Nordic Journal of Library and Information Studies

Vol. 6, No. 2, 2025
ISSN (ONLINE) 2597-0593
DOI https://doi.org/10.7146/njlis.v6i2.158666
© CC BY-NC-ND 4.0

Peter Bauer , PhD in history and history didactics, visiting research fellow at department of Arts and Cultural Sciences Lund University. https://orcid.org/0009-0001-4144-587X
peter.ericsson@hotmail.com

Constructing digital collections

An investigation to legal and ethical frameworks in creating and maintaining digital collections?

Abstract
This article investigates how legal, professional and ethical frameworks shape digital collections at six Swedish cultural heritage organizations by utilizing institutional theory. The study is based on semi-structured interviews with archivists and librarians who work closely with the digital platforms. Findings indicate that the creation of collections is primarily governed by GDPR and copyright law, alongside archival principles of accessibility. At the institutional level, both a practice of digitization-on-demand and informal policies regarding inclusion and exclusion have emerged. These informal policies often stem from contemporary norms as well as the need to protect the integrity of the subjects represented in the records. All six institutions engage in dialogue with surrounding society, which can lead to both the addition and removal of records and metadata. This particularly in cases where records or meta-data is deemed as enforcing social hierarchies. Privacy concerns are frequently addressed by the archivists, yet complaints are rare; instead, many individuals appreciate the publication of records concerning their relatives, even when this challenges contemporary privacy norms. Lastly, professional norms such as maintaining record integrity plays an important role.
Keywords: digitalization, research libraries, digital collections, ethical considerations, post-modern archival theory

Introduction

In recent years, increased internet access and lowered costs for digital storage have enabled some archives and libraries to provide online access to digitized cultural heritage materials. In Sweden this practice is often framed as a means for enhancing democracy, as broader access to cultural heritage

is viewed as a tool to increase civic participation (Thor Tureby & Wagrell, 2022, p. 347). However, the ongoing digitization process has also sparked ethical debates. Internationally this has led to initiatives such as DeCoGLAM which seeks to decolonize digital collections containing Eurocentric narratives (Glamlabs.io, n.d.).

The risk of reproducing Eurocentric narratives is one of many ethical issues associated with digital collections with other issues including infringements of privacy and reinforcement of social hierarchies. These risks can, however, be addressed during collection creation since digital collections are products of selective decisions on what to include or exclude (Alkemade, 2023, pp. 3–4). Such decisions are not made in isolation but are influenced by the organizational and social context in which the archives are created (Schwartz & Cook, 2002, p. 3). Furthermore, the rise of digital collections can be viewed as part of a broader digitalization trend in society, which has introduced "new organizational practices, managerial norms, and social structures across multiple sectors" (Schildt 2022, p. 1). Therefore, digital collections can be understood as outcomes of decisions that are products of institutional logics, shaped by both external legal, economic, cultural frameworks and professional norms.

Despite extensive discussion on digitization ethics, little empirical research has examined how Swedish heritage institutions navigate ethical challenges in practice. This study addresses this gap by investigating how managers at six major Swedish heritage organizations handle ethical considerations when selecting records for online publication. Drawing on institutional theory, the paper also examines how organizational structures, norms, and legal frameworks influence these decisions. The study departs from the following research questions:

- Which ethical challenges arise when creating and maintaining digital collections, and how are these challenges addressed?
- How do organizational context, financial resources, legal frameworks and professional norms shape digital collections?

By focusing on practical experiences, this study aims to provide insight to the mechanisms that shape digital collections in Sweden. Such knowledge is important as digital collections influence public understanding of cultural heritage and in the future might work as training data for machine learning applications (Dahlström et al., 2012, pp. 457–459; Alkemade, 2023, pp. 4–5).

Definitions of central concepts

In this paper the terms digitization, digitalization, digital collection and metadata are used. The first two are often used interchangeably, but it is worth noting that digitization refers to transforming an object from analogue to digital state. This process can take two forms: mass and critical digitization. Mass digitization is an automated process aiming to create digitized versions of as many records as possible, omitting manual, labor-intensive work such as interpretation and context description. In contrast critical digitization focuses on a select number of documents, seeking to maximize interpretation and contextualization (Dahlström et al., 2012, pp. 461–462).

Digitalization is a more ambiguous term but Legner et. al. defines it as "the manyfold sociotechnical phenomena and processes of adopting and using" digitization "in broader individual, organizational and societal contexts" (Legner et. al., 2017, p. 301). In this paper, digitalization is understood as the process of making digitized records available to the public via online databases.

In this paper "digital collection" is utilized instead of "digital archive". This is because archives are defined as an organization that holds records accumulated over time (Thomassen, 2001, p.374). In this study, it is rarely entire archives that are digitized but rather individual records. Thus, the collections discussed consist of fragments of rather than complete archives. Lastly, this study includes a discussion of metadata. This is motivated by previous research which highlights metadata's relevance to ethical considerations in creating digital collections. Further on, metadata also affects how records are found and understood, thus playing an important part in the public's understanding of the collections. In the paper, metadata is understood "data about data" and has been interpreted in the interviews as titles, subject headings, archival descriptions and keywords (Baca, 2016). Apart from these concepts, the concept of institutional logics is also of significance to the paper.

Theoretical perspective – an institutional view on digital collections

Following Schildt (2022), digitalization can be understood as an institution which W. Richard Scott (2001) defines as social structures that guide human behavior. These structures rest on three pillars that Scott labels as the regulative, the normative and the cultural-cognitive. The regulative pillar can here be understood as internal (i.e. policy) and external rules that guide digitization. One example of external rules is legislation such as GDPR, which archives must adhere to when creating digital collections. Sanctions for non-compliance include financial penalties, demonstrating how the state uses economic force to ensure compliance. However, Scott also notes that legislation is not always sufficiently clear enough to provide detailed prescriptions. Instead, law often serves as an "occasion for sense-making and collective interpretation" (Scott, 2001, pp. 47–50, 54).

