunities. On the other hand, data publishing counts as a publication in the interface, which helps to raise the user's metrics. When we looked at how some researchers used the interface for publishing their datasets, we found that simple interfaces such as ResearchGate were not used as a data repository, but rather as a placeholder for a reference to datasets published elsewhere, such as in Dataverse or Zenodo. Thus, this indicates these data repositories are not used as a primary publication channel for sharing data but as means to create greater visibility for the research/researcher.

Metadata for images can be described on several levels. The first is a general one, which is often called the dataset level. Here the collection of files is described as a whole. In the second, metadata can also be described for each file. Here each file can sometimes have its own metadata set. When it comes to images, it is crucial that they are enriched by metadata as there are no other means to understand the file without it, unlike a text file that can itself be searched. When we compare the metadata, where it is possible to link to image files in the interfaces, with the detailed metadata that the researchers are looking for, we see large differences at the dataset level and huge differences at the individual file level as seen in Figure 6. On the left is the available metadata for the dataset as a whole; on the right is the available metadata for each data file. This shows how the files in many of the interfaces are difficult or impossible to categorize individually.

The metadata is found mainly at dataset level; only one, Dataverse, has more than one metadata category on file level. But even here it is very lean and contains only the file name, file description, and keywords. It is therefore very difficult to create qualified metadata for individual image files in these data repositories. But it is possible to describe the whole collection of image files only as one entity. This shows that in the interface, data are treated as a homogeneous entity; nothing can be described in any detail at file level. For example, a collection of portraits uploaded to any of these sites could include information about the type of images (photographs) and the genre (portraits) or the full span of its production period (1850–1950). However, the individual names of the people portrayed, the photographers who have produced these images, their gender, nationality, and the context of their varying production contexts cannot be covered by existing metadata templates.

5 | CONCLUSION OF RESULTS

Analyzing the interfaces with norms of research practices in mind, we found that these interfaces represent a diverse palette of research approaches and practices: from attempts to provide researchers with dynamic CVs where data publishing becomes a step to promotion, where open data are about efficiency rather than criticality, to attempts to create platforms for collaboration and sharing the research process.

While the research paradigms signified are diverse, providing something for almost everyone, what these interfaces all have in common is that they are clearly positioned within an academic setting where the focus is on publication and status. When comparing the metadata available, we find a strong emphasis on publishing metadata in the system, which is very different from the researchers' need for detailed content metadata on file level. The design of the systems also assumes that the actual "data" in the project are homogeneous and can be described on a general level as one quantity. This is why it is difficult to describe individual image files qualitatively in these data repositories.

Thus, the focus and immediate appearance of the interfaces are differentiated for particular user groups in terms of signifiers of research paradigms but, at the core, the system design and functionality are similar, and they determine what it is possible to do in a way that is frequently in sharp contrast to the superficial and apparent affordance.

6 | DISCUSSION

The European Commission regards universal open data repositories as central to the infrastructure that will enable more open research (European Commission, 2016). Emphasis is placed on the opportunities for technology to connect researchers, reuse data, and create more participatory research processes, and thus increase innovation in society. Especially when it comes to the searchability of digital images, data repositories, where users can directly enrich images with metadata, can potentially contribute to

research in the humanities and for our understanding of history. Also, the open data paradigm would promote transparency and scholarly integrity as it would make it easier to evaluate the accuracy and solidity of a dataset.

Several problems have been highlighted as to why humanities scholars in particular have been reluctant to use open data repositories. The problem is situated, foremost, as being with the researcher: that they are too individualistic, as are their research processes; and that they might feel alienated by the built-in norms about the research process that exists in some repositories. Legal issues are another obstacle, such as problems with copyright that limit the right to collect, store, and reproduce the data required because those rights belong to others, as is the case with a considerable body of created artifacts, which are the object of study in the humanities. But above all, the greatest problem when it comes to reuse is the frequent lack of sufficiently qualifying metadata to describe the content of the data online (Borgman et al., 2015; Long & Schonfeld, 2014; Pasquetto et al., 2019).

