

To share or not to share? Image data sharing in the social sciences and humanities

Late, Elina; Skov, Mette; Kumpulainen, Sanna

Published in:
Information Research

DOI (link to publication from Publisher):
[10.47989/ir292834](https://doi.org/10.47989/ir292834)

Creative Commons License
CC BY-NC-ND 4.0

Publication date:
2024

Document Version
Publisher's PDF, also known as Version of record

[Link to publication from Aalborg University](#)

Citation for published version (APA):
Late, E., Skov, M., & Kumpulainen, S. (2024). To share or not to share? Image data sharing in the social sciences and humanities. *Information Research*, 29(2), 386-400. <https://doi.org/10.47989/ir292834>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

Take down policy

If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.



Information Research, Special Issue: Proceedings of the 15th ISIC - The Information Behaviour Conference, Aalborg, Denmark, August 26-29, 2024

To share or not to share? Image data sharing in the social sciences and humanities

Elina Late, Mette Skov, and Sanna Kumpulainen

DOI: <https://doi.org/10.47989/ir292834>

Abstract

Introduction. The paper aims to investigate image data sharing within social science and humanities. While data sharing is encouraged as a part of the open science movement, little is known about the approaches and factors influencing the sharing of image data. This information is evident as the use of image data in these fields of research is increasing, and data sharing is context dependent.

Method. The study analyses qualitative semi-structured interviews with 14 scholars who incorporate digital images as a core component of their research data.

Analysis. Content analysis is conducted to gather information about scholars' image data sharing and motivating and impeding factors related to it.

Results. The findings show that image data sharing is not an established research practice, and when it happens it is mostly done via informal means by sharing data through personal contacts. Supporting the scientific community, the open science agenda and fulfilling research funders' requirements motivate scholars to share their data. Impeding factors relate to the qualities of data, ownership of data, data stewardship, and research integrity.

Conclusion. Advancing image data sharing requires the development of research infrastructures and providing support and guidelines. Better understanding of the scholars' image data practices is also needed.

Introduction

This paper aims to investigate image research data sharing within social sciences and humanities (SSH). Open science agenda advocates for the sharing and reuse of data to enhance research quality and generate cost savings. This initiative encounters challenges within the SSH, primarily due to the often sensitive nature of data pertaining to human behaviour. However, in certain SSH disciplines that utilise quantitative data, the practice of sharing and reusing data has become commonplace (Scheuch, 2003). Indeed, numerous data archives now serve as repositories for research data generated by both organisations and individual scholars (Corti, 2012).

Previous research on data sharing has mostly focused on data sharing on a general level and not analysed practices related to specific data types (Khan et al., 2023; Tenopir et al., 2015). However, sharing qualitative data (Broom et al., 2009; Jeng and He, 2022; Yoon, 2014) and social media data (Bishop and Gray, 2017; Hemphill et al., 2021) have attracted some research. This emphasis has left a gap in the understanding of image data sharing within the SSH as only a few studies on image data sharing have addressed the needs in these fields (e.g. Rodrigues and Lopes, 2023; Fernandes et al., 2020). Yet, the research shows that data sharing is complex and situated, and research on disciplinary data sharing is urgently needed (Akers and Doty, 2013; Cragin et al., 2010; Khan et al., 2023). Also, Thøgersen and Borlund (2022) call for more studies that clearly define what is meant by data sharing and to address the barriers to sharing data.

Visual data, for example, images sourced from social media platforms, have become important empirical data for SSH scholars exploring social behaviour (Ball and Smith, 2017; Chen et al., 2021). Despite many visual data sources being publicly available, numerous impediments, including copyright issues and compliance with the General Data Protection Regulation (GDPR), complicate its utilisation for research purposes. Hence, it is important to delve deeper into scholars' perspectives regarding open data sharing within the realm of visual data.

This study aims to enhance the comprehension of visual data sharing in the SSH. Based on qualitative interview data obtained from 14 Danish and Finnish SSH scholars who utilise digital image data in their research, we aim to address the following research questions:

RQ1: How is image data shared in the SSH?

RQ2: What factors either promote or impede scholars in their image data sharing?

The paper's structure is as follows: we start with a short review of the literature concerning the utilisation of visual data and data sharing within the SSH. We continue by describing our research design, the methods employed for data collection and analysis, and presenting our research findings. The article ends with discussion and conclusions.

Literature review

Images as research data in the SSH

Over the last decades, interest in visual research approaches and visual data has expanded in the SSH as a part of the visual turn in society (Ball and Smith, 2017; Chen et al., 2021). Digital images are produced and reproduced constantly as communication between people has moved towards visual communication. The development of digital technologies has made accessing image data easier and new computer vision technologies allow analysing massive amounts of images (Berg and Nelimarkka, 2023). However, only a few studies on image information behaviour focus on SSH scholars' image data use (Cho et al., 2022) for the purposes of information and illustration (Fidel, 1997; McCay-Peet and Toms, 2009). For example, Beaudoin (2014) studied image use among archaeologists, architects, art historians and artists. Specifically, those in archaeology and art history used images most often for knowledge creation of their lecture presentations, as well as for their research and subsequent publications. According to Rodrigues and Lopes (2023) qualitative analysis of digital photographs was most typical type of image data use in the SSH. Chassanoff's (2018) study showed that for historical research, photographs provide a valuable historical

reference for verification, documentation, or corroboration. Previous research has also focused, for example, on the production and needs for image metadata (Cetinic, 2021; Hanson and Dahlgren, 2022) and studied researchers' image retrieval, especially in the historical domain (Late et al., 2023; 2024).

