



Title	Report on the Survey “Digitization and Artificial Intelligence for Archives and Documentary Heritage Materials”
Working group code	RA03
Study title	AI-Assisted Digitization of Archives and Documentary Heritage Materials
Status	Public
Version	Final
Writers	Jessica Bushey, Marina De Souza, Kailey Fukushima, David Iglésias, Marta Riess, Eng Sengsavang, Hrvoje Stancic, Zeljko Trbusic
Date	27 March 2025



Social Sciences and Humanities
Research Council of Canada

Conseil de recherches en
sciences humaines du Canada

Canada

This work is licensed under [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)



Citation: Jessica Bushey, Marina De Souza, Kailey Fukushima, David Iglésias, Marta Riess, Eng Sengsavang, Hrvoje Stancic, Zeljko Trbusic, *Report on the Survey “Digitization and Artificial Intelligence for Archives and Documentary Heritage Materials,”* InterPARES Trust AI, May 2025, https://interparestrustai.org/assets/public/dissemination/RA03-InterPARESAI-Survey_Report_FI_NAL.pdf

Report on the Survey
“Digitization and Artificial Intelligence for
Archives and Documentary Heritage Materials”
InterPARES Trust AI

Table of Contents

1. Introduction	3
2. Study Background	3
3. Survey Description and Methodology	5
3.1. Description	5
3.2. Methodology	6
4. Summary of Key Findings by Research Questions	8
4.1. Demographics of Survey Respondents	8
4.2. Research Question 1: Key Activities and Best Practices of Digitization Projects	9
4.3. Research Question 2: AI Tools Used in Digitization Projects	12
4.4. Research Question 3: Benefits, Risks, Limitations, and Potential Biases of AI in Digitization Projects	13
4.5. Research Question 4: Future Perspectives for Uses of AI Tools in Digitization Projects	15
5. Survey: Question-by-Question Responses	16
5.1. Section 1: General Information	17
5.2. Section 2: Digitization Activities	24
5.3. Section 3: Digitization Context	41
5.4. Section 4: Digitization and AI	67
6. Conclusion	86
Annex A Dissemination List	88

1. Introduction

The survey *Digitization and Artificial Intelligence for Archives and Documentary Heritage Materials* measures the state of digitization within organizations and assesses the impact of artificial intelligence (AI) on digitization activities. The survey is part of the study “AI-Assisted Digitization of Documentary Heritage Materials.” The broader study is conducted as part of [InterPARES Trust AI](#) (2021-2026), an international, multi-disciplinary research project with over 100 partner institutions and individual collaborators around the world. InterPARES Trust AI is directed by Luciana Duranti and Muhammad Abdul-Mageed of the University of British Columbia, and coordinated by Corinne Rogers. It is funded by the Social Sciences and Humanities Research Council (SSHRC) of Canada.

[Back to Table of Contents](#)

2. Study Background

The main purpose of the broader research study, “[AI-Assisted Digitization of Documentary Heritage Materials](#),” of which the survey forms an integral part, is to investigate the potential benefits, challenges, and risks of using AI during digitization of archives and documentary heritage materials. The study seeks to explore the main ways that AI tools may support digitization processes,¹ and is guided by ethical approaches to the uses and development of AI, as represented in the [UNESCO Recommendation on the Ethics of AI](#).

Digitization refers to the process of creating a digital representation of a physical artefact and/or its associated information. In the study, we focus on digitization of archives - records safeguarded for as long as possible because of their historical, informational, and/or cultural value - with the purpose of creating faithful digital reproductions² of the originals for long-term preservation. Archives hold various types of physical objects, including textual materials, photographs, films, videos, sound recordings, and three-dimensional artefacts (such as albums, books, or other objects). Different processes and tools are used for digitization, including imaging (scanning or photographing), or signal conversion from analogue to digital, in the case of analogue audiovisual media. The result is a digital surrogate of the original physical object, and its associated metadata, which can be managed, stored, published, and disseminated using

¹ The term “documentary heritage materials” refers to artefacts in any format or medium deemed important to safeguard over the long term and kept by an organization, a private individual, or other entity.

² Federal Agencies Digital Guidelines Initiatives (FADGI), Still Image Working Group, 2023. *Technical Guidelines for Digitizing Cultural Heritage Materials: Third Edition*, p. iv. Accessed 07 January 2023, https://www.digitizationguidelines.gov/guidelines/FADGI%20Technical%20Guidelines%20for%20Digitizing%20Cultural%20Heritage%20Materials_3rd%20Edition_05092023.pdf.

software and digital infrastructures. Converting analogue signals to digital signals is a definitive step in the digitization process, and what is most often meant by “digitization.”³ However, digitization involves other significant activities before and after the digital conversion process, which form an integral part of digitization processes. Therefore, by “digitization” we mean not just the moment of digital conversion, but the full suite of activities undertaken during pre-digitization, digitization, and post-digitization processes.

The study research questions include the following:

- What are key activities and best practices carried out in effective digitization of archives and documentary heritage materials?
- What AI tools are currently being used/developed by practitioners and vendors for digitization activities?
- What are the benefits and risks, limitations and potential biases when using AI technologies for digitization?
- What AI-based tools might be developed in the future, particularly to assist archival-related activities during digitization processes, such as metadata creation for archival description?

The study team is comprised of an international group of professionals and scholars, including the following members:

- Shadreck Bayane, University of South Africa
- Jessica Bushey, San José State University (Advisor)
- Marina De Souza, University of British Columbia
- Kailey Fukushima, University of British Columbia
- David Iglésias, City of Girona
- Tomislav Ivanjko, University of Zagreb
- Adam Jansen, Hawaii State Archives
- Petra Lovric (Graduate Academic Assistant)
- Marta Riess, International Atomic Energy Agency
- Eng Sengsavang, UNESCO Archives (Study Lead)
- Hrvoje Stancic, University of Zagreb
- Zeljko Trbusic, University of Zagreb
- Goran Zlodi, University of Zagreb

[Back to Table of Contents](#)

³ Digitization should not be confused with “digitalization,” the process of adopting digital technologies to automate or update business processes, also known colloquially as “going digital.”

3. Survey Description and Methodology

3.1. Description

The survey *Digitization and Artificial Intelligence for Archives and Documentary Heritage Materials* was published in English using the online survey tool Jotform and was open to respondents from 01 March to 15 May 2023. It was disseminated via approximately 66 professional listservs and forums, mostly within the GLAM (Galleries, Libraries, Archives, and Museums) communities, but also, to a much lesser degree, within some AI and digitization domains (see Annex A for a complete dissemination list). The survey was open to professionals working in institutions with archival and documentary heritage collections. Prior experience with AI or digitization was not required to participate in the survey.

The objectives of the survey were threefold:

- 1) To understand the current state of digitization within organizations;
- 2) To understand whether organizations are using AI in digitization projects, and if so, how;
- 3) To understand the barriers and drivers for organizations in relation to digitization and AI.

The survey consisted of four sections, with a total of 50 questions and an estimated duration of 20 minutes to complete. All responses were anonymized and analyzed in aggregate form. Each of the four sections focused on a particular theme:

- **Section 1, “General Information”** (Questions 1 to 6) – Demographic information about respondents and the organizations at which they are employed;
- **Section 2, “Digitization Activities”** (Questions 7 to 23) – Information about the digitization activities taking place at respondents' organizations. The definition of digitization in the Study Background of this document was provided to respondents;
- **Section 3, “Digitization Context”** (Questions 24 to 36) – Information about how respondents plan, develop, and carry out digitization projects at their organizations;
- **Section 4, “Digitization and AI”** (Questions 37 to 50) – Information about the current and/or potential role of AI in respondents' digitization projects or activities. The following definition of AI was provided to respondents in this section: “Artificial intelligence (AI)

refers to “systems which have the capacity to process data and information in a way that resembles intelligent behavior, and typically includes aspects of intelligence, learning, perception, prediction, planning or control.”⁴ AI models, methods, tools, and algorithms include those developed in the AI subfields of machine learning and deep learning. Machine learning models may include text classification, natural language processing, named entity recognition, topic modeling, computer vision, or machine translation, to name a few. In real terms, AI-supported tasks may include handwritten text analysis, automatic translation, automated speech-to-text, image captioning, or more.

[Back to Table of Contents](#)

3.2. Methodology

We adopted survey research methodology and designed a self-completion questionnaire as our primary information gathering instrument. Web-based, self-completion questionnaires afford larger sample sizes (‘n’) and a broader geographic reach than other common survey methods, e.g., structured interviews and focus groups. Our parent project, InterPARES Trust AI, and our study’s research questions are international in scope. As such, this method and instrument aligned with our study objectives and contexts of research.⁵

The research team worked collaboratively and iteratively to compile the research questions. Following internal testing of the questionnaire, we initiated an open testing and feedback round with colleagues at InterPARES Trust AI.⁶ These reviewers were part of our parent project but not involved with our study; they suggested revisions to existing questions as well as additions, deletions, and question groupings. After reviewing and incorporating this feedback, we transferred the questionnaire from a textual document to the online form builder Jotform. The completed Jotform questionnaire was subject to additional rounds of internal testing before we distributed it to pre-selected professional listservs.

The final questionnaire featured a mix of quantitative and qualitative questions. Quantitative questions included ordered/unordered closed responses, dichotomous responses, scaled responses, and ranked responses. Quantitative questions often included an “Other”

⁴ UNESCO, Recommendation on the Ethics of Artificial Intelligence, 2021, p. 10.

⁵ Despite the international scope of our study and the multilingual composition of our research team, we only distributed the questionnaire in English. Our team collaborates in English and we anticipated completing the survey analysis in this same language. We recognize that this decision limits the international reach of our research instruments and the global composition of our sample.

⁶ The InterPARES Trust AI feedback round was not anonymous.

option with a free-form text field for respondents to describe their responses, providing an opportunity to give a qualitative response. By contrast, qualitative questions provided only a free-form text response field for participants to fill in themselves. In addition, the survey included several conditional questions, in which an answer “yes” would direct the respondent to a related sub-question, while an answer “no” would move the respondent on to the next question, thereby skipping the sub-question. For this reason, the total number of responses submitted by each participant varies depending on how they answered the conditional questions. The total number of possible questions ranged from 38 to 58, with 20 questions only appearing if respondents answered “yes” to conditional questions. There were 38 quantitative questions and 12 to 20 qualitative questions.⁷

In total, 219 survey responses were submitted. Two submissions were entirely blank. Thus in our analysis, we considered the responses of 217 participants. While the 217 participants responded to the majority of questions, some participants left blank responses to some questions. In addition, as explained above, some questions were conditional and would only appear if a participant responded “yes” to the previous question. Finally, many questions allowed for multiple responses. Therefore, the number of responses to any given question varies depending on the number of non-blank and conditional responses received, and whether multiple responses were allowed for a given question.

In our analysis of the qualitative survey responses, we used both deductive and inductive approaches. In the initial analysis of the survey responses, study team members used an inductive or “open coding” approach to observe patterns and themes emerging from the data, rather than approaching the data with predetermined themes or coding categories. In a further stage of analysis, study team members used the deductive approach to analyze responses across multiple questions, with the goal of identifying patterns that respond directly to the study research questions.

[Back to Table of Contents](#)

⁷ Respondents’ access to qualitative questions 25b, 26b, 30b, 31b, 32b, 47b, and 48b is contingent upon their responses to questions 25, 26, 30, 31, 32, 47, and 48. This calculation of quantitative and qualitative questions counts question 50 in both categories. Question 50 is a two-part contingency question. First, it asks whether the respondent consents to being contacted for a follow-up interview (quantitative response); if the respondent selects “Yes,” the form provides them with three free-form text fields to record their first name, last name, and email (qualitative response).

4. Summary of Key Findings by Research Questions

This section presents a summary of the demographics of survey respondents, followed by key findings from the survey structured on the four main research questions outlined in Section 2 - Study Background. Question-by-question survey responses are presented in Section 5. Overall, the findings offer a picture of, on the one hand, the state of digitization activities within organizations, and on the other hand, the uses, perceived benefits, limitations, and risks of applying AI tools when digitizing documentary heritage materials.

4.1. Demographics of Survey Respondents

The demographics of survey respondents reflect the working contexts of our research study and reveal an anticipated limitation to the scope of our data collection and analysis. As indicated in Section 3.2 - Methodology, our research team distributed the questionnaire in English. Although our team is international and multilingual, English is the language in which our team collaborates and is also the primary working language of InterPARES Trust AI. We acknowledge that the monolingual composition of this questionnaire limits the global composition of our sample and the generalizability of our data on an international scale. Our research team's geographic locations and professional networks also affect the global distribution of responses. InterPARES Trust AI is an international project based in Canada and the majority of this questionnaire's designers live and work in either North America or Europe. Combined, these linguistic and geographic professional contexts contribute to an overrepresentation of survey responses from North America and Europe, and an underrepresentation of survey responses from people living and/or working in Asia and the global South.