In this study, we have therefore looked at the norms that are assumed to be embedded in the technology, while looking more closely at what the interfaces make possible, in practice, to support sharing and reuse. We found that multiple norms about research are embedded in these five repositories, when it comes to the signifiers of research expressed in graphic display and rhetoric, the perceived affordance for users determines if the users think they belong or not, and whether their needs will be met. However, when it comes to the technical means for metadata production on a system level, what they all share is a focus on the management, concentrating on scholarly publications rather than the sharing of data for transparency or reuse. When counting the metadata categories of publishing defining the data in relation to the academic metric systems, we found more than four times as many as compared with the categories in the two other metadata types together. The interfaces are thus primarily designed to collect metadata that is important for understanding the position of the data within the metrics system of the research world, collecting publication metrics such as funding, academic institution, journal, conference, research subject, and so on.

However, it is difficult to understand how the interface can support the sharing of images in a humanities-oriented collaborative research process, or other research practice where it is important to understand in detail how each image is produced and why, since metadata about the actual content of the material is lacking on item level (data level). In other words, the practices in the interfaces reflect the copyright system, with its focus on intellectual ownership, and the bibliometric system, with its focus on authorship. The metadata provided by some of the larger repositories such as ResearchGate is also often insufficient to identify the data source appropriately and is instead used by researchers as a way to advertise data published elsewhere (e.g., Dataverse or Zenodo). This corresponds with previous research on ASNS showing that the principal use is as a tool, not for sharing research, but for grading and monitoring researcher personas (Barbour & Marshall, 2012; Hahnel et al., 2017; Meishar-Tal & Pieterse, 2017).

This study also reveals a fundamental problem in the system design and the categorization of data, as metadata categorization is primarily about branding the output of the contributing researcher rather than providing datasets for other end users. Basically, the metadata that can be added to images is insufficient for it to be useful to find and reuse images as research data. More detailed metadata is only possible at the overall dataset level, which is in support of an idea of research data as something that is aggregated and homogeneous, rather than particular and specific, at odds with image datasets in the humanities.

The results reveal an ontological gap between the metadata needs and practices of researchers within the humanities using images as source material and the practice that the technical interfaces enable systematically when it comes to using images as research resources. It is therefore hardly surprising that humanities-oriented researchers are hesitant about using open research data repositories. Perhaps it is not so much about researchers' attitudes, as previous research on open data practices in the humanities has shown, but rather the false affordance in the interface design that does not deliver its promises. At the same time, it is easy to see the enormous possibilities that a more open approach to data might bring. In a humanities research tradition, sharing and openness are central to the research culture. In fact, historical data are to a large extent already openly available as deposits in public libraries, archives, and museums. Unlike other faculties where research data are produced by the research process, researchers in the humanities primarily reuse data generated by others. Here, we can develop strategies to provide the data with qualified contextual information, and for reuse to become more than just an ambition. However, sharing images as data on a larger scale requires a different design for data repositories—a design that puts the research material at the center and builds context from the ground in a more dynamic way; where the metadata is the data rather than the image, which for copyright reasons is found elsewhere.

The division into datasets that summarize the contents of a lump sum of images that do not in themselves contain details is not very useful, as the creation of a collection (or dataset) of images is often unique to each

research project and not immediately reusable. In addition, if the data are to be useful and function as the successful mobile research object as Latour (1987) suggests, each image file needs to be enriched with qualified metadata describing historical context, by whom, where, how, and why the image was created.

Hall (2016) warns that a shift in power is taking place through platform reform within academia, which instrumentalizes the research process and creates increased control over researchers. Our results show that these platforms, instead of supporting the reuse of data, primarily support such an instrumentalization of the research process, when the research is commodified and divided into smaller parts. What can mainly be shared and reused is not the content of the data, which is too deficient, rather the priority is metrics about the researcher's activity in the academic system.

ACKNOWLEDGMENTS

The research for this article was conducted within the project *Sharing the Visual Heritage* at the Department of Culture and Aesthetics, Stockholm University, funded by the Swedish Research Council (no. 2018-06057).