Scott's second pillar, the normative pillar, consists of values—preferred or desirable ends—and standards used to assess practices. Norms define goals as well as appropriate means for attaining them. Norms sometimes only apply to specific parts of an organization, producing role expectations that guide professional behavior, which can lead to an internalization of norms (Scott, 2001, p. 55). This is relevant to archival work, as the profession has its own logics and standards. Previous research on the archivist profession is therefore used in this article to discuss these normative aspects.

The third pillar, the cultural-cognitive one, refers to shared interpretive schemas through which individuals make sense of the world. These schemas may operate as habitual routines or as deep-rooted assumptions about social reality. According to Scott this pillar overlaps with the normative pillar and often the two pillars often re-enforce each other (Scott, 2001, pp. 57–58). This is relevant for archival practice, where professional norms and shared assumptions jointly guide decision-making and interpretations of rules and regulations, as well as the archival sector's contacts with the public.

Compliance with the three pillars is, according to Scott, part of organizations' efforts to achieve legitimacy. Legitimacy is not viewed as a commodity but as a "condition reflecting consonance with relevant rules and laws, normative support, or alignment with cultural-cognitive frameworks" (2001, p. 59). In this case, legitimacy can be seen as generated by both the state and the public. By employing this theoretical lens, the aim is to show how selecting material for online publication are shaped by an institutional frame influenced by contemporary beliefs, ideas, and regulations. This perspective aligns with Schwartz & Cooks (2002, p.3) notion of archives as socially constructed. However, by utilizing this framework this study shifts attention from the power-centered themes often emphasized in postmodern archival research. Instead, the focus is on how legal, normative, and cultural forces shape the development of digital collections.

Previous research

This section begins with a presentation of research on privacy in digital collections, before moving onto issues of social justice. Since none of the heritage organizations studied holds any significant colonial collections, research on digitalization of post-colonial collections is not addressed.

Privacy and digital collections

A recurring theme in research on digital collections is the potential infringement of privacy. This is of little surprise since in recent years legal frameworks, such as the GDPR, have been established to protect the integrity of living individuals (Integritetsskyddsmyndigheten, 2021). However, as noted by folklorist Fredrik Skott (2014, pp. 55–56), perceptions of what constitutes private information change over time. For instance, beliefs in hobgoblins may seem harmless today but were considered highly sensitive in the 19th century when some records were created. Consequently, online publication may inadvertently breach privacy without the publisher being aware of doing so.

Another point, made by librarian Paul Dagleish (2011, pp. 71–73), is that online publication bypasses the unconscious and conscious restrictions heritage organizations put on their users. For instance, gaining access to records often requires visiting the archive and spending time in designated reading rooms. Archives can also restrict use in various ways, whereas digitalization removes such barriers. Furthermore, digitalization enables the dissemination of information that the record's creator might not have anticipated.

One way to protect privacy in digital collections is through anonymization, a practice problematized by historians Malin Thor Tureby and Christina Wagrell (2020, pp.112–113). According to the authors, anonymization may contribute to groups being perceived as vulnerable, potentially reinforcing marginalization in certain contexts. Another concern is that anonymization may threaten the integrity of the record. According to librarian David Mindel (2022, pp. 552–556), integrity means that the information represented in a digital replication is the same as in the original record. While Mindel focuses on image-enhancing tools, the practice of blurring or removing names can similarly be said to compromise integrity. Mindel also emphasizes the importance of transparency in describing which technologies were used during digitization.

Excluding records that contain sensitive information has been proposed as a tool to protect privacy. However, Manžuch (2017, pp. 7–12) argues that digitized collections should offer a more detailed and dynamic view of the past than general history writing. Therefore, the inclusion of marginalized groups—whose records often contain sensitive material— is an ethical obligation. This aligns with Mindel's concept of equity, which emphasizes collections need to represent of all segments of society (2022, p.547). Similarly, Skott (2014, pp. 58–59) cautions that omitting the darker aspects of history risks presenting a homogeneous past devoid of social conflict. Achieving equity, however, is challenging, as archival institutions rarely control digitization decisions entirely. This since the high cost of digitization often makes funding reliant on donations, allowing external interests to influence which materials are preserved (Manžuch, 2017).

The risk for strengthening social injustices

Digital collections not only raise privacy concerns but can also impact the records' subjects and broader communities. Archivists Michelle Caswell and Marika Cifor have addressed this through an "ethics of feminist care" framework, which shifts focus from the records themselves to their creators, subjects, users, and affected communities. A central aim in doing this is to "acknowledge, confront, rectify and repair oppressive power structures". In this approach, digital publication is part of a broader process involving engagement with creators, users, and impacted communities. For example,

in a project on Islamophobia, the authors worked closely with the local Muslim community, publishing records online alongside videos of Muslims sharing their perspectives. In another project, *Visual Aids*, artwork by HIV-positive artists from the 1990s was digitized, and for deceased artists, the website included the phrase "In memorial of" to honor their memory. The authors also once halted digitalization of an LGBTQ movement archive due to potential risks for individuals. This was justified by the conclusion that "sometimes, a feminist approach to the ethics of digitalization means learning to accept 'no' as the answer" (Caswell & Cifor, 2019, pp. 159–169)

Caswell has also co-authored an article with Tai et al. exploring how records, beyond evidential value, perform for their users in various ways. The article shows how records can empower marginalized groups and give voice to those silenced in mainstream archives. The authors argue that community archives play a crucial role by providing alternative records to those in traditional archives, which often lack staff diversity and contain embedded racial structures in their descriptions and metadata (Tai et al., 2019, pp. 15–17). It is worth noting that biases in metadata, such as derogatory terms for ethnic groups, have prompted initiatives like *The Metadata Justice Project*, where cataloguers and archivists update subject headings in older catalogues. However, removing such data could conflict with traditional archival principles, which hold that archives should reflect their historical context (Beezley & Scheele Clark, 2024, p. 20; Zhang, 2012, pp. 334–336).