Recently, the use of social media platforms as sources of data has attracted research. Social media data have provided various possibilities for visual research in terms of research topics, and methods. For example, Instagram which is a social media platform focusing on visual communication attracts a growing body of research (Highfield and Leaver, 2016; Rejeb et al., 2022). Image data from social media platforms can be collected in various ways including capturing API data. However, in recent years some platforms have disabled this possibility hindering image data use (McCrow-Young, 2021). According to Chen et al. (2021), there is an increasing interest in disciplines like sociology, cultural studies, communication, and environmental studies to use social media images as research data. Social media images are mostly collected manually and analysed by thematic coding, object recognition or narrative analysis. Although image data often contains copyright and privacy issues, papers analysing image data seldom mention these issues or apply strategies to address any ethical issues. Despite the emerging technologies to analyse large amounts of images, typically, the size of image data for research is relatively small and analysis exploits different manual qualitative methods (Chen et al., 2021).

Data sharing in the SSH

Borgman (2012) defines research data sharing as the '*release of research data for use by others*' (p. 1060). This release can take different forms, including informal or private sharing upon request to formal archiving to data repositories. Borgman (2012) presents four rationales for data sharing: reproducibility, serving public interest, asking new questions with open data, and advancing research. The FAIR principles suggest that merely opening data is not enough, it should be findable, accessible, interoperable, and reusable (Wilkinsson et al., 2016). Although data sharing has been increasingly discussed

due to the digitalisation of information and the open science movement, data sharing has a long history (Fienberg et al., 1985) and information sharing may be integral part of the scholars' research practice (Talja, 2002). Open access to data aims for improving reproducibility, efficiency, collaboration and interdisciplinarity of research assuming that data once archived will be useful and used by others (Borgman et al., 2019).

There is a lack of studies on image data sharing in the SSH. Among the few, Rodrigues and Lopes (2023; 2022a) show that researchers in SSH rarely share their image data in open repositories but a survey study by Fernandes et al. (2020) indicates that images are shared within the research group. Furthermore, Rodrigues and Lopes (2022a) argue that the lack of guidelines for managing image data leads to relying on individual practices. Hansson and Dahlgren (2022) analysed the affordances of data repositories for image data sharing in the context of humanities. They argue that these infrastructures often fail in facilitating data sharing but merely offer possibilities for promoting personas.

Beyond image data, there is a vast amount of literature, witnessing low levels of data sharing in the SSH (Jeng and He, 2022; Jeng et al., 2016; Kim and Stanton, 2016; Tenopir et al., 2015; Tenopir et al. 2011; Zenk-Möltgen et al., 2018). Sharing quantitative data has been more prevalent compared to qualitative data that is often contextual and personal and less easy to reuse (Broom et al., 2009; Jeng and He, 2022; Yoon, 2014). Recently, Khan and colleagues (2023) conducted a large-scale survey and showed disciplinary differences in data sharing. According to their survey scholars in the SSH rely on institutional and journal-supported repositories for data sharing. However, in many cases for example personal websites were also used.

According to Kim and Adler (2015) personal motivations (such as perceived career benefit and risk, perceived effort, and attitude) and normative pressure (collective expectations) support social scientist' data sharing and that institutional pressure has a marginal role. Lilja (2020) on the other hand argues that while

scholars often support the open science agenda there is a gap between open science policy and data practices especially of those using qualitative data that are experienced having contextual and relational character. Participants in Lilja's study saw data sharing as difficult and even leading to weakening the trust in science. For example, anonymization of data was seen as an unreliable practice for protecting participants' personal information and making research data useless for reuse by others. Other reasons for low levels of data sharing include sensitive data, lack of time, effort and skills, fear of not getting credit for authorship, and fear that the data will be misused or misinterpreted (Akers and Doty, 2013; Hemphill et al., 2021; Sayogo and Pardo, 2013; Tenopir et al., 2015). Also, ethical questions arise in all forms of data sharing (Bishop and Gray, 2017).

Research data and methods

This study is based on qualitative semi-structured interviews with 14 scholars in the SSH who incorporate digital images obtained from different sources as their primary research data. The interviews took place in both Finland and Denmark, facilitated either in-person or through an online platform between

February and August 2023. The interviews were conducted by the first author either in Finnish or English. The selection of interviewees followed a multifaceted approach. Participants were approached through personal connections, and a snowballing technique was employed, where interviewees were asked to suggest potential participants. Additionally, a web search of publications that had made use of image data was conducted to identify and reach out to suitable candidates.

Table 1 provides an overview of the interviewees' profiles. These participants encompass diverse fields within the SSH, with a majority specialising in cultural and media studies. As indicated by our recruitment process, visual data usage was most common in these disciplines. History domains were limited outside of the scope. Our interviewees also span a wide spectrum of work titles and seniority levels, ranging from doctoral students to full professors. Notably, every interviewee possessed a minimum of two years of experience in utilising images as research data, with the majority having more than five years of such experience. This broad sample enabled us to glean insights from scholars hailing from various academic backgrounds.

1Country	Finland (8), Denmark (6)
Work organisation	University A (7), University B (4), University C (2), University D (1)
Discipline	Cultural and media studies (8), sociology (3), political studies (1), socio-cultural psychology (1), information studies (1)
Work title	Professor (2), associate professor (5), post-doctoral researcher (5), doctoral student (2)
Experience of image data use	2-5 years (5), 6-10 years (4), 11-15 years (3), over 15 years (2)
Interview format	Face-to-face (11), online (3)

Table 2. Profile of the interviewees. Number of participants in parentheses.