As seen in the responses to question 4 of the survey, the most represented type of professional amongst the respondents are archivists, including audiovisual and digital archivists (31.3%). This is likely due to the type of listservs targeted during dissemination of the survey (see Annex A for the full dissemination list), as well as the framing of the survey focused on archival rather than library materials. Librarians are, nonetheless, the second most represented professional at 17.6%, meaning that altogether, archivists and librarians make up 55.6% of respondents. Digitization specialists, records managers, collections management professionals, and digital engagement/initiatives staff made up the next most represented groups (24.8%) - almost a quarter of participants. Notably, these top six specializations are engaged in the

management of information assets or cultural goods, or in the creation and/or management of digital/digitized assets. Additionally, almost a quarter of respondents (24%) hold either leadership (director, head) or managerial positions. Participants most frequently worked for their organizations for fewer than 5 years, between 5-10 years, or for more than 20 years, for a total of 70.5% across the three categories, with each category receiving an almost equal number of responses.

Participants most frequently worked for public/governmental or private institutions (81.1% total), when compared to non-governmental, intergovernmental, public-private, religious organizations, or other types of organizations, which received far fewer responses. 64% of respondents worked for organizations in either education/research, culture, or memory institutions. However, 11 other types of sectors are represented to a lesser degree, in addition to sectors provided by respondents in the “Other” category, accounting for a significantly diverse representation of sectors in the remaining 36% of responses. More than half of respondents worked for organizations with either 50 or fewer people, or between 100 to 499 people (54.4% total).

[Back to Table of Contents](#)

4.2. Research Question 1: Key Activities and Best Practices of Digitization Projects

The survey responses reveal that digitization is actively practiced within organizations, with an overwhelming 92.2% of respondents affirming that their organizations have engaged in digitization programs or projects in the last five years. Not only is digitization widely practiced, it is highly valued amongst survey respondents, with nearly all participants (94.4%) agreeing that digitization is either a very important or important activity for preservation of and access to documentary heritage materials. The primary motivators for digitization reflect core archival goals and values: increasing external access to collections was the top motivator (18.6%), followed by enhancing visibility and outreach (15.7%), preserving originals (15.5%), and improving internal access (15%). These motivators reflect what participants considered to be the greatest successes of their digitization projects/programs, including increased access to collections (31.5%), enhanced visibility and outreach (11.3%), and the volume or scale of digitized materials (10.9%).

Despite this high rate of digitization and recognition of its importance and positive impact, funding sources for digitization programs and projects vary significantly amongst organizations.

More than one-third of respondents (36.5%) rely on regular funding from their organizations, while 22.2% rely on funding from external public sources, and another 19.3% on special project funds. Private funding sources were also cited (12.9%). Although digitization is undertaken in most organizations, many digitization programs and projects are not funded from organizations' regular budgets. This was echoed in responses regarding the greatest challenges in digitization projects/programs. Respondents identified inadequate funding and resources, including equipment, space, and time, as the greatest challenges (21.8%). Many respondents expressed frustrations over the lack of continuous funding, reliance on grants, and the time-intensive nature of digitization. Staffing issues (16.1%), such as shortages, retention problems, lack of expertise, and work overload, were also significant challenges. Other challenges included issues related to program/project management (10.9%), IT infrastructure (9.5%), digital preservation and sustainability (8.4%), and institutional support (8.1%). Respondents noted a lack of institutional understanding of the resource demands of digitization and the importance of long-term sustainability.

Most participants' organizations have digitized less than 25% of their holdings (63.6%), compared with those that have digitized between 25% to 74% of their collections (16.7%), and a minority that have digitized between 76% to 100% of their holdings (7.1%). Digitization also produces mass amounts of digital files, which can be measured by volume in bytes. One-third of respondents stated that their organizations have produced less than 50 terabytes (TB) of digitized materials. Another one-third of organizations have produced between 51 TB to 999 TB, while 11% of organizations have produced between 1-10 petabytes (PT) of digitized materials. The volume of digitized files accounted for in this survey alone is between 97 to 249 PT (petabytes). These figures point to the mass digital infrastructures and storage space required for digitization efforts, as well as the environmental impact of digitization, an area that requires further study.

More than half of respondents' organizations practice digitization as an ongoing/continuous activity (52.3%), rather than as discrete projects. When practiced as defined projects, the duration of digitization projects is most often two years or less (34.8%). Organizations are digitizing collections across various formats, but most often textual documents, including bound publications and volumes (34.4%), and audiovisual materials combined - including photos, films/videos, and audio recordings - representing altogether 44.4% of materials being digitized within organizations. The results indicate that a given organization is often digitizing several formats. The content and informational value of collections is an important factor when organizations prioritize collections for digitization, whether for research

purposes (21.3%) or for value in terms of public access to information (19.6%) – each slightly ahead of concerns related to the physical state of materials (17.2%), technological obsolescence (14.4%), or age of the materials (9.3%). Remarkably, when the top two responses (digitization for research purposes and digitization for public access to information) are combined, they amount to 40.9%, the same percentage as the next three responses combined (physical condition, age, or technological obsolescence) – 40.9%. This suggests that the content/informational value of documents and the physical state of materials are considered in equal measure when selecting collections to digitize, or that organizations use some combination thereof when prioritizing collections to digitize.

The survey responses confirmed the simple but fundamental fact that digitization involves a suite of activities beyond merely scanning or converting materials to digital formats. When asked what they consider to be the major activities in digitization, respondents most frequently cited digitization and metadata management (~10% each). They also stated that selection, preparation, storage and backup, post-processing, digital preservation, and project management are important elements in digitization processes. 81.6% of respondents indicated that they consider digital preservation requirements in their digitization programs and projects. Digital preservation activities included most often the recording of technical and descriptive metadata, creation of multiple copies of digital files, ingestion into a trusted digital repository, and the creation of digital preservation policies. Further, organizations tend to keep either two or three copies of digitized files (~30% each), either on one (22%) or two (44%) different types of media. While the results are encouraging for the long-term preservation of digital records, they once again highlight the environmental impact and large-scale infrastructure required to sustain this digital ecosystem.

Over half (53.5%) of respondents reported having a digitization policy in place, while 36.4% stated they did not, and 10.1% were uncertain. These figures highlight a heterogeneous approach to digitization across organizations, with a notable portion lacking an internal policy for digitization efforts. In lieu of, or in addition to, internal digitization policies, most respondents (71%) reported following regulations, guidelines, best practices, or standards related to digitization. Participants most frequently cited guidelines directly pertaining to digitization (30.8%), followed by metadata standards (12.1%), digital preservation frameworks (9.8%), and archival laws or standards (4.2%). Specific entities and resources, such as the Federal Agencies Digital Guidelines Initiative (FADGI), the International Standards Organization (ISO), Dublin Core, and the Open Archival Information System (OAIS), were influential. Participants also cited internal regulations within their own organizations (10.7%).

Some organizations may also turn to legal or regulatory frameworks that impact their digitization activities. 44.2% of respondents stated they were aware of digitization-related requirements in their jurisdiction, while 55.8% were either unaware or uncertain. For those aware of relevant regulations, copyright law emerged as the most frequently cited type of regulation (41.6%), followed by privacy and data protection laws, such as the General Data Protection Regulation (GDPR), mentioned in 11.5% of responses. Other significant regulations included national digitization regulations and evidentiary laws related to electronic records, each cited by 6.2% of respondents. The diversity of regulatory sources and standards underscores the varying approaches to digitization practices and governance, as well as the complex web of issues relevant to digitization.

When considering the risks of digitization, 41.2% of respondents reported that their organizations had not identified specific drawbacks of digitization, while 29.6% had, and 29.2% were unsure. Sustainability concerns, such as maintenance costs and IT requirements, were the most frequently mentioned risks (27%). Copyright and data privacy issues were tied for second (15.7%), followed by access challenges (7.9%), and representational bias in selection for digitization (6.7%). Respondents noted that materials selected for digitization are highlighted over other materials, may lose context when published online, and/or may appear to represent the whole collection when they do not.

The responses indicate that digitization is a common activity within organizations and that it plays an important role in safeguarding documentary heritage. Digitization enables broader access to archival materials and strengthens the visibility of institutions and their collections. A suite of digitization policies, standards and guidelines exist within the digitization landscape, which appear to be used heterogeneously and to varying degrees across institutions. Digitization activities are often reliant on provisional funding and are subject to resource constraints. The outcome of digitization projects is both tangible, in terms of the volumes of digital documents and data produced, as well as intangible, in terms of many of the cited benefits, challenges, and risks.

[Back to Table of Contents](#)

4.3. Research Question 2: AI Tools Used in Digitization Projects

Survey results indicate that engagement with AI tools for digitization purposes was relatively low. However, many organizations are curious and some are experimenting with AI tools for digitization purposes. Almost half of respondents stated that their organizations had not

considered using AI tools during digitization projects (47.5%), while a more moderate 35.5% of organizations considered using AI tools during digitization, and 17.1% were not sure. However, in terms of actually using AI tools for digitization, a smaller percentage of respondents indicated that their organizations had done so - 12% used custom-developed AI tools, while 14.3% used off-the-shelf AI tools.

Amongst those who had employed AI tools (26.3%), they were most often used for OCR/HTR (Optical Character Recognition or Handwritten Text Recognition) (17.8%), speech-to-text transcription (12.1%), or computer vision (6.5%) during digitization projects. Topic classification and language identification were the next most frequently cited types of machine learning used (1.9% each). Participants who had used AI also described the types of digitization activities supported by AI, including most often metadata creation, image processing, writing code, and searching/filtering metadata (1.9% each). The specific tools cited included OpenAI's ChatGPT and Tesseract (3.7% each), or Transkribus (2.8%), among many others. Overall, participants named 32 types of AI tools, although many were brand names with no further reference to the AI software or tool, such as Amazon, Google, or Microsoft Azure. Respondents also named specific applications, such as Adobe Bridge, Adobe Acrobat, or Microsoft Word.

Respondents were asked to cite any off-the-shelf AI tools they were aware of, whether or not their organizations had used AI tools. Interestingly, the top response was some form of "I don't know" (17.7%). Looked at another way, 82.3% of respondents provided the name of a tool, such as ChatGPT, Transkribus, Google Bard, and ABBYY Finereader, amongst many others. However, again, some respondents simply named a platform or software application, such as Bing or Microsoft Word, without providing a specific AI tool offered by those products.

The responses indicate that a modest number - just over a quarter - of organizations had used AI tools for digitization, a figure that will no doubt increase over time. A diversity of AI tools and products are known and/or used by respondents, yet at this stage, there is still perhaps some confusion as to what existing IT products are or are not powered by AI, and how.

[Back to Table of Contents](#)

4.4. Research Question 3: Benefits, Risks, Limitations, and Potential Biases of AI in Digitization Projects

When respondents whose organizations had used AI tools (26.3%) assessed how those AI tools helped to achieve their digitization objectives, they replied that AI improved efficiency (10.2%), searchability (10.25), discoverability (10.2%), or generally resulted in greater access to

collections (8.2%). They also responded with the types of machine learning that had positive effects. These included transcription/speech to text (14.3%), OCR/HTR (12.2%), automatic metadata generation (4.1%), and computer vision (4.1%) in the top four responses.

Conversely, respondents described limitations or negative aspects of the AI tools they used. The negative results included inaccuracy (12.1%), poor quality of OCR (12.1%), the time-consuming nature of implementing AI (9.1%), cost (6.1%), the unusefulness of generated metadata (6.1%), and the need for human checking (6.1%). A portion of respondents (9.1%, fourth ranked) stated that they experienced no negative impacts of using AI tools in their digitization processes. However, this could be compared to the 90.9% of respondents who offered various negative aspects of using AI. Continuing with respondents who had deployed AI, the survey asked how the AI tools were selected. Most often, participants stated that the AI tools were either recommended or that they were selected based on ease of use (16.3% each). In addition, AI tools were often selected because they were included in software already used by the organization (11.3%).

All participants were asked to identify any barriers to using AI. Most respondents cited reasons linked to lack of resources or staffing issues, including lack of training or expertise (18.8%) and lack of staff time to integrate or use AI (18.3%). Other reasons, such as lack of funds (15.4%), lack of AI software and tools (15%) and lack of institutional awareness/support (14.5%) were also significant barriers. However, when asked whether their organizations had identified any risks of using AI in digitization, only 19.7% of respondents replied yes, while 44.7% were not sure, and 35.6% had not identified any risks. The responses suggest that organizations are still at an early stage in terms of analyzing the risks of using AI. The percentage of “no’s” suggests that those organizations have either determined that there are no/low risks to using AI for digitization, or that they are not analyzing the risks of AI.

When those who had identified risks were asked to elaborate, they most often described bias and ethical concerns (29.6%), as well as the inaccuracy of AI tools (16.7%). Lack of transparency (7.4%), the need for human quality control (7.4%), and data privacy (5.6%) were also rated as risks to lesser degrees. The responses show that participants are concerned with broader ethical questions pertaining to the development of AI. For example, they raised questions around bias in training sets and AI algorithms, such as racial, ableist, and gender bias, including transphobic bias; the ethics of using facial recognition tools; and the lack of oversight within tech companies in developing AI tools/models. One respondent cited the fact that Google fired its AI ethics team after they raised concerns to management.

[Back to Table of Contents](#)

4.5. Research Question 4: Future Perspectives for Uses of AI Tools in Digitization Projects

The survey findings indicate that, if the risks of using AI tools are properly addressed, machine learning has potentially beneficial applications for future digitization projects and programs. In particular, many participants are motivated by the potential of AI to help automate and bolster certain kinds of human labour. When asked about motivators for using AI tools during digitization, respondents expressed a wish for AI tools to help increase the effectiveness and efficiency of workflows (25.5%) and supplement existing staff efforts (20.5%). They also saw the potential for AI to maximize the impact and reach of organizational resources (13.6%). Fewer participants stated that they were motivated by AI to support deeper analysis (8.5%) or to support analysis or decision-making (7.8%).