Racial issues in a digital context are further discussed in Peterson Brink et al.'s (2016, pp. 21–22) study on digitizing the early 20th-century campus humor paper *The Awgwan*. The magazine's racist and sexist stereotypes created tension between its research value and the risk of reinforcing oppressive structures. The authors ultimately halted digitalization, as potential harm outweighed benefits. Disclaimers were rejected because not all users would read them and because they derive from a "pseudo-legal" tradition that could make the library appear self-protective. Ideally, the material would be presented with educational context explaining the period's attitudes. But since the project was lacking adequate resources to develop this context, online publication was postponed until sufficient contextualization could be provided.

The research presented in this and previous sections illuminates both ethical challenges and possible strategies for managing digital collections. A key difference between these studies and this study is that this one primarily engages with institutions at the practical level. Furthermore, the use of institutional theory also illustrates how factors outside of the organizations affect digitization practices.

Methodology – data collection

To investigate the practices that shape collections similar to those presented in previous research, this study employed semi-structured interviews with representatives from six major Swedish heritage organizations. Focus groups could have been an alternative, but they would have required more resources and risked lower participation (Denscombe, 2010, pp. 175–178). Managers were chosen as informants because they are central to decisions regarding what to include or exclude from collections. An initial survey of Swedish archival institutions with an online presence identified eight institutions, of which six agreed to participate in the final study. Despite some declines, these organizations still hold some of Sweden's largest digital collections (Table 1.1).

Table 1.1. Overview of informants

Informant	Organization description	Role	Objects in collection
A	Informant A is employed by a municipal board that maintains a portal containing material from both municipal and private archives and museums.	Web-editor	30,000 pictures and records
B	Informant B is employed by a municipal archive and oversees a large database of photographs from the municipality.	Picture archivist	90,000 pictures
C	Informant C is employed at one of Sweden's largest university libraries, which holds material from antiquity to the present. This material is published on a nationwide platform shared by several stakeholders.	Picture-archivist	10,000 objects
D	Informant D is employed at one of Sweden's largest university libraries (a different institution from informant C), which holds material from antiquity to the present. This material is published on a nationwide platform shared by several stakeholders.	Head of digitization	490,000
E	Informant E works at a language and folklore archive operated by a government authority.	Head of department	20,000 objects
F	Informant F is employed at a folklore and traditions archive at a university. They oversee two large databases containing both music and images.	Music archivist	20,000 objects

The informants represent three types of heritage organizations: municipal, research-based, and university libraries. Recruitment followed Adams' (2015, pp. 412–413) approach, with the organizations suggesting suitable managers for interviews. All informants are responsible for digital collections, although titles vary. The limited number reflects that Sweden has approximately 950 archivists, with only a subset working with public digital collections which limits the number of

potential informants (utbildningssidan.se, n.d.). The study aimed to reach thematic saturation, typically achieved after six interviews (Guest et al., 2020, pp. 13–14).

The interview guide was based on four themes: institutional frameworks for digitalization, selection of records, management of ethical challenges, and mediation of records, with prioritized sub-questions for limited time (Adams, 2015, pp. 496–497). Questions were sent one week in advance. Following Haukås and Tishakov (2025, pp. 64–66), this allowed informants to reflect and discuss practices with colleagues, enhancing response quality. A limitation is that answers may reflect institutional norms rather than actual behavior, but such reflections as well as collegial inputs is beneficial when exploring managerial beliefs. Additional measures, inspired by Denscombe, such as follow-up visits to digital collections, were applied to verify responses. The consistency of informants' accounts across interviews further supports credibility (Denscombe, 2010, pp. 188–189).

A potential source of bias is the researcher's positionality (Ravisan Goundar, 2025, pp. 2–3). The researcher holds a PhD in history, is a trained archivist, and has experience as a municipal archivist. These experiences provide insight into historical value, municipal practice, and legislation. However, the position of the research as academic might lead informants to feel pressure to provide "academically correct" answers. This was one of the reasons to submitting questions in advance, since this potentially reduce interview stress (Haukås & Tishakov, 2025, pp.64–66). Interviews were conducted on Zoom, lasted 50–60 minutes, and informants were briefed on the study's aims and the researcher's background. After initial revisions, the number of informants was expanded to six; interviews with B and F were conducted a year later without changes in overarching interview frameworks. All interviews were recorded, transcribed manually in Audacity, and imported into NVivo.

Data analysis

The material was analyzed using Braun & Clarke's (2006) framework for thematic content analysis, allowing the identification of recurring patterns and division of interviews into comparable units. Familiarization occurred during transcription and an additional read-through, followed by coding using categories such as legal, financial, and moral dimensions derived from previous research.

Initially, material was grouped into five main themes (legal, moral, financial, and organizational) identified deductively based on prior theoretical considerations. As these themes were part of the interview questionnaire, they were apparent at the semantic level and not treated as latent meanings. Coding progressed from whole segments (e.g., discussions on law) to line-by-line coding, capturing multiple aspects within each theme. Subthemes such as specific legislation (GDPR, copyright) and moral positions (integrity, racial injustice), were then identified inductively as they emerged from the data. This iterative process continued until all relevant data (excluding irrelevant parts such as the researcher's introduction or anecdotes) were sorted into themes.