Prior to the interviews, informed consents were collected, and participants were provided with comprehensive project details and an outline of the data collection process. The overarching themes of the interviews were clarified, and interviewees were asked to prepare themselves to discuss about one recent research project in

which they had employed image data. The interview questions encompassed initial background inquiries, delving into aspects such as the interviewees' current status and research field. Subsequently, a modified version of the critical incident technique (Flanagan, 1954) was employed, prompting interviewees to describe

their utilisation of image data in recent research tasks. This approach facilitated the collection of narratives pertaining to critical incidents associated with image data usage. The questions were designed to revolve around three distinct themes: the characteristics of visual research data, information practices related to visual research data, and openness in the context of image data. While the interview guide is openly accessible (Late and Kumpulainen, 2024), it is worth noting that the interviews didn't strictly follow the order of the guide. Instead, the guide served as a comprehensive checklist, ensuring that all relevant topics were addressed during the interviews.

The interviews were audio-recorded and transcribed into text for the analysis. On average, each interview spanned approximately 68 minutes, resulting in a total of 15 hours and 50 minutes of recorded audio data. Furthermore, several interviewees displayed their research materials and provided demonstrations of their data practices and shared their publications and presentation materials where they had employed image data. This aided the researcher in gaining a deeper and more comprehensive understanding of the interviewees' research projects and their work practices.

Qualitative content analyses were executed using a combination of Atlas.ti software and Microsoft Excel. Initial coding was done by one scholar but later discussed with the research team to find consensus. This was done to avoid biases in the analyses. The analysis process involved a series of iterative readings of the interview transcripts, open coding, and selective coding, following the methodology outlined by Strauss and Corbin (1997). The initial step encompassed multiple readings of the data to establish familiarity. During this initial coding phase, all instances related to data sharing (n=87) were identified within the research data. For the purposes of this study, we define data sharing broadly by following the definition by Borgman (2012) as releasing of research data for the use of others. This definition can include any kind of data sharing and is not restricted to depositing data in repositories, for example.

Citations linked to data sharing were subsequently extracted from the Atlas.ti software and transferred to Excel for further coding and analysis that were done data driven. Links to the original data were retained to facilitate subsequent reference if necessary. Descriptive information concerning the critical incidents (research tasks) discussed during the interviews were traced and described to provide background information pertaining to the tasks in which image data were utilised (see Table 2). Quotations were selected from the interviews to illustrate the findings. If needed, quotations were translated from Finnish to English.

Results

Characteristics of research tasks

The interviewees were asked to talk about one recent or ongoing research project where they used digital images as research data. Central characteristics of the research tasks are presented in Table 2. Described research tasks varied according to their phase as some were already finalised and some were in earlier stages of the research process. This variation provided a rich description of the research process. Image data were collected from various sources, social media platforms being the most typical one. In several cases (n=6) images were collected from various sources. Most scholars used images as primary data, but some integrated images with other data types, such as interviews, other social media data or register data. For example, images could be used as a part of interviews to aid the discussion with the participants. A majority (n=9) of the interviewees applied qualitative approaches as quantitative or mixed methods were utilised less frequently. Quantitative methods concerned applying machine learning and computer vision methods to analyse the data.

Research tasks represented diverse research topics that were grouped into four categories. Projects on *visual communication* focused for example on how images were used as a part of digital communication in specific context or by some specific group of people. These studies usually utilise small size of image data sets from social media platforms and apply qualitative

methods. Projects on *visual culture* or *visual representation* studied how certain phenomenon have been visualised and how, for example, this could be automatically detected. These studies utilised various image sources and sizes and both qualitative and quantitative methods. Projects on *visual practices* looked at how images or other visual materials have been

created. These projects used different image data sources, small and medium size data sets and combined images with different data types such as interviews. Finally, projects on *visual development* studied how visual appearance or use of visuals had changed over time. These studies relied on large data sets collected from different sources and quantitative methods.

Task phase	Finalised (6), reporting (4), analysis (1), data collection (3)
Image data sources	Social media images (15), internet images (3), digital publications (4), archival images (2), satellite images (1), corporate images (1), street art (1)
Size of image data set	Small 1-100 (4), medium 101-10000 (5), large 10000+ (5)
Purpose of image data use	As primary data (10), part of data (4)
Research methods	Qualitative methods (9), quantitative methods (2), mixed methods (3)
Research topics	Visual communication (5), visual culture/representation (5), visual practices (2), visual development (2)

Table 2. Characteristics of research tasks (n=14). Number of incidents in parenthesis.

Image data sharing

According to the interviews image data sharing is not a common practice. Only four interviewees had shared their data, four were about to share their data, and the rest (n=6) did not share their image data by any means. Interviewees' data sharing was categorised into *formal* and *informal* data sharing. *Formal* data sharing means storing materials in data archives specialised in housing research data. None of the interviewees had yet shared their data formally, however, some (n=2) had plans to do so. In both cases where image data sharing was planned, qualitative methods were applied to analyse the data of medium size. However, sharing codes or derived data can take place for projects handling large amounts of data. Formal image data sharing requires careful planning from the very beginning of the research process and collecting informed consents from the participants.

We will share a certain share of the data that we have agreed that everyone will collect. We have agreed on how to do it, and we worked it out for a long time, and we came

up with this. One participating university's information security system, to which the material can be shared and through which the material can then be accessed. P14

Some scholars felt pressure for data sharing from the research funders side but none of them recognised pressure from journal data policies. Many interviewees using qualitative methods considered formal data sharing out of their reach and mainly as practice for sharing quantitative data.