Participants were also asked which AI tools in particular could be most helpful for digitization. The AI tools most cited were OCR/HTR (21.4%), AI tools for metadata management (20.6%), automatic transcription of audio/moving images (17.3%), and AI tools for classification or tagging (16%). Quality control (10.1%) and translation (6.2%) were also cited. When asked to elaborate, similar to responses for motivators of using AI, participants remarked on the potential for AI to save staff time, including the potential for AI to alleviate tedious or repetitive tasks, or to enhance access to collections even at the item level.

The main motivators for using AI to support digitization, and the AI tools considered to be most helpful for digitization, align with responses to the question of what respondents would do differently in their next digitization efforts. When reflecting on lessons learned for future digitization projects/programs, respondents expressed a desire to improve various aspects of project management (32.9%), as well as staffing (12.5%) and metadata creation (12.5%). For project management, participants stated that they would establish more realistic planning in terms of timelines and scope, more clear policies around digitization and selection of materials, and improved strategic or methodological approaches, such as better quality control, improved workflows, or the creation of inventories to better understand collections before digitization. In terms of staffing, many respondents stated that they would hire more staff or more qualified staff, especially for tasks such as metadata creation or management. These lessons learned reveal some of the current needs and gaps of digitization programs/projects, and suggest ways in which AI could help to address these gaps. In some instances, they also directly reflect respondents' stated motivations for using AI for digitization, such as the above-cited reasons of increasing workflow efficiency (25.5%) and supplementing existing staff efforts (20.5%).

Moreover, the question of which AI tools could be most helpful in digitization processes is revealing when compared to the question of the perceived benefits of digitization and the greatest successes of digitization projects/programs in which respondents were involved. For both of these questions, the top responses primarily focused on increased access to collections as the top motivator (33.6% when increased external and internal access are combined) and similarly, increased access to collections as the greatest success of digitization projects (31.5%). In parallel, participants most often cited AI tools that have the potential to enhance access, search, and discoverability of collections, including OCR/HTR, metadata management, transcription, classification and tagging, and translation.

The findings suggest that in future, the most pertinent applications of AI tools for digitization purposes center on AI tools to automate digitization workflows, particularly in the areas of quality control or metadata creation and management, for example, thus alleviating staff time and work. AI tools that enhance searchability and discoverability of digitized collections, whether OCR/HTR, metadata creation and tagging, or transcription or translation will also continue to be key uses of AI tools in digitization projects. The answers are in line with core digitization activities, many revolving around the creation or enrichment of metadata, and show that participants perceive AI as mostly supporting already-existing digitization tasks or outcomes, rather than offering completely new features. However, in some cases, such as translation, item-level description, or image enhancement, AI may afford previously impossible or near-impossible offerings.

[Back to Table of Contents](#)

5. Survey: Question-by-Question Responses

The following section provides question-by-question results, including a narrative description and visual representation of responses for each survey question. The section is organized following the four survey sections: “Section 1: General Information,” “Section 2: Digitization Activities,” “Section 3: Digitization Contexts,” and “Section 4: Digitization and AI.”

5.1. Section 1: General Information

1) In what country is your organization located?

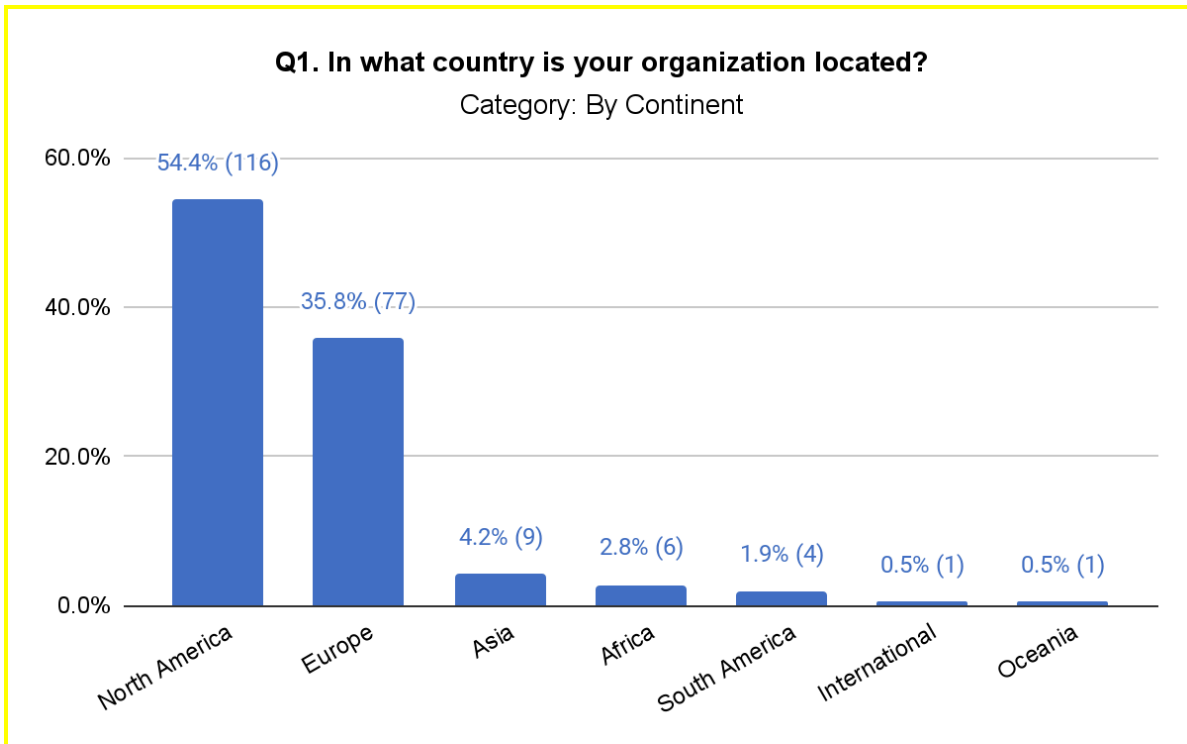
Of 217 submissions included in the overall analysis, this question received 214 responses.

Respondents work for organizations located in 41 countries and six continents. The top countries were the United States of America (39.7%), Canada (13.6%), the United Kingdom (10.7%), Spain (5.1%), the Czech Republic (3.3%), Switzerland (2.3%), and France, Germany, and the Netherlands (1.9% each). All other countries on the list received only one or two responses (0.9% or 0.5%).

By continent, the majority of respondents' organizations are located in North America (54.4%) and Europe (35.8%), followed by Asia (4.2%), Africa (2.8%), South America (1.9%), and Oceania (0.5%). Two respondents wrote "International." One of the two respondents qualified that their organization was international but headquartered in Switzerland; this response was coded as "Switzerland."

Q1. In what country is your organization located?		
Country	Count	%
United States of America	85	39.7%
Canada	29	13.6%
United Kingdom, including Ireland (1) and Scotland (3)	23	10.7%
Spain, including Catalonia (3)	11	5.1%
Czech Republic	7	3.3%
Switzerland	5	2.3%
France	4	1.9%
Germany	4	1.9%
Netherlands	4	1.9%
Austria	2	0.9%
Chile	2	0.9%
Costa Rica	2	0.9%
Croatia	2	0.9%
Greece	2	0.9%
Italy	2	0.9%
Lebanon	2	0.9%
Slovenia	2	0.9%
South Africa	2	0.9%
Albania	1	0.5%
Argentina	1	0.5%
Australia	1	0.5%
Bangladesh	1	0.5%
Belgium	1	0.5%
Botswana	1	0.5%
Indonesia	1	0.5%
International	1	0.5%
Israel	1	0.5%
Kenya	1	0.5%
Luxembourg	1	0.5%
Macedonia	1	0.5%
Nigeria	1	0.5%
Norway	1	0.5%
Peru	1	0.5%
Philippines	1	0.5%

Poland	1	0.5%
Qatar	1	0.5%
Romania	1	0.5%
Serbia	1	0.5%
Slovak Republic	1	0.5%
South Korea	1	0.5%
Thailand	1	0.5%
Zimbabwe	1	0.5%
Total	214	100.0%



2) For which type of organization do you work?

Of 217 submissions included in the overall analysis, this question received 217 responses.

A majority of respondents work for governmental or public organizations (59.4%), followed by organizations in the private sector (21.7%) and non-governmental organizations (7.4%). The degree of difference between the top three types of organizations is substantial: more than 35% between the top two responses, and almost 15% difference between the top 2nd and 3rd choices.

Q2. For which type of organization do you work?		
Type of organization	Count	%
Public/governmental	129	59.4%
Private	47	21.7%
Non-governmental	16	7.4%
International/intergovernmental	11	5.1%
Other - Academic/University	7	3.2%
Other - Non-profit	5	2.3%
Other - Religious	1	0.5%
Other - Public/private partnership	1	0.5%
Total	217	100.0%

3) To which sector(s) does your organization belong? Select all that apply.

Of 217 submissions included in the overall analysis, 216 participants provided 406 responses to this multiple choice question, which enabled participants to select multiple answers.

This question permitted multiple responses per participant, recognizing that some organizations belong to multiple sectors. The top three sectors in which respondents work are “Education/ Research” (31%), Culture (17.5%), and “Memory Institution” (15.5%). Respondents working in these three sectors make up 64% of all survey responses combined. It is important to note that this question asks respondents to classify the sector to which their organization belongs and not the sector most closely related to their individual work. Respondents who selected the “Other” option were given a free-form text field to explain their selection. When provided, the descriptions in the “Other” responses (6.4%) are included in the table below.

Q3. To which sector(s) does your organization belong? Select all that apply.		
Sector	Count	%
Education/Research	126	31.0%
Culture	71	17.5%
Memory institution	63	15.5%
Sciences	26	6.4%
Communications/Media	21	5.2%
Information technology	20	4.9%

Services Sector/Tourism	12	3.0%
Finance	11	2.7%
Events	7	1.7%
Health	7	1.7%
Agriculture	6	1.5%
Legal	4	1.0%
Transportation	4	1.0%
Management/Human Resources	2	0.5%
Other: Beverage Company, Broadcast media, Energy, Government, Health sciences, Humanitarian, Intellectual property rights protection, Local government, Manufacturing, Nuclear, Procurement regulatory authority, Religious organization (2), Tribal archive, UN	26	6.4%
Total	406	100.0%

4) What is your role within the organization (e.g., archivist, librarian, records manager, supervisor of archives, digitization specialist, audiovisual archivist, administrator, technician, consultant, etc.)? If you have more than one role, please specify.

Of 217 submissions included in the overall analysis, 217 participants provided 250 responses to this free-form question.

The majority of respondents self-identified as archivists (38%). The next largest group of respondents, librarians, accounted for just under half that number (17.6%). The third-largest composition of respondents was digitization specialists or technicians (11.2%). Many respondents indicated that they hold a leadership or management/supervisory role within their organizations. Overall, 9.7% of respondents across all work areas hold leadership positions within their organizations, including acting as Director, Chief, Head, University/City Archivist, Founder/Owner, or Deputy Head. 14.3% of respondents indicated that they are in management, supervisory, or coordinator roles.

Q4. What is your role within the organization? If you have more than one role, please specify.		
Role/Work area	Count	Percentage
Archivist	95	38.0%
Librarian	44	17.6%
Digitization specialist/technician	28	11.2%
Records manager	18	7.2%

Collections management	9	3.6%
Digital engagement/initiatives	7	2.8%
Leadership (when no work area mentioned)	7	2.8%
Photography/photographer/imaging	6	2.4%
Curator	4	1.6%
Administrator	3	1.2%
Information management/governance	3	1.2%
Manager (when no work area mentioned)	3	1.2%
Research/researcher	3	1.2%
Educator	2	0.8%
Historian	2	0.8%
Neuropsychiatrist	2	0.8%
Registrar/registry	2	0.8%
Restorer	2	0.8%
Artist	1	0.4%
Board of Directors	1	0.4%
Conservator	1	0.4%
Consultant	1	0.4%
Digital humanities specialist	1	0.4%
Documentation	1	0.4%
IT	1	0.4%
Museum specialist	1	0.4%
Publishing	1	0.4%
Technician	1	0.4%
Total	250	100%

The table below shows the total number of respondents (24%) who indicated that they hold a leadership or management role within their organization. Leadership roles include Directors, Heads or Chiefs, Founders or Owners, and Deputy Heads. Management roles include Managers, Supervisors, and Coordinators.