Once themes were established, the material was reread and linked to the theoretical framework. A final recoding connected subthemes to Scott's three pillars. Some themes were merged, such as financial and organizational, while others became "headline themes". These headline themes were a grouping of sub-themes that the final text structured around: context in the work with digitalization, material and metadata reflecting darker social orders, and material that violates privacy. Overall, coding involved at least three main rounds: segment-based coding, line-by-line coding with subtheme identification and final recoding linking subthemes to theory.

To enhance credibility and reduce misinterpretation, informants reviewed the draft and provided feedback (Denscombe, 2010, p. 180). No objections were raised, validating the interpretations and returning some interpretative authority to the informants. The researcher's background as a social-historian and archivist informed the coding and analysis, which might for instance have led to an increased interest in both community archives and the inclusion of marginalized groups. It should, however, be noted that the respondents were also allowed to highlight aspects that we had missed during the interviews, hence reducing the risk for overstating these circumstances.

Context in the work with digitalization

The interviews described in the previous section started with a discussion on organizational structures, economic conditions and legal frameworks. This was done to investigate the context that according to Scott (2001) shapes institutional practice. Informant A's organization is placed under the educational board, while Informant B's organization is placed under the city archive, which in turn is subordinate to the city's cultural board. Informant A's organization does not hold collections of its own but provides a web portal for municipal, state-, and privately-run heritage organizations with material concerning the municipality's history. Organizations outside the municipality pay a fee of 1,500 euros a year to utilize the portal.

The main aim of Informant A's organization is to publish material relevant for schools, thus reflecting their placement under the school board. Informant B, on the other hand, views their organization's mission as providing material relevant to as many of the municipality's inhabitants as possible, reflecting their placement under a municipal board. This has led Informant B to focus on pictures of buildings and streets rather than historical portraits, as the latter would mostly interest genealogists. According to the informants, their organizations aim to create collections that cover the entire city rather than only the city center. Utilizing Scott's three-pillar model, this illustrates a normative expectation where the material should be relevant for a broad segment of the population. This seems to have influenced the development of informal in-house policies regarding what to publish. In the case of Informant A, there also exists a regulative element of needing to be relevant for schools.

The regulative element is also reflected in legislation. Informant B stated that GDPR has led the organization to refrain from publishing material that is "too recent" or depicts identifiable individuals. Informant B further noted that an EU directive requires that visually impaired individuals have access to assistance in interpreting the material, necessitating the creation of textual descriptions for images. Finally, both informants indicated that their organizations are funded internally and are therefore not dependent on external sources of financing.

Informants C and D work at university libraries and are therefore subject to different directives than municipal organizations. Both describe a change in the regulative element, with digitalization recently incorporated into their institutional policies. This amendment reflects a broader trend in recent years, where digitalization is increasingly regarded as a tool to enhance public access (Thor Tureby & Wagrell, 2022, p. 347). From a legal perspective, both informants report that GDPR and copyright legislation often result in material being excluded from their collections. Informant D emphasizes this by noting that their organization must ensure that "people are dead enough" for copyright restrictions to have expired. Overall, regarding regulative elements, the libraries studied appear to have internalized the increased demand for digitalization while simultaneously complying with external legal pressures such as GDPR and copyright legislation.

Both informants C and D emphasize that when selecting material to digitize, they often prioritize material at risk of decay. This reflects a normative element of their role as archivists, since preserving

material for future generations is a central professional responsibility. Although the libraries are primarily funded through their general budgets, they also practice digitalization on demand, meaning material is digitized for a fee when requested by a researcher. Informant D describes this fee as "very reasonable," as it covers only part of the costs since "neither the public nor the individual research could pay the actual cost". Because material digitized under this model is co-financed, the library retains the right to publish it online, avoiding the risk that the financier could restrict access (Manžuch, 2017, p. 6). This practice of keeping material publicly accessible may illustrate an overlap between regulative and normative elements, since archival law mandates accessibility to conventional archives while professional norms historically emphasize the archivist's dual role as custodian and facilitator (SFS 1990:782 §3; Cook, 2011, p. 608).

Informants E and F both work at government-run tradition archives and express that digitization primarily is a tool for preservation. Informant F notes that material is often digitized to prevent physical deterioration, while Informant E adds that digitalization increased during the COVID-19 pandemic to mitigate archive closures. Both examples reflect the normative element of the archival role, emphasizing the custodian-role of the archivist. However, Informant E also points out that, in the long term, digitalization reduced demand for physical visits to the research hall, thereby lowering staffing costs. This could be understood in the context of Swedish government organizations being subject to annual budget cuts aimed at increasing productivity (Riksrevisionen, 2022, p. 4). A possible interpretation is thus that a normative value element may exist, reflecting that cost-cutting practices have been internalized within the archival organization. Finally, both Informants E and F mention that digitization is mainly funded from internal budgets, although Informant F notes that digitization can also be a positive side effect of external projects with other primary aims.

Informants E and F also describe how regulative elements affect digitalization. Informant F states that material younger than fifty years is not published unless explicit consent is obtained from the records subjects. This is based on the assumption that people who were middle-aged in the 1970s are likely deceased and thus no longer protected under GDPR. This reasoning is since one consequence could be that some material could include living individuals. In this regard the interpretation of GDPR might be seen as a risk-assessment, rather than a strict interpretation. Hence it seems like in this case, GDPR functions primarily as a starting point for discussion which, according to Scott, is typical for laws that are ambiguous or difficult to implement (Scott, 2001, p. 54).