If I would have a big, automatically collected data set and I would use only a part of it, then I would consider sharing it. It would be amazing to find an archive, where I could find image data that suits to my topics. P5

Informal data sharing means sharing data directly through personal networks. Four interviewees had shared their data informally and some (n=2) expressed willingness if someone would ask. Informal data sharing concerned projects that analysed small or medium size data with qualitative methods. Personal contacts were the key issue in this

type of data sharing as information needs appeared during conversations with colleagues, social media discussions and in conferences, for example. Interviewees often wanted to know who would be using the data and for what purposes. When sharing, scholars also explained how data was collected, how it was used and what were the limitations. They also expected acknowledgement of data collection but did not follow if the shared data was at the end used.

If anyone asks, we are happy to share it [the data set] with people who know what they're doing and how to work with it. But to make it openly available is problematic. P12

Image data was shared informally by email or even by storing images to open Dropbox files,

where anyone having the access could download data. Many scholars also received image data from their colleagues and from other personal networks.

I've also created some small databases consisting of different strongly themed stuff that people might be interested in. [...] I just keep it in the Dropbox. Then everyone can kind of have access to it. P1

Motivating factors for data sharing

Although image data sharing was not a common data practice, interviewees expressed various motivators for doing so. Motivations applied for both formal and informal data sharing. Expressed motivators were categorised into three groups presented in Figure 1 below.

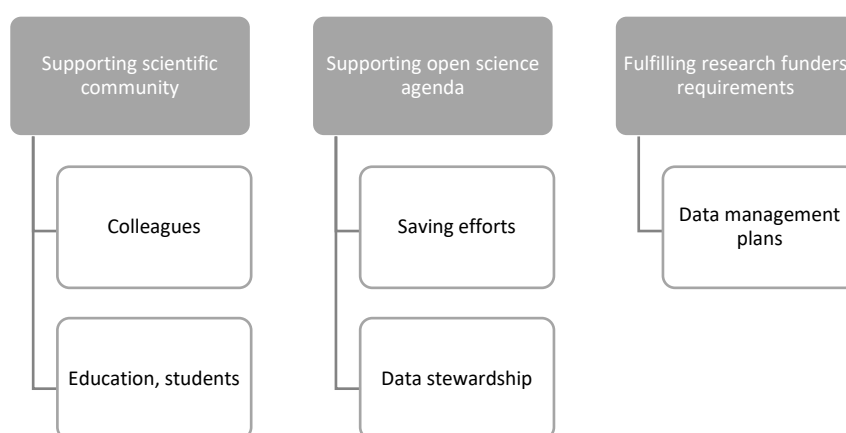


Figure 1. Motivating factors for image data sharing.

First, *supporting scientific community* was seen as an important motivator for image data sharing. Data sharing was seen as a part of research collaboration. Some also recognised the lack of modern image research data preventing students to use image data for their thesis.

For students the situation would be very different if there would be data available for their theses. Of course, you can find some in museums, but if you want to have something more modern. P4

Promoting the open science agenda was also a clear motivator for data sharing. By open access more image data could be easily accessed. Many interviewees collected image data constantly (for example from social media) even if they did not know how or when to use it. Therefore, they might have large number of images that they did not use or analyse by themselves. Sharing data might possibly save efforts from others and provide credits from data stewardship for the data collector.

It [getting merits from data collection] would be definitely something attractive to

consider. [...] But in most cases I think it would be really helpful if just across academia, but also in other platforms, to have more openness of what one can do with those, because it's a shame all the things that are lost because of gatekeepers all around. P7

Fulfilling research funders' requirements was also seen as one motivator for data sharing. Many research funders recommend open data sharing and require scholars to provide data management plans. Data policies may motivate scholars to plan their data management in the future in a way that enables them to share their image data openly.

I think it's possible if I have it in mind for a new project where I think from the beginning, I would like this [...] to be a shared [data set], and then from the beginning I would also keep different information more carefully about each image. P6

Impeding factors for data sharing
As most of the interviewees did not share their image data openly, they expressed various reasons for this. Expressed impeding factors were categorised into four groups presented in Figure 2 below.

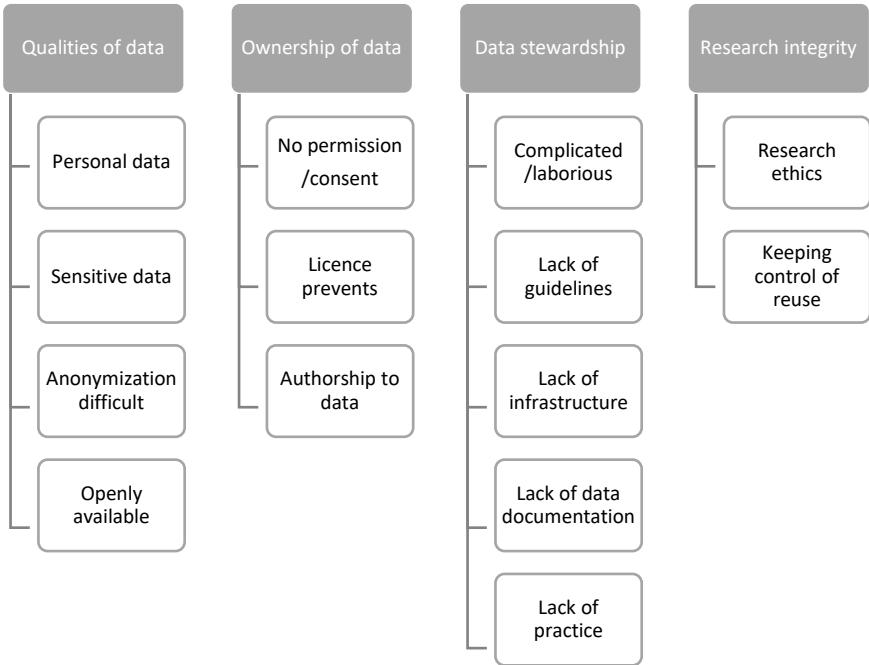


Figure 2. Impeding factors for image data sharing.