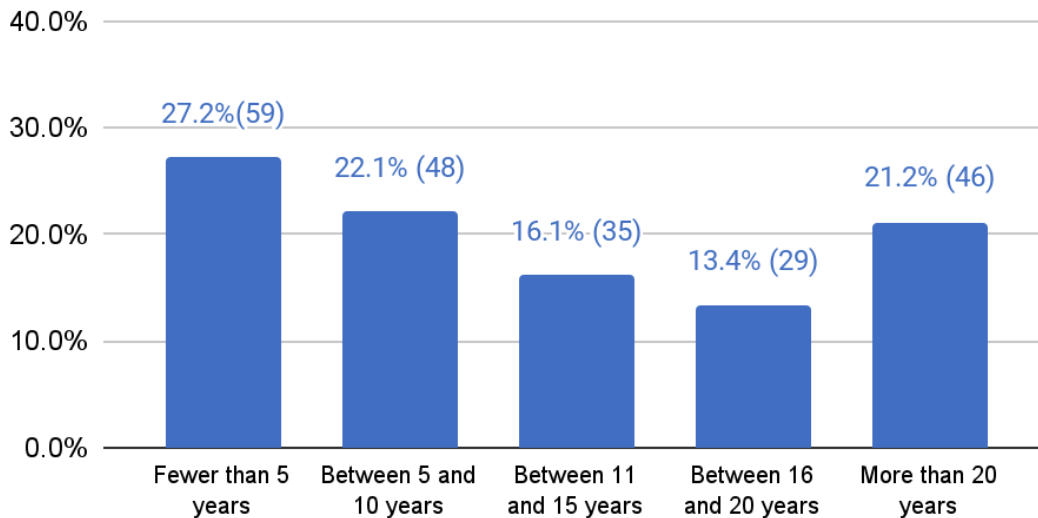
Q4. By Level of Responsibility - What is your role within the organization? If you have more than one role, please specify.		
Level of Responsibility	Count	Percentage
Leadership - Chief/Head, Director/Deputy, Founder/Owner	21	9.7%
Managerial - Coordinator, Manager, Supervisor	31	14.3%
Total	52	24%

5) How many years have you worked within the organization?

Of 217 submissions included in the overall analysis, this question received 217 responses.

This question asked respondents how long they have worked within their organizations, and not how long they have worked within their professions. Responses to this question were well distributed. The largest group of respondents reported working within their organization for fewer than five years (27.2%). The next largest categories of responses were from those who had worked within their organizations for between five and ten years (22.1%) and for more than twenty years (21.2%); these latter categories received almost equivalent responses.

Q5. How many years have you worked within the organization?

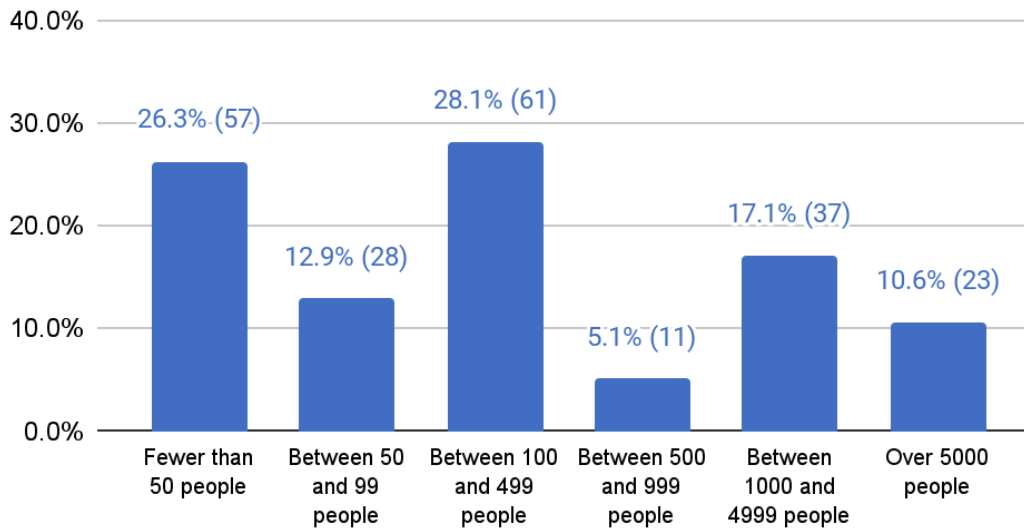


6) What is the approximate size of your organization in your particular location?

Of 217 submissions included in the overall analysis, this question received 217 responses.

Among respondents, 28.1% reported working in an organization with between 100 and 499 people, while 26.3% reported working in an organization with fewer than 50 people. The next most frequently reported organization sizes were “Between 1000 and 4999 people” (17.1%) and “Between 50 and 99 people” (12.9%). Organizations with “Over 5000 people” (10.6%) and “Between 500 and 999 people” (5.1%) were the least well represented.

Q6. What is the approximate size of your organization in your particular location?



[Back to Table of Contents](#)

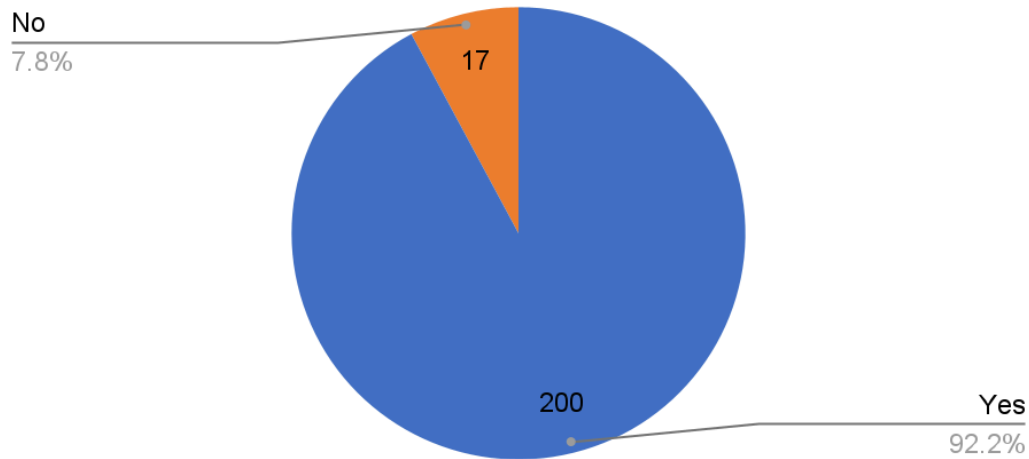
5.2. Section 2: Digitization Activities

7) Has your organization undertaken at least one digitization project or program in the past five years (i.e., a defined project or program to digitize selected archival holdings or documentary heritage collections, as opposed to ad-hoc, everyday digitization)?

Of 217 submissions included in the overall survey analysis, this question received 217 responses.

200 people (92.2%) responded “Yes” to the question of whether their organizations have undertaken at least one digitization project or program in the past five years, while 17 people (7.8%) responded “No.” The responses show that overwhelmingly, respondents’ organizations have undertaken digitization projects in the last five years.

Q7. Has your organization undertaken at least one digitization project or program in the past five years?



8) What is/was your role in your organization's digitization project(s) or program(s)? Please describe briefly.

This question was conditional upon a “Yes” answer to question no. 7, which received 200 “Yes” responses. Of the 200 possible responses, this question received 195 responses.

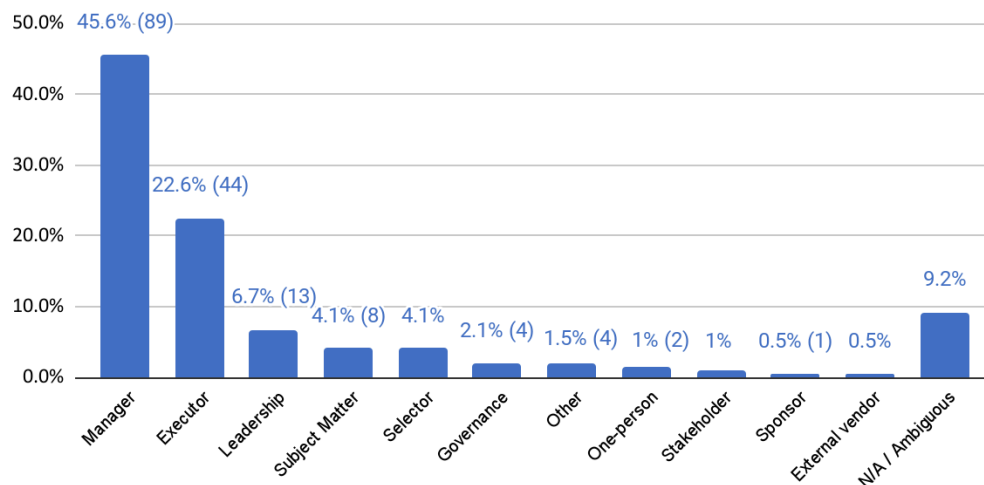
The free-form responses were coded into the following types of roles undertaken by participants in relation to digitization activities within their organizations:

- *Leadership* - Executive/leader within the organization - sets priorities for the digitization program, oversees the digitization program at a high level. May include establishing budgets and staffing.
- *Digitization manager/coordinator* – Coordinates, manages, and supervises digitization projects and programs; includes project leads. Sometimes raises funds or also executes digitization tasks/activities, such as metadata creation (see below under Digitization executor).
- *Digitization selector* - Selects materials to be digitized. Participants are categorized here when no other tasks/responsibilities are mentioned except for selection of materials.
- *Digitization executor* – Performs digitization tasks and activities across one or multiple digitization phases (pre-digitization, during, and post-digitization), including, but not limited to: selection of materials (when one of many tasks executed by the participant), imaging, creation or management of metadata, technical set-up and implementation, quality control, file management, publication for access, and post-production. In some cases, digitization executors are also responsible for fundraising (writing grant proposals). Includes persons who undertake ad-hoc digitization by request.

- *Governance* – Creator of relevant policies, procedures, guidelines.
- *Subject-matter expert* – Advises or consults on some aspect of the digitization project/program, such as digital collections management or digital preservation, but may not be directly involved in executing the digitization.
- *Stakeholder* – Is directly affected by the digitization project or program, either upstream or downstream. Includes those responsible for digital assets management systems (DAMs) or digital preservation.
- *One-person project* – Digitization projects or programs coordinated, prepared, and undertaken by one person.
- *Sponsor* – Financial supporter for the digitization project or program.
- *External vendor* – External digitization service provider.

Survey respondents most often participated in the digitization process as digitization managers or coordinators (45.6%) or as digitization executors (22.3%), that is, persons directly involved in the digitization. In coding the responses, it emerged that the line between digitization manager/coordinator roles and digitization executors (those carrying out digitization activities) are often blurred, given different organizational contexts such as size and structure of staffing. Many individuals in a management role also stated that they took part in digitization activities. In addition, our definition of “digitization executor” is broad because it encompasses activities across all stages of digitization (pre-digitization, digitization, and post-digitization). Several participants (9.2%) provided responses that were either not specific enough, or the question was not applicable to them, for example, because they arrived in their organizations after a digitization project or program took place. Four respondents (1.5%) were categorized as “Other” roles, including Administrator (2), Curator (or person who selects materials to be highlighted for promotional purposes once digitized), and one respondent who stated that they provide materials to be digitized when requested.

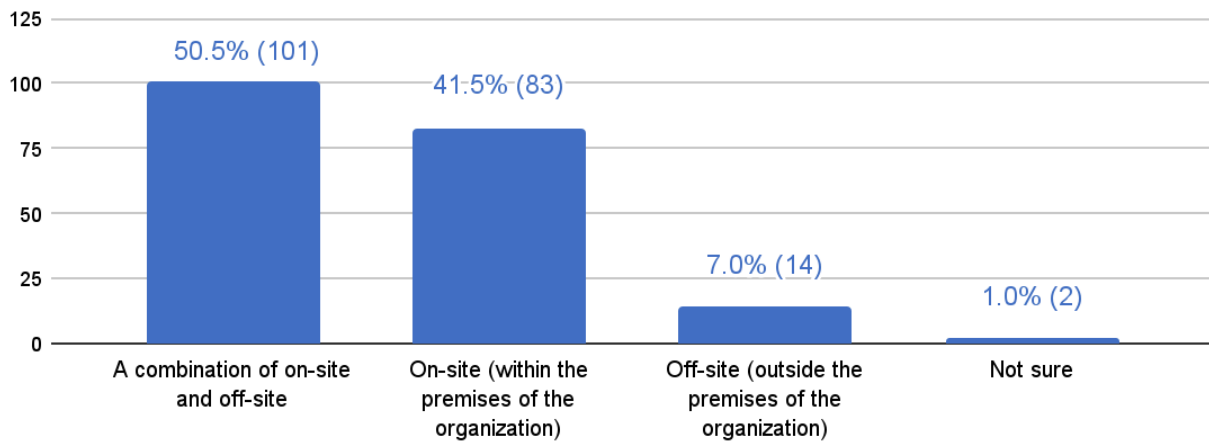
Q8. What is/was your role in your organization's digitization project(s) / program(s)?



9) Where does/did digitization take place in your organization?

This question was conditional upon a “Yes” answer to question no. 7, which received 200 “Yes” responses. Of the 200 possible responses, this question received 200 responses.

Q9. Where does/did digitization take place in your organization?