Furthermore, Informant F states that material such as "books, photographs, and literature" is not published unless it is certain that copyright has expired. According to Informant E, as a government authority, their organization acts in accordance with legislation, but laws do not provide ethical guidance. This perception aligns with the argument made by legal scholar Lixinsky (2022, pp. 66–67), who notes that legal frameworks seldom contain ethical guidance for heritage institutions, even though they sometimes operate in legally ambiguous areas. This in turn produce a challenge for establishing legitimacy, since a public authority have to adhere to legal norms to not lose its legitimacy (Scott, 2001, p.59).

Lastly, it should be noted that informal ethical frameworks exist in some of the studied organizations. Informant A states that they do not publish images of deceased persons. Informants B and D emphasize that they avoid including pictures of nude children in their collections. According to Informant D, most archivists have a "spinal reflex" to avoid such depictions, while Informant B acknowledges that such material might appear during mass digitization but is removed once identified. This may suggest that mass digitization may prompt critical evaluation when the material is published, thereby blurring the distinction between mass and critical digitization described by

Dahlström et al. (2012). Informant F adds that their organization sometimes excludes indecent songs from digital publication, instead only publishing the metadata to indicate the song's existence in the physical archive. These practices can be interpreted as part of the cultural-cognitive pillar, where certain types of material are initially perceived as too sensitive or controversial to be published by the informants.

Material and metadata mirroring older social orders

The previous section described the frameworks guiding decisions on what to publish online, but even if a record can be published within these frameworks, it may still raise ethical concerns. However, informant E emphasizes that sensitive material remains historically important:

> *"What I see in that material, which is a bit more difficult and which I don't want to publish just any way, is still an important part of our history. It is also important to show that it hasn't always been rosy as one might sometimes think. […] It is a balancing act between protecting personal integrity and creating security for the people in the material, while also showing what it was like for a young girl out on a summer pasture. Or a boy, for that matter. So, there it is. It must be someone else's task, not ours as an archive."* (informant E)

This quote reflects the traditional view of the archivist as a neutral custodian, aiming to present history without moral judgement (Cook, 2011, p. 608), illustrating a normative element of the profession. However, Informant F notes that maintaining neutrality can be challenging when creating digital collections, for instance with photographs of deceased persons. In such cases, the archivists at the organization together assess whether the images contribute to cultural understanding and therefore have historical value that justifies publication. Informant F's account illustrates how the two normative elements—respecting individuals' integrity while acting as a neutral custodian—are negotiated and reinterpreted within an organizational context.

Informant C describes how publishing material related to racial biology sparked public debate. The material had been digitized because it was frequently requested and at risk of deterioration, but public criticism of re-enforcing past injustices led to its removal from the website. It is now accessible only on a computer at the library, which the informant finds problematic, as many users cannot visit in person. Informant C therefore hopes to in the future create a middle ground by enabling online access through a user profile that would monitor use. This case may illustrate how the archivist's normative role as custodian came into conflict with cultural assumptions of those in the material is needing protection. It is also noteworthy that the proposed solution further separates the individuals in the records from the majority, reinforcing their status as "victims" and as distinct from the social majority (Thor Tureby & Wagrell, 2020, pp.112–113). This choice may, however, be seen as a crisis management where the organization risks being seen as what Scott describes "illegit" by reproducing old power structures even though the public does not consent.

Another example of a collection that might effect society is described by informant D:

> *"For example, I thought we should thoroughly search our print collections for all the Nazi publications produced in Sweden during the war years. There is a significant gap, both among the public and researchers, regarding how much support the Nazis had. They basically printed death lists of well-known cultural figures and propaganda. In one way, it would be an incredible service to publish this and make it publicly known. But then, as*

> *someone pointed out, it is also material that could unfortunately be misused by right-wing extremists today."* (informant D)

Here the informant stated that, although one could argue that publishing the material would shed light on darker aspects of history, there remains a significant risk of misuse. Informant D therefore had to assess the material and make a calculation of the potential risks, leading to its exclusion. It may here be noted that increasing democracy is often highlighted as one of the key aims with increasing access to cultural heritage (Thor Tureby & Wagrell, 2022, p. 347). This may further explain why the informant ultimately decided not to digitize the material, as doing so could undermine the normative value-dimension of digitalization, which is to increase democracy; a goal that right-wing extremist use undermines.

The municipal organizations operate under different directives than the university libraries and have to adapt to the supervising municipal board. Informant A's web portal is placed under the educational board, which requires that published material aligns with the school curriculum. As a result, the portal has published material that contains denigratory descriptions of minorities, since this supports critical classroom discussions. The informant explains that they have published material in which Romani people are described as "gypsies". However, the author of the record is anonymized, and a disclaimer contextualizes the terminology. According to Mindel (2022, p.547), anonymization can be controversial from an archival perspective because it affects the integrity of the material. That the organization nonetheless adopts this approach may reflect a regulative element. This since their placement under the educational board requires them to prioritize educational over evidential values.

Informant B explains that their organization avoids publishing material from city festivals in the 1980s and 1990s because these often contain stereotypical portrayals of non-Swedish cultures. When racialized representations appear, such as black face during theater performances, only a limited selection is made available. According to the informant, such images are considered to have "low historical value," and researchers who need further access can contact the archive directly. This may be understood as a normative judgement shaped by professional as well as cultural values of what constitutes legitimate historical representation. On one hand, professional norms stress the importance of documenting uncomfortable aspects of the past (Mindel, 2022, p.547). On the other hand, the organization must maintain public legitimacy and avoid being perceived as reproducing racist imagery. In connection to Scott's terms, reflects a negotiation between normative elements of archival professionalism and cultural-cognitive assumptions about public expectations, where external legitimacy concerns shape what is considered publishable.