Firstly, the *qualities of the image data* impeded data sharing. Images often included personal and/or sensitive data that could not be shared. Images were also very difficult or even impossible to anonymise which made sharing hard and time consuming.

We thought about it [data sharing], but then GDPR came. And after GDPR it's not possible to share image data with people inside. P12

There should be only bananas and onions in the pictures so we could share them. [...] I hope that some solutions can be found, but right now it feels like there are just a lot of questions. P4

Many interviewees also thought that image data sharing was not even needed if images were collected from open sources. Thus, the data was openly available anyway for anyone to collect.

I might have some that are not anymore online, but I think that openness is also realized through the fact that it is public material. P2

Ownership of data was another factor hindering data sharing. In many cases scholars did not have permission to share image data or data sharing was prevented in the licence. For example, social media platforms or other sources where data was collected did not allow data sharing even when the data was openly available. Consent to share image data should be collected from various actors including persons in the images, the photographer and the organisation publishing the images. In many cases this was not possible.

Potentially there can be a confidential information in this material, so I'm not allowed to share it. [...] I have signed all sorts of consents that I will keep it private. I need to store it in a safe place. I can't share it with anyone but people who help me to analyse. P8

Some interviewees hesitated image data sharing because of their sense of authorship to the data. These feelings were mostly expressed by those using qualitative approaches. They felt that data collection was part of their intellectual work that they did not want to share with others.

Maybe it is possessive... especially if the data is small, and it's been terribly hard to put it together. Then I'd like to use it myself. Could someone else write about it? It would be a bit like that would be written from my point of view because that material is already collected from my point of view. P5

Lack of data stewardship also prevents data sharing. Sharing image data was considered as complicated and laborious. Additionally, interviewees perceived guidelines and support for image data sharing and data management incomplete or even missing. Many interviewees were insecure where to archive their image data if they were about to do so.

If I would anonymise image data from social media...I have no instructions how to do it.

What metadata and image data I should delete? P4

Another reason preventing image data sharing was the fact that open sharing was not planned during the data collection phase. Data collection was not documented in a way that would allow data sharing to take place at the end of the project. Some interviewees explained that data sharing was not part of their data practice and they had not considered data sharing during their research project. One participant discussed about the usefulness of archiving small individually collected data sets. He preferred larger curated image collections for future research needs.

I don't think that these general archive dumps...well they might be useful but not in the long run. P14

Finally, image data was not shared because of research integrity. Research ethics and scholars' own values or stance were sometimes against image data sharing.

I don't want to repeat the fact that those [images] are circulating on some platform again [...]. I must do it in a way that I can stand behind it. It's somehow clear though. Some texts about research ethics have also been written, and I am really in line with it. P3

Interviewees also wanted to keep control of who could use their data and for what purposes. Sharing data that could end up for example for commercial or political uses was something interviewees wanted to prevent. Some scholars also expressed a lack of trust in data archives' abilities to monitor the data use according to the licence.

But I'm also constricted because I think some of the information that I'm actually able to retrieve, I'm not sure I want someone to have access to it. P9

Discussion

This study analysed image data sharing within the SSH. Our qualitative interview data shows that image data sharing is not an established research practice, and when it happens it is mostly done via informal means by sharing data

through personal contacts. Indeed, scholars' social networks are important places for data sharing (Talja, 2002). This is an important observation since often only the formal data sharing is considered in open science policy (Reichmann, 2023; Thoegersen and Borlund, 2022). For instance, this could lead to open science and research support services aligning with the specifications of data archives and repositories rather than serving the preferences of researchers.

Informal data sharing can also be considered as information sharing characterised as a collaborative information behaviour (Talja and Hansen, 2006). Informal data sharing is not only about sharing the data itself but also practices related to data collection and use which is important especially when sharing qualitative data. When data is shared informally, researchers may have better control over who can use the data and for what purposes (Bishop and Gray, 2017). In formal data sharing, there is a fear of losing this control. Therefore, digital services should be designed to support scholars' work and collaboration practices that are socially constructed and develop much more slowly than technology does (Fry and Talja, 2007; Talja, 2002).

Although image data is rather rarely shared, scholars recognise various motivators for doing so. These motivators mostly relate to collegial and institutional issues. Kim and Adler (2015) witnessed the normative pressure for data sharing related to the collective expectation of the community. Our results also support findings from earlier studies (Jeng and He 2022; Zhu, 2019) that many researchers support the idea of open data, but they rarely share their own data. While the open science agenda was brought up in the discussions, scholars mainly consider data sharing for the benefit of the scholarly community, not for the larger public (see also Lilja, 2020). Research funders' data policies were seen as a motivating factor, and they might also shape the research practices in the future.

An important part of our findings concerns the impeding factors for sharing image data. These factors relate to the qualities of data, ownership of data, data stewardship, and research

integrity. The results suggest that the data itself is a major factor influencing data sharing. Images often include personal or even sensitive data. For instance, if data sharing is prohibited by the GDPR or by the data license, sharing is not possible. Collecting consents from various actors is often challenging. Other studies have also discussed the strict interpretations of GDPR and copyrights in the European Union and their influences on data sharing (Lilja, 2020; Chawinga and Zinn, 2019). Rodrigues and Lopes (2022a) also point out that data life cycle models do not contemplate data security and privacy procedures as a main phase of data management although the issue is relevant in all cases where personal data is been processed.