ORCID

Karin Hansson https://orcid.org/0000-0002-5962-1536

REFERENCES

Assante, M., Candela, L., Castelli, D., & Tani, A. (2016). Are scientific data repositories coping with research data publishing? *Data Science Journal*, *15*(6), 1–24. https://doi.org/10.5334/dsj-2016-006

Barbour, K., & Marshall, D. (2012). The academic online: Constructing persona through the world wide web. *First Monday*, *17*(9), 1–20. https://doi.org/10.5210/fm.v0i0.3969

Bhardwaj, R. K. (2017). Academic social networking sites: Comparative analysis of ResearchGate, academia.Edu, Mendeley and Zotero. *Information and Learning Science*, *118*(5–6), 298–316. https://doi.org/10.1108/ILS-03-2017-0012

Borgman, C. L. (2015). *Big data, little data, no data: Scholarship in the networked world* (pp. 1–383). Cambridge, MA: MIT Press.

Borgman, C. L., Darch, P. T., Sands, A. E., Pasquetto, I. V., Golshan, M. S., Wallis, J. C., & Traweek, S. (2015). Knowledge infrastructures in science: Data, diversity, and digital libraries. *International Journal on Digital Libraries*, *16*(3–4), 207–227.

Busch, L. (2017). *Knowledge for Sale: He neoliberal takeover of higher education* (2nd ed., pp. 1–154). Cambridge, MA: MIT Press.

Chassanoff, A. M. (2018). Historians' experiences using digitized archival photographs as evidence. *American Archivist*, *81*(1), 135–164.

Dahlgren, A., & Hansson, K. (2021). What an image is: The ontological gap between researchers and information specialist. *Art Documentation: Journal of the Art Libraries Society of North America*, *40*(1), 21–32.

Dataverse (n.d.). About | The Dataverse Project – Dataverse.org. Retrieved from https://dataverse.org/about.

Earley-Spadoni, T. (2017). Spatial history, deep mapping and digital storytelling: archaeology's future imagined through an engagement with the digital humanities. *Journal of Archaeological Science*, *84*, 95–102. https://doi.org/10.1016/j.jas.2017.05.003

European Commission. (2016). *Open innovation, open science, open to the world – a vision for Europe*. Retrieved from https://ec.europa.eu/digital-single-market/en/news/open-innovation-open-science-open-world-vision-europe.

Figshare. (n.d.). Figshare - credit for all your research. Retrieved from https://figshare.com/.

Gaver, W. W. (1991). Technology affordances. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 79–84).

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

Given, L. M., & Willson, R. (2018). Information technology and the humanities scholar: Documenting digital research practices. *Journal of the Association for Information Science and Technology*, *69*(6), 807–819. https://doi.org/10.1002/asi.24008

Hahnel, M., Treadway, J., Fane, B., Kiley, R., Peters, D., & Baynes, G. (2017). The state of open data report 2017. Figshare. doi:https://doi.org/10.6084/m9.figshare.5481187.v1

Hall, G. (2016). *The uberfication of the university*. University of Minnesota Press.

Harrower, N., Maryl, M., Biro, T., & Immenhauser, B. (2020). Sustainable and FAIR data sharing in the Humanities: Recommendations of the ALLEA Working Group E-Humanities. doi: https://doi.org/10.7486/DRI.tq582c863

Humanities Commons. (n.d.). About Humanities Commons. Retrieved from https://hcommons.org/about-humanities-commons/.

Hutchby, I. (2001). Technologies, texts and affordances. *Sociology*, *35*(2), 441–456. https://doi.org/10.1177/s0038038501000219

Kidwell, M. C., Lazarević, L. B., Baranski, E., Hardwicke, T. E., Piechowski, S., Falkenberg, L.-S., ... Nosek, B. A. (2016). Badges to acknowledge open practices: A simple, low-cost, effective method for increasing transparency. (M. R. Macleod, Ed.). *PLoS Biology*, *14*(5), e1002456. https://doi.org/10.1371/journal.pbio.1002456

Kraker, P., Lex, E., Gorraiz, J., Gumpenberger, C., & Peters, I. (2015). Research Data Explored II: the Anatomy and Reception of figshare. Retrieved from http://arxiv.org/abs/1503.01298.

Lakoff, G. (1980). *Metaphors we live by*. University of Chicago Press.

Latour, B. (1987). *Science in action: How to follow scientists and engineers through society*. Harvard University Press.

Leonelli, S. (2015). What counts as scientific data? A relational framework. *Philosophy of Science*, *82*(5), 810–821.