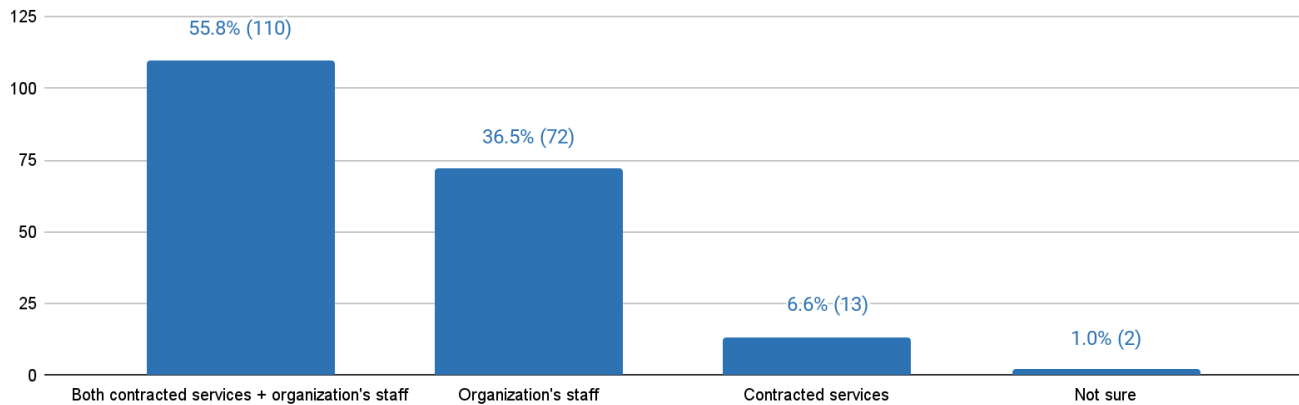
Half of respondents reported that their organization’s digitization activities take place in a combination of on-site and off-site premises (50.5%). A significant portion of respondents reported that their organization carries out digitization activities entirely on-site, i.e., within the premises of the organization (41.5%). Only 7% stated that digitization takes place off-site.

10) Who performs the digitization work in your organization?

This question was conditional upon a “Yes” answer to question no. 7, which received 200 “Yes” responses. Of the 200 possible responses, this question received 200 responses.

More than half of respondents reported that both contracted services and organizational staff perform digitization work within their organizations (55.8%). More than one-third of respondents reported that their organization’s digitization work is performed entirely by organizational staff (36.5%). Conversely, only 6.6% of respondents indicated that contracted services perform all digitization activities within their organization.

Q10. Who performs the digitization work in your organization?

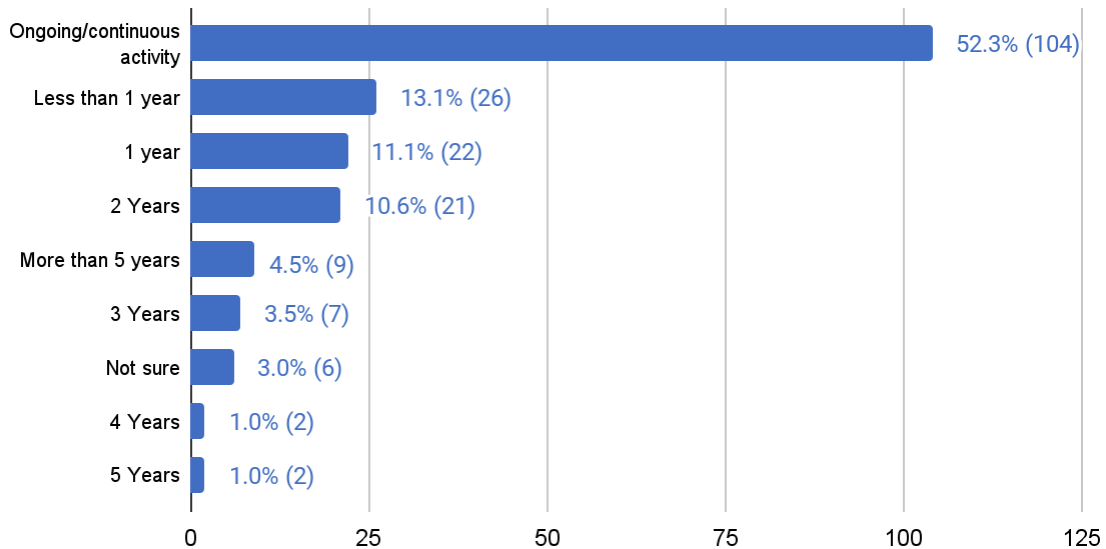


11) On average, what is the duration of a digitization project/program in your organization?

This question was conditional upon a “Yes” answer to question no. 7, which received 200 “Yes” responses. Of the 200 possible responses, this question received 200 responses.

More than half of respondents reported that digitization is an ongoing or continuous activity within their organization (52.3%). This was closely followed by durations of less than one year (13.1%), one year (11.1%) and two years (10.6%), with the lowest response rates being four and five years (1% each). The responses suggest that participants who work at organizations without ongoing/continuous digitization programs take on shorter term and/or ad-hoc digitization projects.

Q11. On average, what is the duration of a digitization project/program in your organization?



12) What type(s) of formats has your organization digitized? Select all that apply.

This question was conditional upon a “Yes” answer to question no. 7, which received 200 “Yes” responses. Of the 200 possible responses, this question received 200 responses.

This question enabled respondents to check all multiple choice answers that applied, as well provide an “Other” response. 200 respondents selected a total of 904 answers, including seven “Other” responses. On average, respondents to this question selected 4.5 formats from a predetermined list of eight formats, suggesting that most respondents’ organizations have digitized (or are digitizing) multiple formats. The most commonly reported formats are “Unbound textual documents” (18.1%) and “Photographs” (18.0%), followed closely by “Bound textual documents” (16.3%), “Films/videos” (13.5%), “Two-dimensional graphic materials” (13.3%), and “Audio recordings” (12.9%). Comparatively few respondents indicated that their organizations had digitized “Three-dimensional objects” (7.1%).

Seven respondents who selected the “Other” option indicated in their free-form answers that their organizations are digitizing the following types of materials:

- 3D scans of building structures
- Surfaces using RTI (Reflectance Transformation Imaging)
- Framed art works
- Textiles, fabrics, clothing (x2)
- Manuscripts

- Old and handwritten documents, old and rare books that need special treatment before preparation for digitization, large format documents not adapted to information equipment (photo cameras and scanners)

Q12. What type(s) of formats has your organization digitized? Select all that apply.		
Type of format	Count	%
Unbound textual documents	164	18.1%
Photographs (negatives, prints, slides, glass plate negatives, etc.)	163	18.0%
Bound textual documents (volumes or publications)	147	16.3%
Films/videos	122	13.5%
Two-dimensional graphic materials (maps, posters, engravings, etc.)	120	13.3%
Audio recordings	117	12.9%
Three-dimensional objects	64	7.1%
Other (please describe)	7	0.8%
Total	904	100.0%

13) What primary criteria does your institution use to prioritize materials for digitization? Select the top three.

This question was conditional upon a “Yes” answer to question no. 7, which received 200 “Yes” responses. Of the 200 possible responses, this question received 200 responses.

The top criteria when prioritizing materials for digitization were “Potential research value” (21.3%) and “Value in terms of public access to information” (19.6%). These criteria were followed by factors relating to the physical condition of materials, including: “State of conservation of the materials” (17.2%), “Technological obsolescence of the physical format” (14.4%), and “Age of the materials” (9.3%). The criteria “Administrative or evidential value” (7.6%) and “Outreach value” (7%) were the least prioritized.

Among the 19 respondents who selected “Other” (3.5%), 7 people noted that digitization takes place as needed, for example in response to a user request, as part of an exhibit, or to support programming efforts. Other answers indicated that priority was given according to legislative mandate, historical significance, cultural and scientific significance, physical storage capacity, profit potential, and materials produced by the organization. One respondent reported that their institution aims to digitize its entire collection. Another respondent from a medical archive stated that the organization selects items with no privacy restrictions, such as the U.S. Health

Insurance Portability and Accountability Act (HIPAA) and the U.S. Family Educational Rights and Privacy Act (FERPA), and that priority is also determined by institutional policies.

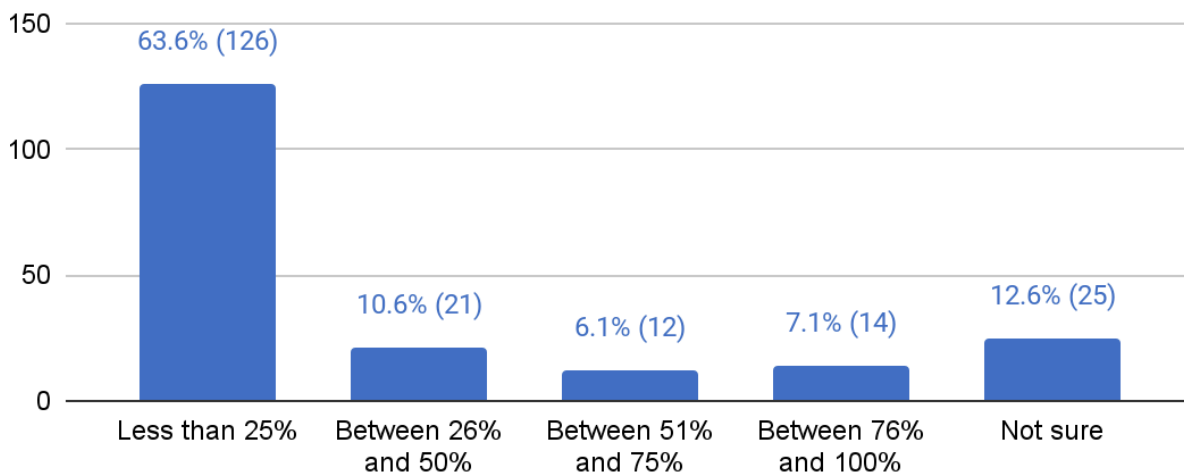
Q13. What primary criteria does your institution use to prioritize materials for digitization? Select the top three.		
Selection Criteria	Count	Percentage
Potential research value	115	21.3%
Value in terms of public access to information	106	19.6%
State of conservation of the materials	93	17.2%
Technological obsolescence of the physical format	78	14.4%
Age of the materials	50	9.3%
Administrative or evidential value	41	7.6%
Potential outreach value	38	7.0%
Other	19	3.5%
Total	540	100.0%

14) What is the estimated percentage of materials that have been digitized within your department/unit?

This question was conditional upon a “Yes” answer to question no. 7, which received 200 “Yes” responses. Of the 200 possible responses, this question received 198 responses.

The majority of respondents reported that “Less than 25%” of their organization’s materials have been digitized (63.6%). Others reported that they were either unsure (12.6%) or that their organizations had digitized “Between 26% and 50%” (10.6%) of their collections. Higher rates of digitization were notably much less common, including organizations that had digitized “Between 76% and 100%” (7.1%) of materials, or “Between 51% and 75%” (6.1%) of materials.

Q14. What is the estimated percentage of materials that have been digitized within your department/unit?

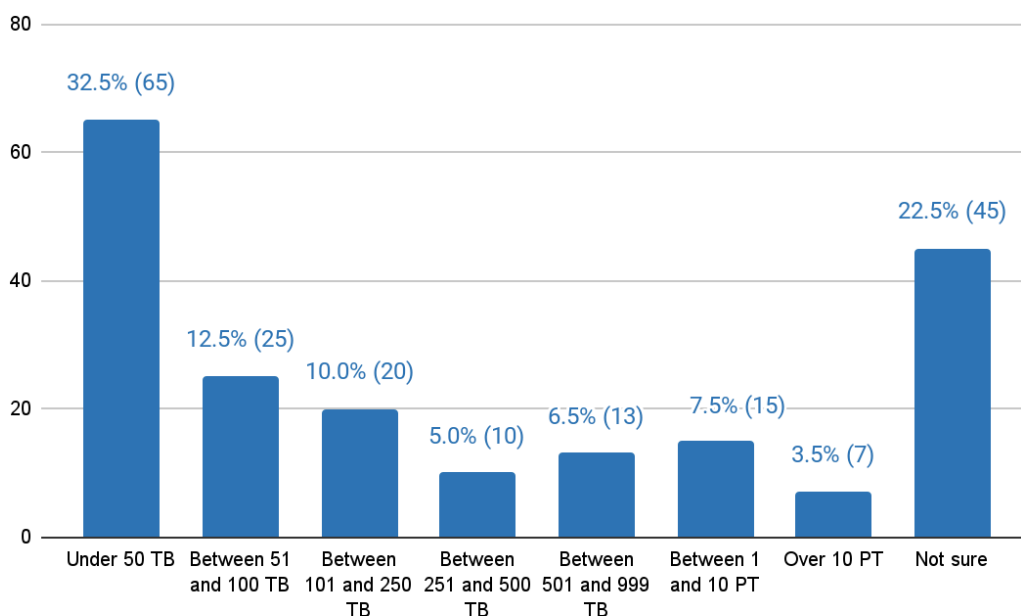


15) What is the approximate quantity of digitized materials?* (*Reminder: digitized materials refer to analogue materials converted into digital form and not born-digital materials).

This question was conditional upon a “Yes” answer to question no. 7, which received 200 “Yes” responses. Of the 200 possible responses, this question received 200 responses.