Further, many scholars connected formal data sharing mostly with quantitative data. However, scholars handling large amounts of data did not their share data, because licenses prevented it. Indeed, ownership of big data might be clearer compared with smaller hand selected data sets. Sometimes scholars want to hold their ownership of the data or ownership is unclear (see also Jeng and He, 2022). Impeding factors related to data stewardship are often related to unclear situations. In fact, many scholars were unaware of how and where they could share their data. For example, anonymization of image data was seen as difficult or even impossible. Therefore, support and guidelines for image data sharing are desperately needed. Luckily, development towards this direction has already taken place as a new metadata model and vocabulary have been recently developed for image data (Rodrigues and Lopes, 2022b). Providing such instruments helps researchers in their data management and saves time. However, research processes that are not always straightforward in SSH fields, may complicate data documentation making data sharing at the end of the project difficult or even impossible. Jeng and He (2022) argued that social scientists rarely receive formal training in data management or sharing. Yet, training, and open communication within the community are essential for research infrastructure development (Sendra et al., 2023). Data sharing can also be impeded because of research integrity issues as scholars carry great responsibility for the data they have

collected. Fear of misuse or misinterpretation has been recognised in earlier studies as well (Sayogo and Pardo, 2013). Scholars need to trust data archives' abilities to secure responsible reuse of data (see also Bishop and Gray, 2017). Therefore, it is important to engage scholars in the development of digital research infrastructures (Foka et al., 2018).

As our sample of interviews was relatively small and included only some disciplines in the SSH it is likely that our results do not necessarily capture all issues around image data sharing. However, our results can provide an important beginning for the discussion about image data sharing as it is likely that the use of image data will increase in the SSH research. As most of our observations are consistent beyond image data, some concerning the qualities of data, ownership and data stewardship are more prevalent in image data. In many cases, image data sharing is not at the hands of the scholar, but legal issues prevent sharing. Support, guidelines and discussion around data management practices and ethical issues are surely needed. Thus, there is an urgent need for more research on data sharing considering a range of data types.

Conclusions

This paper provides the first insights into image data sharing within SSH. The results show that image data is rarely shared, and that sharing happens mostly via informal means relying on scholars' personal contacts. By sharing image data informally scholars can control the access

to the data and also share their research practices related to the data. Although data sharing is often considered more prevalent for bigger data sets, according to the findings image data sharing concerns mainly small or medium size data. Formal data sharing through data repositories is rare in the context of image data. Because open science agenda usually identifies only formal data sharing, further discussion about the importance of recognising informal data sharing practices is needed.

Although scholars are often committed to supporting their scientific community and advancing open science, these ideals do not always realise in their data practices. Reasons for this concern the qualities of the data, ownership issues, data stewardship, and research integrity. For advancing openness to image data more support in the form of research infrastructures and guidelines is needed –but keeping the researchers' needs in mind. This can be achieved by focusing on researching scholars' visual data practices. This could yield a new understanding of how image data is collected and used and how open science practices should be applied in the context of image data.

Acknowledgements

This work was supported by the Research Council of Finland, grant number 351247. We would like to thank the Information studies seminar participants at Tampere University and two anonymous referees for their valuable comments and support.

About the authors

Elina Late (<https://orcid.org/0000-0002-3232-1365>) works as a Senior Research Fellow at Tampere University (Finland) in the Faculty of Information Technology and Communication Sciences. She holds a PhD in information studies. Her research interests include scholar's information and data practices, scholarly communication, and open science. Currently she studies open access to image information and its use as research data. She can be contacted at elina.late@tuni.fi.

Mette Skov (<https://orcid.org/0000-0002-8821-0314>) is Associate Professor in Information Behaviour at Aalborg University, Department of Communication and Psychology, Denmark. She holds a PhD in Library and Information Science from the Royal School of Library and Information Science, Denmark. Her main research interests include everyday life information seeking, user studies, user experience and interaction design. She can be contacted at skov@ikp.aau.dk

Sanna Kumpulainen (<https://orcid.org/0000-0002-7016-257X>) holds the position of Associate Professor in Information Studies at Tampere University, Finland. Her research centers on exploring the dynamics of human interaction with information and developing strategies for its facilitation. Her research interests encompass also interactive information retrieval, digital libraries, and the promotion of open science. She can be contacted at sanna.kumpulainen@tuni.fi.