Metadata as resource and problem

Even when material is not fully included in a digital collection, its metadata is sometimes added to alert users to its existence. However, metadata can be sensitive when it contains denigratory terms (Beezley & Scheele–Clark, 2024, p. 20). Organizations handle this differently. Informant E leaves the metadata unchanged but adds a disclaimer on a separate page of the database. Informant F instead uses quotation marks in archival descriptions to signal sensitive terms. Informant D notes that negative wording can appear in titles, preventing changes. This is further developed by Informant C who explains that eugenic material is not altered as changing titles, such as "crossbreds", would be reclassifying the records' subjects. At the same time, they sometimes add new metadata, like "Romani", to make the material findable for modern users. These practices reflect a cultural-cognitive element: organizations recognize that potentially racist material needs careful handling. As a result, it seems that they developed strategies, such as disclaimers or quotation marks, while still adhering to professional norms by preserving original titles.

In 2025, Klareld et al. (p. 40) noted that the archivist profession increasingly involves public interaction, which becomes evident in the informants' discussions of metadata. One informant describes correcting information about a person in the database after being contacted by a member of the public regarding the individual's past as an ultra-right politician. Similarly, informant F regularly receives public suggestions for correcting inaccurate metadata. Informant B explains that their organization employs young people to photograph local environments and maintains contact with former employees of the town's largest factory, who provide relevant metadata for related images. These practices may differ from Caswell and Cifor's (2019) vision of digital collections as part of an ongoing process between archives and communities. However, they can be seen as organizational adaptations to shifting professional norms and, in informant B's case, as responses to municipal expectations to document the entire city.

Material that violates privacy

The previous section discussed the risk of reproducing social injustices, but the informants also emphasize the need to protect individuals represented in the records. This relates to a cultural-cognitive element, as digitization has intensified debates on personal integrity (SOU, 2019:58, p. 119). Informant A notes that the web portal sometimes publishes material revealing identities in records concerning crime, mental health, and similar issues. Editors therefore differentiate between public and private individuals, redacting the names of the latter. This regulative practice may be shaped by cultural-cognitive assumptions in Swedish media discourse, where public figures are considered to have reduced privacy protection due to public interest (Journalistförbundet, 2025). The informant also claims that identifiable cases never have resulted in complaints; instead, users often express appreciation for locating relatives in the collections. This echoes Tai et al.'s (2019, p. 14) study which shows how archival records can help individuals situate themselves within broader histories.

Informant B states that they mainly upload material from the 1960s and earlier, as this avoids many privacy challenges. Their starting point is therefore primarily regulative rather than normative. The informant also notes that photographs taken by youths employed for the collection are sometimes removed upon request, while the metadata is kept. This shows a normative conflict. On one hand, the public might expect the images to remain, since they were produced within municipal employment. On the other hand, recent ethical work stresses the importance of accepting "no" as an answer (Caswell & Cifor, 2019, p. 169). Keeping the metadata can therefore be seen as a compromise between these competing norms.

Informant E explains that their traditions archive anonymizes material even when the subjects are deceased, out of respect for living relatives who may feel exposed. This conflicts with Mindel's (2022, p. 547) view of integrity since the digital copy is altered, but the organization manages this by being transparent and providing on-site access to the unedited record. Informant F also notes that privacy concerns restrict what they publish with interviews often being withheld because they may contain sensitive personal or third-party information. Privacy concerns are also raised by Informant D, who sometimes encounters sensitive content in personal archives. Such material may initially be published through mass digitization but is removed once identified. This suggests that mass digitization, as described by Dahlström et al. (2012, p. 457), may trigger subsequent rounds of qualitative appraisal. Across discussions of privacy, the organizations have developed internal regulative policies to handle material deemed invasive with care. This aligns not only with external frameworks, such as GDPR, but also with cultural-cognitive elements, reflecting privacy's longstanding significance and its heightened relevance in digitization debates (SOU, 2019:58, p. 119).

An interesting notion is, however, that it is not always clear when to protect privacy. This, for instance is raised by informant D:

> *"It became apparent that the wrong names were redacted. This is because the interviews conducted with Holocaust survivors in 1945/46 were not sensitive for the survivors, but for some of the perpetrators mentioned by name. And those names had not been redacted. So, okay, that was a bit strange, and at the same time, the fact that they redacted the names of the informants drastically reduced the research value, because you couldn't identify who said what, and the whole point of the documentation project was that the interviews were supposed to be true to reality and follow the same format so they could be compared and support each other. It became impossible to work with them in that way when you didn't have the names of the informants*." (informant D)

The names of the perpetrators were the most sensitive, as they were accused of war crimes, which could constitute slander under Swedish law. Many informants were also quite young when interviewed, and data protection legislation conflicted with prior consent for publication. The university library consulted legal advisors, who concluded that the research interest outweighed the need to protect informants and perpetrators, allowing publication with names included. According to the informant, survivors appreciated this decision and viewed it as a form of historical redemption. From a theoretical perspective, this choice reflects a regulative element enabling publication (research interest) while being constrained by other legal frameworks (GDPR and slander laws). Informant D notes that anonymization can be counterproductive, as "a public authority, in their effort not to offend anyone, actually offends them by blacking out their names." This aligns with Thor Tureby and Wagrell (2020, pp.112–113), who argue that anonymization may perpetuate the marginalization of minorities.

Discussion

At the studied sites, digitization appears to be shaped not only by ethical decisions, but also by organizational factors. Municipal organizations differ by board: Informant B aims to create a city-wide collection, while Informant A must align material with the school curriculum. University libraries and traditional archives follow classic archival norms, prioritizing preservation of at-risk materials. It thus seems that selection for digitization reflects organizational goals rather than external funding. One exception is digitalization on demand at university libraries. However, the material gets publicly accessible, thereby circumnavigating the risk highlighted by Manžuch (2017, p.6). This also reflect the normative principle of access in Swedish archival legislation (SFS 1990:782 §3).