32.5% of respondents reported that their organization had under 50 terabytes (TB) of digitized materials. The next most popular selections were “Between 51 and 100 TB” of digitized materials (12.5%) and “Between 101 and 250 TB” of digitized materials (10%). The remaining possible given quantities, all above 250 TB, each received fewer than 10% of responses. However, when counted together, 22.5% of respondents indicated that their organizations have from 250 TB to over 10 PB of digitized materials. The data indicates that 55% of respondents’ organizations have 250 TB or less of digitized materials, as compared to 22.5% with more than 250 TB of digitized files. Significantly, 22.5% of respondents were unsure about the quantity of digitized materials within their organizations. The data also suggests that almost a quarter of respondents either do not know how much their organization has digitized, do not know how to express this information using a quantitative measure (TB and petabytes (PB)), or that they were unprepared to make an estimate while filling out the questionnaire.

Q15. What is the approximate quantity of digitized materials?



16) What do you consider to be the major activities and/or functions* carried out in digitization projects and programs? Please select all that apply and add any that you think are missing. (*Activities and functions are taken from the Croatian State Archives' Guidelines for Digitization of Archival Materials (2021) and from a draft model of digitization activities created by this study team).

Of 217 submissions included in the overall survey analysis, this question received 216 responses.

This question enabled respondents to check all multiple choice answers that applied out of the 14 possible choices, including an option for "Other." The average respondent selected 11 answers. Respondents most often selected "Digitization (i.e., imaging, quality control, file naming, file versioning)" (10.4%) and "Metadata management (e.g., identification, creation, classification, editing, etc.)" (10.3%) as major digitization activities. The next two most checked answers were "Selection of materials to be digitized" (9.7%) and "Preparation of materials to be digitized" (9.1 percent). "Storage and backup" (8.4%) and "Post-processing" (8.3%) received practically the same number of responses. Only 0.7% of respondents selected "Other"; most of their responses are included in the table below. Some "Other" answers were not included in the description due to the uncertainty of their meaning.

It should be noted that the number of selections decreases according to the order that each given answer appeared on the list. In other words, the top listed answers were the most often selected. This response pattern suggests a possible questionnaire design limitation, in which information overload may have contributed to respondent fatigue and overly determined respondents' choices.

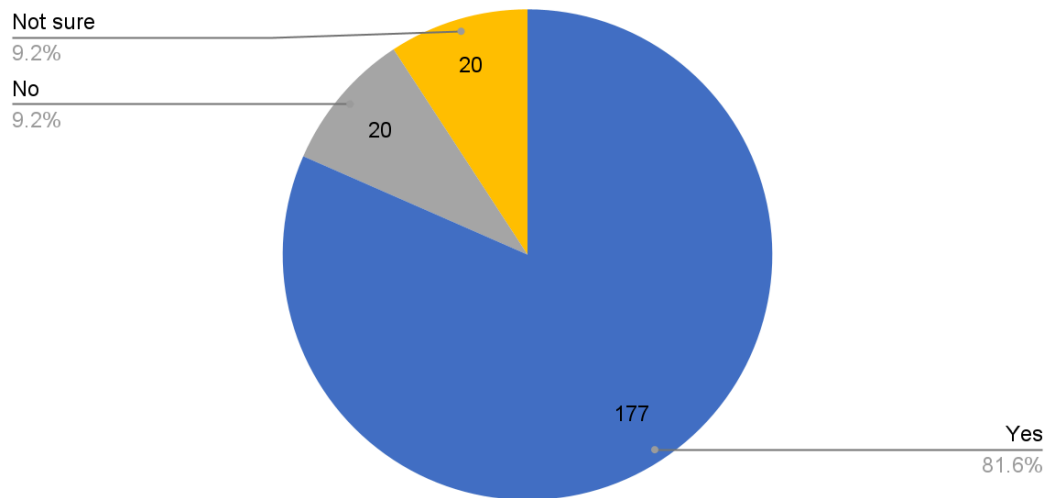
Q16. What do you consider to be the major activities and/or functions* carried out in digitization projects and programs? Please select all that apply and add any that you think are missing.		
Activities and/or Functions	Count	Percentage
Digitization (i.e., imaging, quality control, file naming, file versioning)	190	10.4%
Metadata management (e.g., identification, creation, classification, editing, etc.)	188	10.3%
Selection of materials to be digitized	177	9.7%
Preparation of materials to be digitized	165	9.1%
Storage and backup	153	8.4%
Post-processing (e.g., file editing or re-digitization if necessary, optical character recognition (OCR) if applicable, etc.)	152	8.3%
Long-term digital preservation	145	8.0%
File management and publication (i.e., in database system/access interface)	135	7.4%
Project management	116	6.4%
Validation (i.e., information package creation, ingest, integrity check)	112	6.2%
Rights management	98	5.4%
Staff training	96	5.3%
Outreach and communications	81	4.4%
Other: Digital preservation (x 2); Access; Accessibility including alt text and extended descriptions, trans- and gender-expansive/inclusive descriptions; Copyright and licensing; Different projects will have different needs and activities depending on scope and aims; Digital infrastructure maintenance; Digitization equipment maintenance; Moving items to premises where they will be digitized; Project management such as budgeting and payments; Protection of personal data and privacy, especially if the digital archive is used or processed by AI models; Strategic planning; Transcription and image description.	13	0.7%
Total	1821	100.0%

17) Do you consider digital preservation requirements in your digitization project(s) or program(s)?

Of 217 submissions included in the overall survey analysis, this question received 217 responses.

Overwhelmingly, respondents consider digital preservation requirements in their digitization project(s) or program(s). 81.6% of people responded “Yes” to this question, while 9.2% responded “No,” and another 9.2% were “Not sure.” This suggests a high level of professionalization and archival methodology in the field of digitization - an encouraging sign, given the importance of digital preservation as an essential part of an organization’s overall digitization strategy.

Q17. Do you consider digital preservation requirements in your digitization project (s) or program(s)?



18) Which digital preservation practices or activities are incorporated into your digitization project(s) or program(s)? Check all that apply.

This question was conditional upon a “Yes” answer to question no. 17, which received 177 “Yes” responses. Of the 177 possible responses, this question received 176 responses.

This question enabled respondents to check all multiple choice answers that applied out of the 9 possible choices, including an option for “Other.” On average, respondents selected 4 multiple choice answers. The three most commonly reported digital preservation practices or activities were “Recording of technical and preservation metadata” (17.7%), “Creation of multiple copies of digital files” (16.1%), and “Ingestion [of materials] into a trusted digital repository” (14.2%). 1.3% of respondents selected “Other”; their responses are included in the table below.

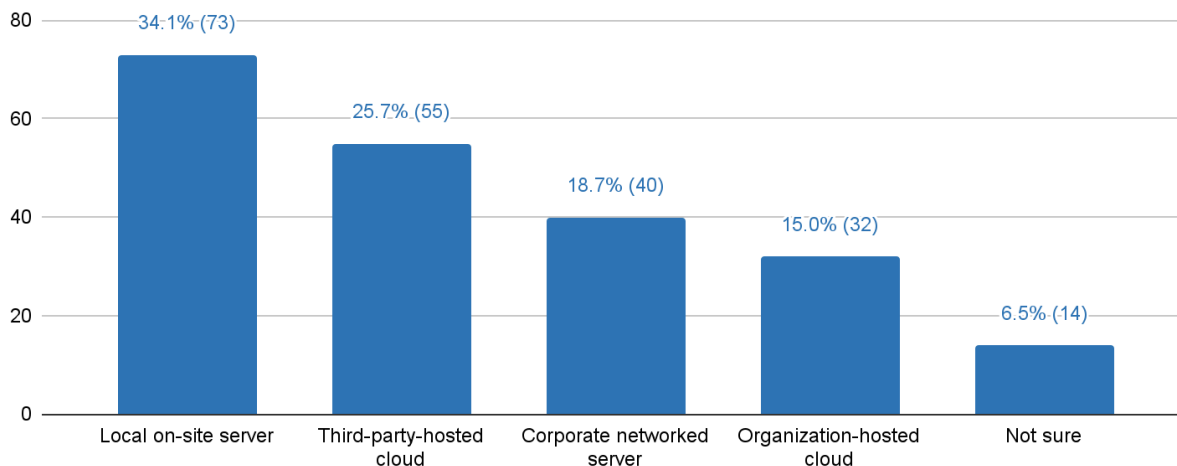
18. Which digital preservation practices or activities are incorporated into your digitization project(s) or program(s)? Check all that apply.		
Practice or Activities	Count	Percentage
Recording of technical and preservation metadata	131	17.7%
Creation of multiple copies of digital files	119	16.1%
Ingestion into a trusted digital repository	105	14.2%
Digital preservation policy	103	13.9%
Integrity checks	79	10.7%
Format migration	74	10.0%
Maintenance of digital file register	61	8.2%
Creation of hash values	59	8.0%
Other: Monitoring and use of long-term preservation file formats (x 2); Association with existing descriptive metadata; Intellectual metadata and structures are also important in the long term; Organization will incorporate LOCKSS but not for another year; Policy creation; Still building a digital preservation policy; We know all of these are important, but we don't currently have resources for all of these activities; Don't know.	10	1.3%
Total	741	100.0%

19) Where is your organization's digitized material located?

Of 217 submissions included in the overall survey analysis, this question received 214 responses.

Over one third of respondents indicated that their organization's digitized materials are located on a "Local on-site server" (34%). 26% reported that their digitized materials are located in a "Third-party-hosted cloud," 19% reported that they are stored in a "Corporate networked server," and 15% reported that they are stored in an "Organization-hosted cloud." 15% of respondents indicated that they were not sure where their organization's digitized materials were located.

Q19. Where is your organization's digitized material located?



20) According to the 3-2-1 backup rule, organizations should create 3 copies of their digital data (one primary copy and two backups); save backups on 2 different kinds of media; and keep at least 1 backup copy off-site (Peter Krogh, *The Dam Book: Digital Asset Management for Photographers*, 2005). How many copies of digitized files does your organization keep?

Of 217 submissions included in the overall survey analysis, this question received 216 responses.

The majority of respondents indicated that their organizations keep either two copies (28%) or three copies (31%) of their digitized files - a total of 59% combined. A significant number of respondents were not sure (17%), suggesting a possible lack of communication with IT. Another substantial group of respondents believe that “It is the cloud service provider’s responsibility to ensure backups” (10%), in other words, that backing up copies may be outsourced to a cloud service provider. No respondents brought up potential environmental issues or storage costs related to keeping multiple copies of digitized files.

Q20. How many copies of digitized files does your organization keep?		
Number of copies	Count	%
1	7	3.2%
2	61	28.2%
3	68	31.5%
More than 3	13	6.0%
It is the cloud service provider's responsibility to ensure backups	21	9.7%

Not sure	37	17.1%
Other	9	4.2%
Total	216	100.0%

Respondents who selected “Other” (4%) provided the following descriptions:

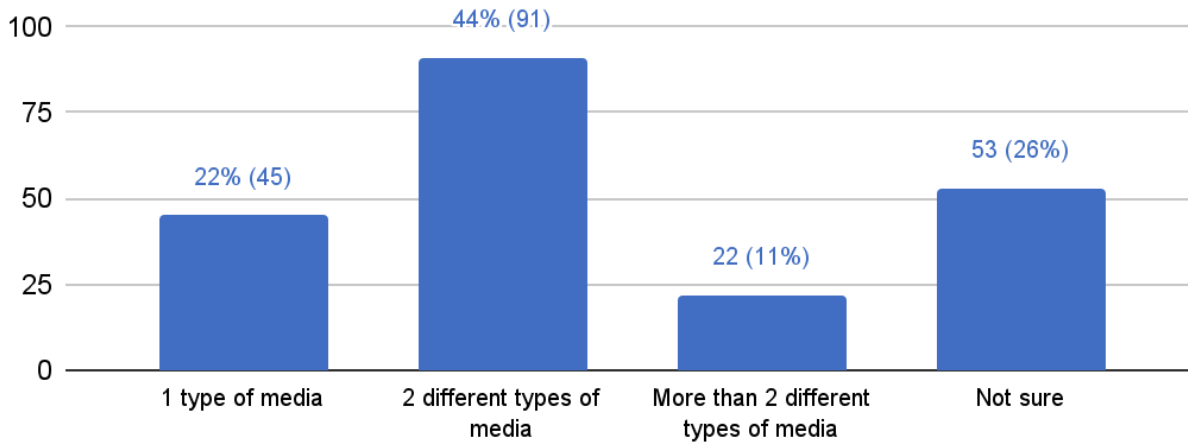
- We are currently migrating systems, so there is no single answer now.
- We have files on our servers that are backed up offsite and some files have a third cloud backup through our digital platform.
- For in-house digitized items, we have one file in our local server which is backed up at our campus; that server is then backed up at a remote server. So technically 3 copies, but they are not geographically disparate.
- We currently have 1 but are actively trying to enact a second. We will have two in the next two months.
- We have one copy that is accessible to us but is replicated at another university. We're also working on getting a copy in a third party-hosted cloud.
- For one project we will have 3 copies, for others not.
- We're in the process of acquiring a digital asset management tool.
- Varies based on asset format (still or moving image).
- One backup is on hard drives kept by organization, and cloud service provider responsible to backup everything stored in cloud.

21) If your organization keeps more than one copy of digitized files, how many different kinds of media do you keep the copies on (e.g., disk, tape, cloud or network storage, microfiche, etc.)?

Of 217 submissions included in the overall survey analysis, this question received 211 responses.

Almost half of respondents indicated that their organization keeps digitized files on “2 different types of media” (44%). The remaining 66% of respondents indicated that they were either “Not sure” (26%), that their organization keeps digitized files on “1 type of media” (22%), or “More than 2 different types of media” (11%).