References

- Akers, K. G., & Doty, J. (2013). Disciplinary differences in faculty research data management practices and perspectives. *International Journal of Digital Curation*, 8(2), 5–26. <https://doi.org/10.2218/ijdc.v8i2.263>
- Ball, M., & Smith, G. (2017). Working with visual data: Practices of visualization and representation. *International Review of Qualitative Research*, 10(2), 119–127. <https://eprints.staffs.ac.uk/id/eprint/3831>
- Beaudoin, J. E. (2014). A framework of image use among archaeologists, architects, art historians and artists. *Journal of Documentation*, 70(1), 119–147. <https://doi.org/10.1108/JD-12-2012-0157>
- Berg, A., & Nelimarkka, M. (2023). Do you see what I see? Measuring the semantic differences in image-recognition services' outputs. *Journal of the Association for Information Science and Technology*, 74(11), 1307–1324. <https://doi.org/10.1002/asi.24827>
- Bishop, L. & Gray, D. (2017). Ethical challenges of publishing and sharing social media research data. In Woodfield, K. (Ed.), *The Ethics of Online Research. Advances in Research Ethics and Integrity* (2nd ed., pp. 159–187). Emerald Publishing Limited. <https://doi.org/10.1108/S2398-601820180000002007>
- Borgman, C. L. (2012). The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology*, 63(6), 1059–1078. <https://doi.org/10.1002/asi.22634>
- Borgman, C. L., Scharnhorst, A., & Golshan, M. S. (2019). Digital data archives as knowledge infrastructures: Mediating data sharing and reuse. *Journal of the Association for Information Science and Technology*, 70(8), 888–904. <https://doi.org/10.1002/asi.24172>
- Broom, A., Cheshire, L., & Emmison, M. (2009). Qualitative researchers' understandings of their practice and the implications for data archiving and sharing. *Sociology*, 43(6), 1163–1180. <https://doi.org/10.1177/003803850934570>
- Cetinic, E. (2021). Towards generating and evaluating iconographic image captions of artworks. *Journal of Imaging* 7(8), 123. <https://doi.org/10.3390/jimaging7080123>
- Chassanoff, A. M. (2018). Historians' experiences using digitized archival photographs as evidence. *The American Archivist*, 81(1), 135–164. <https://doi.org/10.17723/0360-9081-81.1.135>
- Chawinga, W. D., & Zinn, S. (2019). Global perspectives of research data sharing: A systematic literature review. *Library & Information Science Research*, 41(2), 109–122. <https://doi.org/10.1016/j.lisr.2019.04.004>
- Chen, Y., Sherren, K., Smit, M., & Lee, K. Y. (2021). Using social media images as data in social science research. *New Media & Society*, 25(4), 849–871. <https://doi.org/10.1177/14614448211038761>

- Cho, H., Pham, M. T., Leonard, K. N., & Urban, A. C. (2022). A systematic literature review on image information needs and behaviors. *Journal of Documentation*, 78(2), 207-227. <https://doi.org/10.1108/JD-10-2020-0172>
- Corti, L. (2012). Recent developments in archiving social research. *International Journal of Social Research Methodology*, 15(4), 281-290. <https://doi.org/10.1080/13645579.2012.688310>
- Cragin, M. H., Palmer, C. L., Carlson, J. R., & Witt, M. (2010). Data sharing, small science and institutional repositories. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 368(1926), 4023-4038. <https://doi.org/10.1098/rsta.2010.0165>
- Fernandes, M., Rodrigues, J., Lopes, C. (2020). Management of research data in image format: An exploratory study on current practices. In Hall, M., Merčun, T., Risse, T., Duchateau, F. (Eds), *Digital Libraries for Open Knowledge: TPD 2020*. (Lecture Notes in Computer Science vol 12246). Springer. https://doi.org/10.1007/978-3-030-54956-5_16
- Fidel, R. (1997). The image retrieval task: implications for the design and evaluation of image databases. *New Review of Hypermedia and Multimedia*, 3(1): 181-199. <https://doi.org/10.1080/13614569708914689>
- Fienberg, S.E., Martin, M.E. & Straf, M.L. (Eds) (1985). *Sharing research data*. National Academies Press.
- Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin*, 51(4), 327. <https://doi.org/10.1037/h0061470>
- Foka, A., Misharina, A., Arvidsson, V., & Gelfgren, S. (2018). Beyond humanities qua digital: Spatial and material development for digital research infrastructures in HumlabX. *Digital Scholarship in the Humanities*, 33(2), 264-278. <https://doi.org/10.1093/llc/fqx008>
- Fry, J., & Talja, S. (2007). The intellectual and social organization of academic fields and the shaping of digital resources. *Journal of Information Science*, 33(2), 115-133. <https://doi.org/10.1177/0165551506068153>
- Hansson, K., & Dahlgren, A. (2022). Open research data repositories: Practices, norms, and metadata for sharing images. *Journal of the Association for Information Science and Technology*, 73(2), 303-316. <https://doi.org/10.1002/asi.24571>
- Hemphill, L., Hedstrom, M. L., & Leonard, S. H. (2021). Saving social media data: understanding data management practices among social media researchers and their implications for archives. *Journal of the Association for Information Science and Technology*, 72(1), 97-109. <https://doi.org/10.1002/asi.24368>
- Highfield, T. & Leaver, T. (2016) Instagrammatics and digital methods: studying visual social media, from selfies and GIFs to memes and emoji. *Communication Research and Practice*, 2(1), 47-62. <https://doi.org/10.1080/22041451.2016.1155332>
- Jeng, W., & He, D. (2022). Surveying research data-sharing practices in US social sciences: A knowledge infrastructure-inspired conceptual framework. *Online Information Review*, 46(7), 1275-1292. <https://doi.org/10.1108/OIR-03-2020-0079>