GDPR and copyright act as external regulative elements across all sites. Privacy concerns predated GDPR (Skott, 2017), but the regulation has strengthened protective measures. Informant A uses anonymization, which may compromise records' integrity (Mindel, 2022, p.457), reflecting a prioritization of educational over evidential value. Informant E also applies anonymization but redirects users to the physical copy on site. Other organizations generally avoid publishing "too recent" or identifiable material, employing risk-minimization strategies to estimate whether subjects are likely deceased, illustrating Scott's (2001, pp. 54–55) point that legislation often serves as a starting point for discussion rather than rigid compliance.

According to the informants, material containing identifiable individuals rarely causes complaints and can provide value for users locating relatives. This is for instance exampled by holocaust records that according to informant D were seen as contributing to historical redemption (Informant D), aligning

with Tai et al. (2019) on records' performative role and Thor Tureby & Wagrell (2020, pp.112–113) on anonymization potentially marginalizing.

The informants have also faced material challenging cultural-cognition elements, including images of deceased, nude children, and metadata with negative wording. Approaches vary: some material is reflexively excluded, others prompt internal deliberation. Racialized material is often published "as is," justified by educational value or archival conventions. However, informant E utilizes contextualization, echoing Peterson Brink et al. (2016) and informant F quotation marks as form of disclaimer. Despite previous research, such as Caswell and Cifor (2019) and Alkemade et. al. (2023) none of the informants in this study reported using community engagement in collection creation, though other Swedish institutions may follow different practices. Yet, one informant states that public criticism has led to the removal of a eugenic collection. This illustrates cultural-cognitive elements on vulnerability and what is appropriate to publish in digital collections, which if violated might lead to delegitimization (Scott, 2001, pp. 57–58).

Finally, regarding generalization and future research, the study's small sample limits broad conclusions. Nevertheless, consistent patterns suggest avenues for further study on ethical issues, especially in countries without Sweden's tradition of public access. Such research might also be essential given that efforts like The Cultural Heritage Cloud, which aims to create digital infrastructure across European cultural heritage organizations, may lead to international linking of collections. This might in turn create a need for contextualization of digital collections (European Commission, n.d.).

Conclusion:

This study shows how ethical and organizational aspects shape digital collections at six Swedish heritage sites. The main empirical findings are listed in table 1.2. These findings indicate that digitization practices vary notably across the studied organizations. Their internal organizational goals appear to be a central factor shaping selection and handling of material. Legislative requirements also influence decision-making, at times creating tensions with professional archival norms and serving as reference points in ambiguous cases. Future research should examine how such norms and legal frameworks differ internationally, as contextual factors are crucial for understanding and interpreting digital collections.

Table 1.2 main conclusions

Aspect	Informant A (Educational municipal board)	Informant B (Cultural municipal board)	Informant C (University library)	informant D (University library)	Informant E (Traditions and folklore archive)	Informant F (Traditions and folklore archive)
Economic conditions and effect of organizational context	In-house budget. Prioritize material relevant for school curriculums.	In-house budget. Aims to create collections relevant to all the city	In-house budget, digitization on demand. Mainly prioritize material risking decaying. Prioritize material requested of researchers.	In-house budget, digitization on demand. Mainly prioritize material risking decaying. Prioritize material requested of researchers.	In-house budget. Mainly prioritize material risking decaying.	In-house budget, digitization positive side-outcome of other projects. Mainly prioritize material risking decaying.
Copyright	Not discussed	No owns their material.	avoid publishing too young material	Avoid publishing too young material	Not discussed	Avoid publishing too young material
GDPR	Utilize anonymization.	Avoid publishing identifiable material	Avoid publishing identifiable material	Avoid publishing identifiable material	Avoid publishing identifiable material	Does not publish material before 1970's
Examples of material deemed unpublishable and consequences	Deceased persons, policy against publishing	Nude children, policy against publishing	Eugenics collection, removed after public complaints	Nazi magazines not published.	Material too exposing. Avoid publishing but highlights metadata	Indecent songs. Highlighted by metadata.
Material that is sensitive due to risk of reproducing negative social structures incl. metadata	Published due to educational value, authors anonymized.	Small samples published, refers to physical collections	Publish as is.	Publish as is.	Publish as is. Utilize contextualization and disclaimers.	Publish as is, negative wording in metadata in quotation marks. Internal discussions.
Material violating privacy	Publish, nonpublic figures anonymized	Publish, but might remove at request	Not developed	Case to case decision based on research value	Avoids publishing material to young. Publish metadata to direct to physical archives. Anonymization if necessary	Avoids publishing material. Publish metadata to direct to physical archives.

Interviews

1. Web-editor of a heritage platform. (2024, May 27). Personal communication.

2. Archivist, university library. (2024, May 31). Personal communication.

3. Head of department at a language and folklore archive. (2024, June 4). Personal communication.

4. Head of digitization at a university library. (2024, June 12). Personal communication.

5. Picture archivist, municipal platform. (2025, June 11). Personal communication.

6. Music archivist at a traditions archive. (2025, June 11). Personal communication.

References

Adams, W. C. (2015). Conducting semi-structured interviews. In K. E. Newcomer, H. P. Hatry, & J. S. Wholey (Eds.), *Handbook of practical program evaluation* (pp. 492–506). Wiley. https://doi.org/10.1002/9781119171386.ch19

Alkemade, H., Claeyssens, S., Colavizza, G., Freire, N., Lehmann, J., Neudecker, C., Osti, G., & van Strien, D. (2023). Datasheets for digital cultural heritage datasets. *Journal of Open Humanities Data, 9*, 17. https://doi.org/10.5334/johd.124

Baca, M. (Ed.). (2016). *Introduction to metadata*, 3rd ed. Getty Publications. https://www.getty.edu/publications/intrometadata/introduction/

Beezley, S., & Scheele Clark, H. (2024). When catalogers and archivists come together: The creation of the Metadata Justice in Oklahoma Libraries & Archives Symposium. *Collaborative Librarianship, 14*(1), Article 4. https://digitalcommons.du.edu/collaborativelibrarianship/vol14/iss1/4

Caswell, M., & Cifor, M. (2019). Neither a beginning nor an end: Applying an ethics of care to digital archival collections. In M. Caswell, & M. Cifor (Eds.), *The Routledge handbook of new digital practices in galleries, libraries, archives, museums and heritage sites* Routledge.