Q21. If your organization keeps more than one copy of digitized files, how many different kinds of media do you keep the copies on?

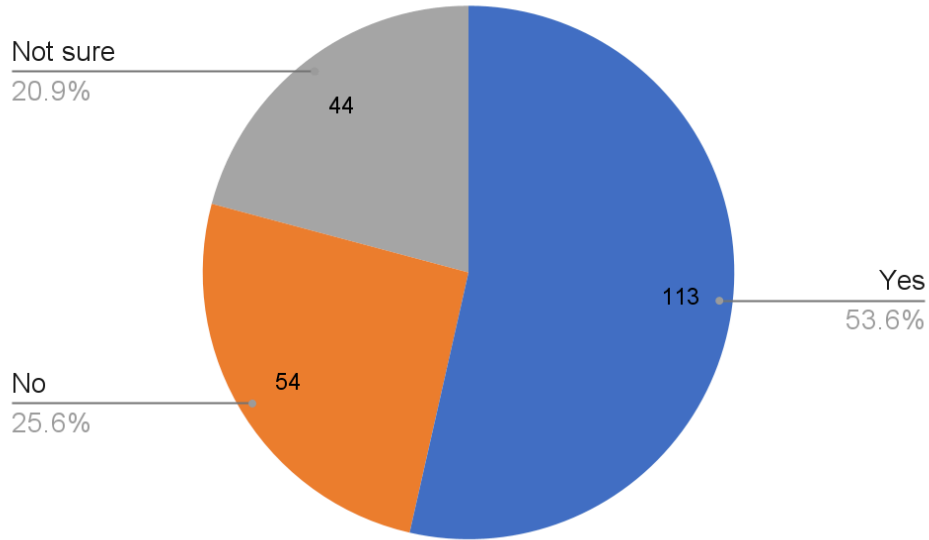


22) If your organization keeps more than one copy of digitized files, do you store at least one copy off-site (e.g., in a different building, town, region, country, or continent from the primary copy)?

Of 217 submissions included in the overall survey analysis, this question received 211 responses.

48.1% of respondents stated “Yes” to the question of whether their organization keeps more than one copy of digitized files off-site. 23.8% of respondents indicated “No” and 28.0% were “Not sure.”

Q22. If your organization keeps more than one copy of digitized files, do you store at least one copy off-site?

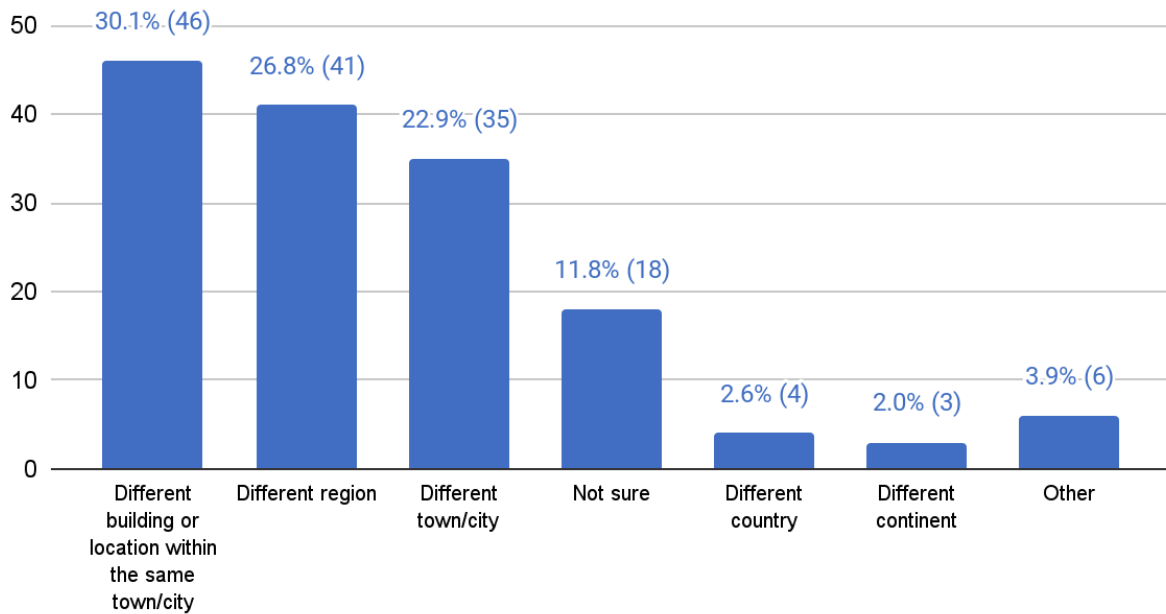


23) Where are your organization's backup copies located? Select all that apply.

This question was conditional upon a "Yes" answer to question no. 22, which received 113 "Yes" responses. Of the 113 possible responses, this question received 113 responses.

Most respondents indicated that their organization's backup copies are located in a different building or location within the same city/town (30.1%), in a different region (26.8%), or in a different town/city (22.8%). The least number of respondents noted that their organizations keep backup copies in a different country (2.6%) or in a different continent (2%). A significant number of respondents were not sure (11.8%) while some provided an "Other" response (3.9%).

Q23. Where are your organization's backup copies located? Select all that apply.



In the “Other” category, respondents provided the following answers:

- Cloud.
- Both hosted cloud backup and server back up, with offsite tapes.
- Different server.
- Geo-remote backup is coming soon.
- Working on setting up in a different location, possibly also different country storage for parts of the collection.

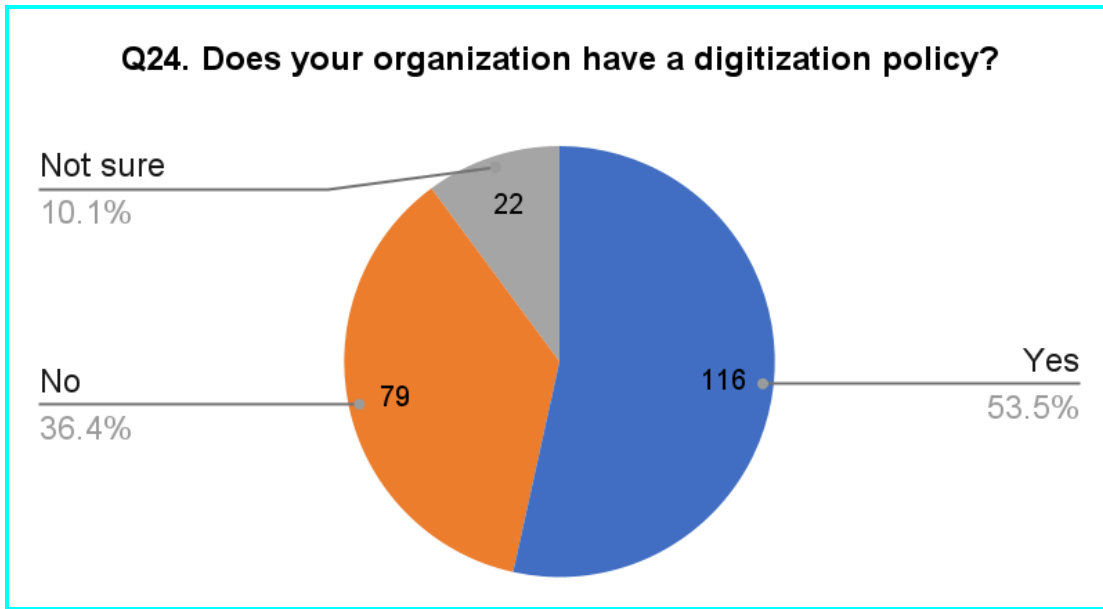
[Back to Table of Contents](#)

5.3. Section 3: Digitization Context

24) Does your organization have a digitization policy?

Of 217 submissions included in the overall analysis, this question received 217 responses.

53.5% of respondents indicated that their organization has a digitization policy, while 36.4% stated their organization does not have a digitization policy. 10.1% responded that they are not sure.

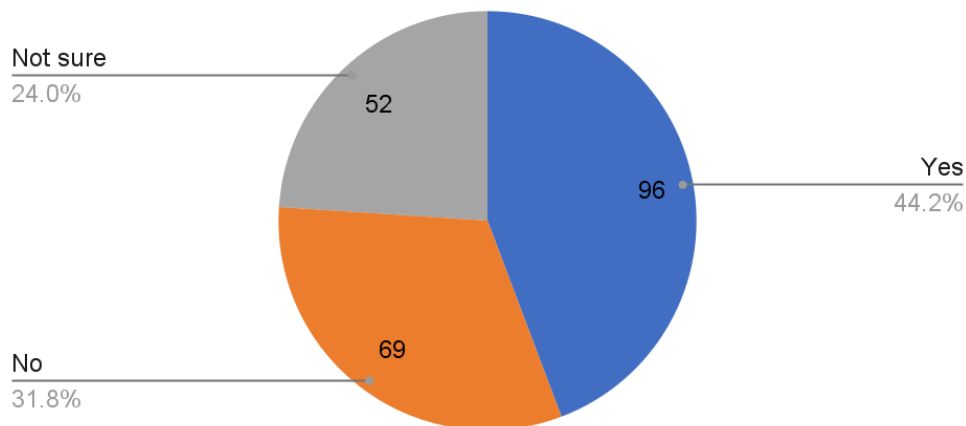


25(a) Are you aware of any legal or regulatory requirements pertaining to digitization in your country or jurisdiction?

Of 217 submissions included in the overall analysis, this question received 217 responses.

44.2% of respondents confirmed they are aware of legal requirements pertaining to digitization in their country or jurisdiction, while 31.8% are not aware of any legal requirements, and 24% are not sure. The results suggest different possible meanings, such as that legal requirements related to digitization either do not exist or are not well known in some jurisdictions.

Q25(a) Are you aware of any legal or regulatory requirements pertaining to digitization in your country or jurisdiction?



25)(b) Please briefly list the most pertinent legal or regulatory requirements.

This question was conditional upon a “Yes” answer to question 25(a), which received 96 “Yes” responses. Of 96 possible responses, 88 participants provided 113 responses to this free-form question permitting multiple responses.

Respondents cited copyright law most often as pertinent to digitization (41.6%), followed by privacy and data protection law, including GDPR and its article 17, Right to erasure ('Right to be forgotten') (11.5%). The third most cited types of regulations were digitization regulations at national level or as represented in standards/guidelines (6.2%), and evidentiary law, particularly legislation on electronic records as evidence (6.2%).

25)(b) Please briefly list the most pertinent legal or regulatory requirements.		
Type of Legal or Regulatory Requirement	Count	Percentage
Copyright/Intellectual property	47	41.6%
Privacy/Data protection - GDPR (4) including art. 17 Right to erasure ('Right to be forgotten')	13	11.5%
Digitization - national legislation (3), standards (3)	7	6.2%
Evidence - electronic records as evidence (5)	7	6.2%
Archival legislation - national level (6)	6	5.3%
Digital legislation - national level	5	4.4%
Records management/Records retention	5	4.4%
Access to information	4	3.5%
Accessibility for people with disabilities	4	3.5%
Contractual - donor contracts (2)	4	3.5%
Institutional regulations	3	2.7%
Canon law (religious)	1	0.9%
Community ethics	1	0.9%
Digital libraries - national legislation	1	0.9%
Other	5	4.4%
Total	113	100.0%

Several respondents provided the names of particular legislations that they consider to be most pertinent for digitization, listed below. A few respondents also cited international standards.