- Jeng, W., He, D., & Oh, J. S. (2016). Toward a conceptual framework for data sharing practices in social sciences: A profile approach. *Proceedings of the Association for Information Science and Technology*, 53(1), 1-10. <https://doi.org/10.1002/pr2.2016.14505301037>
- Khan, N., Thelwall, M. and Kousha, K. (2023). Data sharing and reuse practices: Disciplinary differences and improvements needed. *Online Information Review*, 47(6), 1036-1064. <https://doi.org/10.1108/OIR-08-2021-0423>
- Kim, Y., & Adler, M. (2015). Social 'scientists' data sharing behaviors: Investigating the roles of individual motivations, institutional pressures, and data repositories. *International Journal of Information Management*, 35(4), 408-418. <https://doi.org/10.1016/j.ijinfomgt.2015.04.007>
- Kim, Y., & Stanton, J. M. (2016). Institutional and individual factors affecting scientists' data-sharing behaviors: A multilevel analysis. *Journal of the Association for Information Science and Technology*, 67(4), 776-799. <https://doi.org/10.1002/asi.23424>
- Late, E., & Kumpulainen, S. (2024). Interview guide for SSH scholars about their image data use. Zenodo. <https://doi.org/10.5281/zenodo.10807674>
- Late, E., Ruotsalainen, H., & Kumpulainen, S. (2023). In a perfect world: Exploring the desires and realities for digitized historical image archives. *Proceedings of the Association for Information Science and Technology*, 60(1), 244-254. <https://doi.org/10.1002/pr2.785>
- Late, E., Ruotsalainen, H. & Kumpulainen, S. (2024). Image searching in an open photograph archive: search tactics and faced barriers in historical research. *International Journal on Digital Libraries*. <https://doi.org/10.1007/s00799-023-00390-1>
- Lilja, E. (2020). Threat of policy alienation: Exploring the implementation of Open Science policy in research practice. *Science and Public Policy*, 47(6), 803-817. <https://doi.org/10.1093/scipol/scaa044>
- McCay-Peet, L., & Toms, E. (2009). Image use within the work task model: Images as information and illustration. *Journal of the American Society for Information Science and Technology*, 60(12), 2416-2429. <https://doi.org/10.1002/asi.21202>
- McCrow-Young, A. (2021). Approaching Instagram data: Reflections on accessing, archiving and anonymising visual social media. *Communication Research and Practice*, 7(1), 21-34. <https://doi.org/10.1080/22041451.2020.1847820>
- Rejeb, A., Rejeb, K., Abdollahi, A., & Treiblmaier, H. (2022). The big picture on Instagram research: Insights from a bibliometric analysis. *Telematics and Informatics*, 101876. <https://doi.org/10.1016/j.tele.2022.101876>
- Reichmann, S. (2023). Mobile researchers, immobile data: Managing data (producers). *Social Studies of Science*, 53(3), 341-357. <https://doi.org/10.1177/03063127231156862>
- Rodrigues, J., & Lopes, C. (2023). Research image management practices reported by scientific literature: An analysis by research domain. *Open Information Science*, 7(1), 20220147. <https://doi.org/10.1515/opis-2022-0147>
- Rodrigues, J., Lopes, C. (2022a). Research data management in the image lifecycle: A study of current behaviors. In Guizzardi, R., Ralyté, J., Franch, X. (Eds), *Proceedings of the International Conference on Research Challenges in Information 2022*. (Lecture Notes in Business Information Processing, vol 446). Springer. https://doi.org/10.1007/978-3-031-05760-1_3

Rodrigues, J., & Lopes, C. (2022b). Describing data in image format: Proposal of a metadata model and controlled vocabularies. *Journal of Library Metadata*, 22(3-4), 213-234. <https://doi.org/10.1080/19386389.2022.2117511>

Sayogo, D. S., & Pardo, T. A. (2013). Exploring the determinants of scientific data sharing: Understanding the motivation to publish research data. *Government Information Quarterly*, 30, S19-S31. <https://doi.org/10.1016/j.giq.2012.06.011>

Scheuch, E. K. (2003). History and visions in the development of data services for the social sciences. *International Social Science Journal*, 55(177), 385-399. <https://doi.org/10.1111/j.1468-2451.2003.05503004.x>

Sendra, A., Late, E., & Kumpulainen, S. (2023). More than data repositories: perceived information needs for the development of social sciences and humanities research infrastructures. *Information Research*, 28(4), 83-101. <https://doi.org/10.47989/ir284598>

Strauss, A., & Corbin, J. M. (1997). *Grounded theory in practice*. Sage.

Talja, S. (2002). Information sharing in academic communities: Types and levels of collaboration in information seeking and use. *New Review of Information Behavior Research*, 3(1), 143-159.

Talja, S., & Hansen, P. (2006). Information sharing. In Spink & Cole (Eds.), *New directions in human information behavior* (pp. 113-134). Springer.

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., Manoff, M. & Frame, M. (2011). Data sharing by scientists: practices and perceptions. *PLoS One*, 6(6), e21101. <https://doi.org/10.1371/journal.pone.0021101>

Tenopir, C., Dalton, E. D., Allard, S., Frame, M., Pjesivac, I., Birch, B., Pollock, D. & Dorsett, K. (2015). Changes in data sharing and data reuse practices and perceptions among scientists worldwide. *PLoS One*, 10(8), e0134826. <https://doi.org/10.1371/journal.pone.0134826>

Thoegersen, J. L., & Borlund, P. (2022). Researcher attitudes toward data sharing in public data repositories: A meta-evaluation of studies on researcher data sharing. *Journal of Documentation*, 78(7), 1-17. <https://doi.org/10.1108/JD-01-2021-0015>

Zenk-Möltgen, W., Akdeniz, E., Katsanidou, A., Naßhoven, V., & Balaban, E. (2018). Factors influencing the data sharing behavior of researchers in sociology and political science. *Journal of documentation*, 74(5), 1053-1073. <https://doi.org/10.1108/JD-09-2017-0126>

Zhu, Y. (2019). Open-access policy and data-sharing practice in UK academia. *Journal of Information Science*, 46(1), 1-12. <https://doi.org/10.1177/0165551518823174>

Yoon, A. (2014). "Making a square fit into a circle": Researchers' experiences reusing qualitative data. *Proceedings of the American Society for Information Science and Technology*, 51(1), 1-4. <https://doi.org/10.1002/meet.2014.14505101140>

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. & 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>