Long, M. P., & Schonfeld, R. C. (2014). Preparing for the future of research services for art history: Recommendations from the Ithaka S+R report. *Art Documentation: Journal of the Art Libraries Society of North America*, *33*(2), 192–205. https://doi.org/10.1086/678316

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

Meishar-Tal, H., & Pieterse, E. (2017). Why do academics use academic social networking sites? *International Review of Research in Open and Distance Learning*, *18*(1), 1–22. https://doi.org/10.19173/irrodl.v18i1.2643

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., ... Yarkoni, T. (2015). Scientific standards. Promoting an open research culture. *Science (New York, N.Y.)*, *348*(6242), 1422–1425. https://doi.org/10.1126/science.aab2374

Ochsner, M., Hug, S. E., & Daniel, H. D. (2016). *Research assessment in the humanities: Towards criteria and procedures. Research assessment in the humanities: Towards criteria and procedures*. Springer International Publishing. https://doi.org/10.1007/978-3-319-29016-4

Open Research Data Task Force. (2018). Case Studies: Annex to the final report of the Open Research Data Task Force. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/775379/Case-studies-ORDTF-July-2018.pdf.

Ortega, J. L. (2015). Disciplinary differences in the use of academic social networking sites. *Online Information Review*, *39*(4), 520–536. https://doi.org/10.1108/OIR-03-2015-0093

Ovadia, S. (2014). ResearchGate and academia.Edu: Academic social networks. *Behavioral and Social Sciences Librarian*, *33*(3), 165–169. https://doi.org/10.1080/01639269.2014.934093

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), 1–34.

Plantin, J. C., Lagoze, C., & Edwards, P. N. (2018). Re-integrating scholarly infrastructure: The ambiguous role of data sharing platforms. *Big Data and Society*, *5*(1), 1–14.

Poole, A. H., & Garwood, D. A. (2020). Digging into data management in public-funded, international research in digital humanities. *Journal of the Association for Information Science and Technology*, *71*(1), 84–97. https://doi.org/10.1002/asi.24213

re3data. (n.d.). Home | re3data.org. Retrieved from https://www.re3data.org/.

ResearchGate. (n.d.). About ResearchGate. Retrieved from https://www.researchgate.net/about.

Roche, D. G., Lanfear, R., Binning, S. A., Haff, T. M., Schwanz, L. E., Cain, K. E., ... Kruuk, L. E. B. (2014). Troubleshooting public data archiving: Suggestions to increase participation (J. A. Eisen, Ed.). *PLoS Biology*, *12*(1), e1001779. https://doi.org/10.1371/journal.pbio.1001779

Sicilia, M. A., García-Barriocanal, E., & Sánchez-Alonso, S. (2017). Community curation in open dataset repositories: Insights from Zenodo. *Procedia Computer Science*, *106*, 54–60. https://doi.org/10.1016/j.procs.2017.03.009

Thelwall, M. (2019). Do Mendeley reader counts indicate the value of arts and humanities research? *Journal of Librarianship and Information Science*, *51*(3), 781–788. https://doi.org/10.1177/0961000617732381

Thelwall, M., & Kousha, K. (2016). Figshare: A universal repository for academic resource sharing? *Online Information Review*, *40*(3), 333–346. https://doi.org/10.1108/OIR-06-2015-0190

Trace, C. B., & Karadkar, U. P. (2017). Information management in the humanities: Scholarly processes, tools, and the construction of personal collections. *Journal of the American Society for Information Science and Technology*, *68*(2), 491–507. https://doi.org/10.1002/asi

Van den Eynden, V., Knight, G., Vlad, A., Radler, B., Tenopir, C., Leon, D., ... Corti, L. (2016). *Towards Open Research: practices, experiences, barriers and opportunities*. Retrieved from https://doi.org/10.6084/m9.figshare.4055448

Vasquez, F. K., & Caicedo Bastidas, C. E. (2015). Academic Social Networking Sites: A Comparative Analysis of their Services and Tools. iSchools. Retrieved from http://hdl.handle.net/2142/73715.

Wuttke, U. (2019). "Here be dragons": Open access to research data in the humanities. https://doi.org/10.5281/zenodo.2630155

How to cite this article: Hansson, K., & Dahlgren, A. (2021). Open research data repositories: Practices, norms, and metadata for sharing images. *Journal of the Association for Information Science and Technology*, 1–14. https://doi.org/10.1002/asi.24571