Cook, T. (2011). The archive(s) is a foreign country: Historians, archivists, and the changing archival landscape. *The American Archivist, 74*(2), 600–632. https://doi.org/10.17723/aarc.74.2.xm04573740262424

Dahlström, M., Hansson, J., & Kjellman, U. (2012). 'As we may digitize'—Institutions and documents reconfigured. *LIBER Quarterly, 21*(3–4). https://doi.org/10.18352/lq.8036

Dalgeish, P. (2011). The thorniest area: Making collections accessible online while respecting individual and community sensitivities. *Archives and Manuscripts, 39*(2), 71–82.

Denscombe, M. (2010). *The good research guide for small-scale social research projects*, 4th ed. Open University Press.

European Commission. (n.d.). *The Cultural Heritage Cloud*. https://research-and-innovation.ec.europa.eu/research-area/social-sciences-and-humanities/cultural-heritage-and-cultural-and-creative-industries-ccis/cultural-heritage-cloud_en#what-is-the-cultural-heritage-cloud

GLAM Labs Community. (n.d.). *Projects*. https://www.glamlabs.io/projects
Integritetsskydsmyndigheten. *Personuppgifter.* https://www.imy.se/verksamhet/dataskydd/det-har-galler-enligt-gdpr/introduktion-till-gdpr/personuppgifter/

Guest, G., Namey, E., & Chen, M. (2020). A simple method to assess and report thematic saturation in qualitative research. *PLOS ONE, 15*(5), e0232076. https://doi.org/10.1371/journal.pone.0232076

Journalistförbundet. (2025, December 8). *Publicitetsregler*. https://www.sjf.se/yrkesfragor/yrkesetik/spelregler-press-radio-och-tv/publicitetsregler

Klareld, A., Pihl Skoog, E., & Söderström, K. (2025). Managing the past and the future: Evolving professional roles of archivists in contemporary Sweden. *Nordic Journal of Library and Information Studies, 5*(1), 29–48. https://doi.org/10.7146/njlis.v5i1.146726

Legner, C., Eymann, T., Hess, T., et al. (2017). Digitalization: Opportunity and challenge for the business and information systems engineering community. *Business & Information Systems Engineering, 59*, 301–308. https://doi.org/10.1007/s12599-017-0484-2

Lixinski, L. (2022). Digital heritage surrogates, decolonization, and international law: Restitution, control, and the creation of value as reparations and emancipation. *Santander Art and Culture Law Review, 6*(2), 65–86.

Mindel, D. (2022). Ethics and digital collections: A selective overview of evolving complexities. *Journal of Documentation, 78*(3), 552–569. https://doi.org/10.1108/JD-11-2020-0193

Peterson Brink, M., Ducey, E., & Lorang, E. (2016). The case of the Awgwan: Considering ethics of digitization and access for archives. *The Reading Room, 2*, 21–22. https://doi.org/10.69772/trrajsc.13

Riksrevisionen. (2022). *Räkna med mindre – den årliga omräkningen av myndigheternas anslag*. https://www.riksrevisionen.se/granskningar/granskningsrapporter/2022/rakna-med-mindre---den-arliga-omrakningen-av-myndigheternas-anslag.html

Schildt, H. (2022). The institutional logic of digitalization. In T. Gegenhuber, D. Logue, C. R. Hinnings, & M. Barrett (Eds.), *Digital transformation and institutional theory*. Emerald.

Schwartz, J., & Cook, T. (2002). Archives, records, and power: The making of modern memory. *Archival Science, 2*, 1–19 https://doi.org/10.1007/BF02435628

Scott, W. R. (2001). *Institutions and organizations,* 2nd ed.. Sage.

Skott, F. (2014). Finns den på nätet? Nätpublicering som kulturarvsproduktion. *Tidskrift för kulturforskning, 13*(3), 55–60.

Tai, J., Zavala, J., Gabiola, J., Brilmyer, G., & Caswell, M. (2019). Summoning the ghosts: Records as agents in community archives. *Journal of Contemporary Archival Studies, 6*, Article 18. https://elischolar.library.yale.edu/jcas/vol6/iss1/18

Thomassen, T. (2001). A first introduction to archival science. *Archival Science, 1*, 1–19. https://doi.org/10.1023/A:1011523616841

Thor Tureby, M., & Wagrell, K. (2020). Digitization, vulnerability, and Holocaust collections. *Santander Art and Culture Law Review, 2*, 119–138.

Thor Tureby, M., & Wagrell, K. (2022). Crisis documentation and oral history: Problematizing collecting and preserving practices in a digital world. *The Oral History Review, 49*(2), 347–366. https://doi.org/10.1080/00940798.2022.2101933

Utbildningssidan.se. (n.d.). Information about the labour market of archivists in Sweden. https://utbildningssidan.se/.../arkivarie

Zhang, J. (2012). Archival context, digital content, and the ethics of digital archival representation. *Knowledge Organization, 39*(5), 332–339. https://doi.org/10.5771/0943-7444-2012-5-332