Q25)(b) Please briefly list the most pertinent legal or regulatory requirements				
Authority	Title of Regulation (when given)	Year	Level	Count
Canada	Copyright Act	1985	National	2

Canada	Electronic records as documentary evidence (CAN/CGSB-72.34-2017)	2017	National	2
Canada	Policy on Service and Digital	2020	National	1
Canada	Library and Archives of Canada Act	2004	National	1
Canada	Accessible Canada Act	2019	National	1
Canada	Access to Information and Privacy (ATIP)	1983	National	1
Canada - provinces	Freedom of Information and Protection of Privacy (FOIP)	2000	Provincial	1
Canada - provinces	Personal Information Protection Act	2003	Provincial	1
Costa Rica	Guidelines for the digitization of textual documents with the purpose of replacing the original support (National Technical Standard NTN-004)	2021	National	1
Croatia	Standard for Digital Libraries (NN 103/2021)	2021	National	1
Czechia	Copyright Act (Decree No. 488/2006 Coll.)	2006	National	1
European Union	General Data Protection Regulation - Right to erasure ('right to be forgotten') (Article 17)	2014	Intergovernmental	1
France	Definition and specification of services for the faithful digitization of paper documents and control of these services (NF Z42-026)	2023	National	1
ISO - International Standards Organization	Information and Documentation – Records Management, Section 5 - Records and Records Systems (ISO 15489-1 (2016-04-15))	2016	International	1
ISO - International Standards Organization	Information and documentation - Implementation guidelines for digitization of records (ISO/TR 13028:2010)	2010	International	1
Lebanon	Law on the Protection of Literary and Artistic Property	1999	National	1
Macedonia	Law on Archival Material and Law on Amendments to the Law on Archival Material	2012	National	1
Macedonia	Decree on office and archival operations	2014	National	1
Macedonia	Decree for determining the criteria for longer	2012	National	1

	terms for the use of archival material			
Peru	Legislative Decree No. 681	1991	National	1
Peru	Supreme Decree No. 029-2021-PCM (Regulations of the Digital Government Law)	2021	National	1
Slovenia	Protection of Documents and Archives and Archival Institutions Act (No. 30/06 and 51/14)	2006	National	1
South Africa	Protection of Personal Information Act (POPIA)	2020	National	1
South Africa	Electronic Communications and Transactions Act	2002	National	1
South Africa	Promotion of Access to Information Act 2 of 2000	2000	National	1
Spain	Technical Interoperability Standard for Document Digitalisation (translation)	2012	National	1
Spain	Organic Law 2/2023, of 22 March, on the University System (BOE-A-2023-7500)	2023	Institutional - university	1
Spain	Electronic administration laws		National	1
United Kingdom	Evidential Weight and Legal Admissibility of Electronic Information (BS 10008)	2008	National	1
United States of America	US National Archives and Records Administration - Federal Records Management: Digitizing Permanent Records and Reviewing Records Schedules	2023	National	1
United States of America	Transition to Electronic Records (M-19-21) and Update to Transition to Electronic Records (M-23-07)	2019, 2022	National	1
United States of America	Digitizing Temporary Federal Records (36 CFR Chapter XII, Subchapter B, Part 1236, Subpart D)	2019	National	2
United States of America	Copyright Act of 1976	1976	National	1
United States of America	Copyright Law of the United States and Related Laws Contained in Title 17 of the United States Code	2022	National	2
United States of America	Americans with Disabilities Act (ADA)	1990	National	1

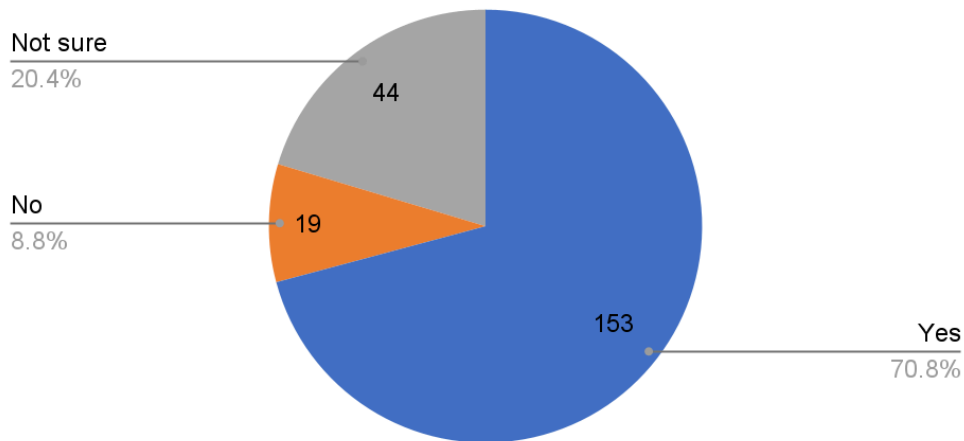
United States of America	Copyright Law of the United States - Fair Use in Section 107	2022	National	1
Total				40

26(a) Does/did your organization follow any regulations, guidelines, best practices, and/or standards during digitization project(s) or program(s)?

Of 217 submissions included in the overall analysis, this question received 216 responses.

71% of respondents confirmed that they follow regulations, guidelines, best practices, or standards during digitization, while 9% do not, and 20% are not sure.

Q26(a) Does/did your organization follow any regulations, guidelines, best practices, and/or standards during digitization



26(b) Please briefly list the most pertinent regulations, guidelines, best practices, and/or standards.

This question was conditional upon a “Yes” answer to question 26(a), which received 153 “Yes” responses. Of 153 possible responses, 125 participants provided 214 responses to this free-form question permitting multiple responses.

Respondents provided either the title of regulations, guidelines, best practices, and/or standards used by their organizations during digitization, or the category/subject of guidelines used (such as “copyright” or “digital preservation”), without reference to a specific guideline. The table below provides a summary of responses by type/subject of guidelines. Not surprisingly, the type of guideline most often referred to is digitization (30.8%). Within the category of digitization, some respondents referred to particular sub-topics, such as for example file editing, file format standards, or file resolution standards. Many participants named an entity or authority, such as the Library of Congress or Library and Archives Canada, often without reference to a specific

type of guideline. This included references to the respondents' own organizations. These types of responses were categorized under "Entity/Authority" (29%). Metadata guidelines (12.1%) received the 3rd most frequent citations, followed by digital preservation (9.8%) and archival laws and standards (4.2%). We have included a table each for responses by subject and responses by title of guideline/name of entity.

Q26(b) Please briefly list the most pertinent regulations, guidelines, best practices, and/or standards.		
Responses by Type of Regulation, Guideline, Best Practices, or Standards	Count	%
Digitization - subtopics when cited: file formats (4), file resolution (4), file editing (2), file versions (1)	66	30.8%
Entity/Authority - types: Professional Association/Coalition (19), Internal/own institution (18), National Library (13), Non-Governmental (5), Governmental/Intergovernmental (4), Private or Public/Private (3) - <i>list of names in table below</i>	62	29.0%
Metadata	26	12.1%
Digital preservation	21	9.8%
Archival laws/standards	9	4.2%
Copyright	6	2.8%
Records management	5	2.3%
File naming	4	1.9%
Library standards	3	1.4%
File storage/backup	2	0.9%
Privacy/Data protection	2	0.9%
Not sure	2	0.9%
Data management	1	0.5%
National legislation - general	1	0.5%
Open information law	1	0.5%
Accessibility	1	0.5%
Audiovisual standard	1	0.5%
Evidence (electronic records as evidence)	1	0.5%
Total	214	100.0%

Many respondents provided the titles of the regulations, guidelines, best practices, and/or standards used by their organizations during digitization, or the names of particular entities. The Federal Agencies Digital Guidelines Initiative (FADGI) suite of guidelines were the most frequently cited (21.3%), followed by internally-produced guidelines within respondents' own institutions (10.7%), and thirdly by both the International Standards Organization and the Library of Congress (5.9% each). Guidelines specific to audiovisual materials included those by the

International Association of Sound and Audiovisual Archives (IASA) (3.6%) and Metamorfoze (1.8%), among others. The descriptive metadata standard Dublin Core and the digital preservation model OAIS Open Archival Information System were the 4th (tied with IASA) and 5th most cited guidelines (3.6% and 3%, respectively).

Q26(b) Please briefly list the most pertinent regulations, guidelines, best practices, and/or standards.		
Responses by Name of Guideline or Entity (when given)	Count	%
FADGI - Federal Agencies Digital Guidelines Initiative	36	21.3%
Internal guidelines - topics when cited: file format; metadata standards; technical specifications; accessibility, gender, race, and ethnicity guidelines; layout structuring for newspapers; quality control; collaboration with internal Digital Initiatives team	18	10.7%
ISO - International Standards Organization - ISO/TR 13028:2010 Information and documentation—Implementation guidelines for digitization of records (3); ISO 15489-1:2016 Information and documentation—Records management (2); ISO/TS 19264-1:2017 Photography—Archiving systems—Image quality analysis (1); ISO 30301:2019 Information and documentation—Management systems for records—Requirements (1)	10	5.9%
Library of Congress - digitization of newspapers from microfilm; digitization; Library of Congress Subject Headings; audio digitization; Sustainability of Digital Formats (2)	10	5.9%
IASA - International Association of Sound and Audiovisual Archives - audio production, IASA TC-03, IASA TC-04 (2), IASA TC-06 (2)	6	3.6%
Dublin Core: for descriptive metadata	6	3.6%
OAIS - Open Archival Information System	5	3.0%
Library and Archives Canada guidelines - guidelines on digitization (2)	4	2.4%
SAA - Society of American Archivists guidelines and best practices	4	2.4%
NARA - United States National Archives and Records Administration (Guidelines for Digitizing Archival Materials for Electronic Access; Technical Guidelines for Digitizing Archival Materials)	3	1.8%
Metamorfoze (Netherlands)	3	1.8%
NDSA - National Digital Stewardship Alliance - Levels of Digital Preservation	3	1.8%
DACS - Describing Archives: A Content Standard	3	1.8%
UK National Archives best practices	2	1.2%
ALCTS - Association for Library Collections and Technical Services (former division of the ALA American Library Association)	2	1.2%

DFG - Deutsche Forschungsgemeinschaft (German Research Foundation) standards	2	1.2%
NEDCC - Northeast Document Conservation Center	2	1.2%
Smithsonian Institute directives	2	1.2%
American Library Association	2	1.2%
Memoriav (Switzerland)	2	1.2%
ISAD-G - General International Standard Archival Description	2	1.2%
METS - Metadata Encoding and Transmission Standard	2	1.2%
Not sure	2	1.2%
WCAG - Web Content Accessibility Guidelines	1	0.6%
Archival laws of the State of Macedonia	1	0.6%
Library and Archives Canada Act	1	0.6%
Protection of Documents and Archives and Archival Institutions Act (Slovenia)	1	0.6%
NTSC - National Television System Committee broadcast standard	1	0.6%
Core Trust Seal (trustworthy data repositories)	1	0.6%
Digital Curation Centre	1	0.6%
Digital Preservation Coalition guidelines	1	0.6%
LOCKSS - Lots of Copies Keep Stuff Safe (Stanford University)	1	0.6%
Nestor Preservation of Digital Materials	1	0.6%
OPF - Open Preservation Foundation guidelines	1	0.6%
SHA-1 - Secure Hash Algorithm 1 (U.S. Federal Information Processing Standard)	1	0.6%
UNESCO Charter on the Preservation of the Digital Heritage	1	0.6%
DIN 15587:2019-04 - Recommendations for digitization of cinematographic film	1	0.6%
FIAF - International Federation of Film Archives Digital Statement	1	0.6%
IFLA/ICA for UNESCO: Guidelines for Digitization Projects for collections and holdings in the public domain, particularly those held by libraries and archives	1	0.6%
NISO Framework of Guidance for Building Good Digital Collections	1	0.6%
United Nations digitization guidelines	1	0.6%
Europeana standards	1	0.6%
British Library	1	0.6%
Library of Catalonia	1	0.6%
National Library of Spain	1	0.6%
Harvard best practices	1	0.6%

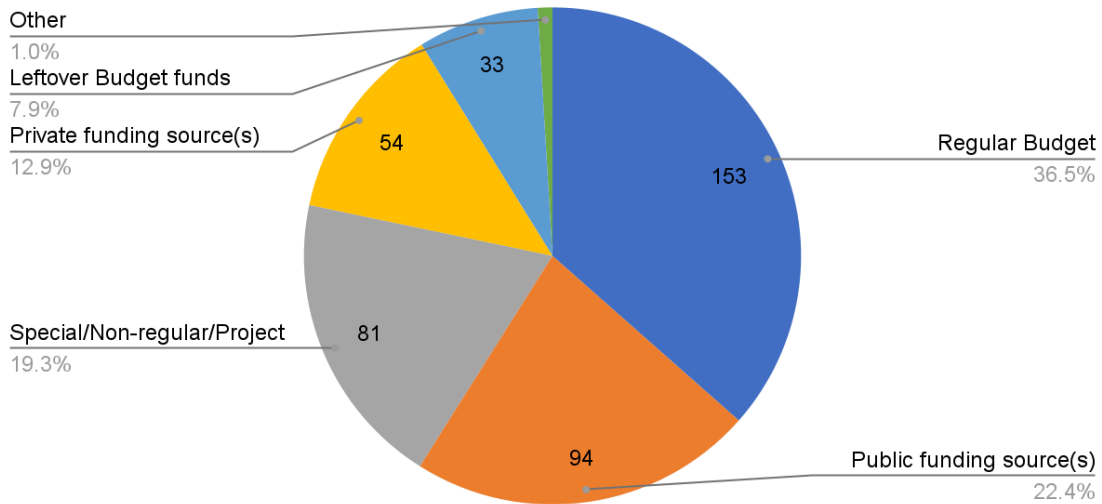
AMIA - Association of Moving Image Archivists	1	0.6%
FIAT/IFTA - International Federation of Television Archives	1	0.6%
BS10008 - Evidential Weight and Legal Admissibility of Electronic Information	1	0.6%
Standard for Digital Libraries, NN 103/2021 (Croatia)	1	0.6%
Standards of the National Library of the Czech Republic	1	0.6%
AAT - Art and Architecture Thesaurus (Getty Research Institute)	1	0.6%
ALTO - Analyzed Layout and Text Object (XML schema)	1	0.6%
DPLA - Digital Public Library of America	1	0.6%
EAD - Encoded Archival Description	1	0.6%
FITS - File Information Tool Set/MIX - Metadata for Images in XML Standard	1	0.6%
METS/ALTO (XML standards - METS + ALTO = standard for newspaper digitization)	1	0.6%
MODS - Metadata Object Description Schema	1	0.6%
PBCore Metadata Standard	1	0.6%
VRA CORE - Visual Resources Association data standard for description of visual works/images	1	0.6%
GDPR - General Data Protection Regulation	1	0.6%
Total	169	100.0%

27) What are the principal funding source(s) for digitization project(s) or program(s) in your organization? Select the top three that most apply.

Of 217 submissions considered in the overall analysis, 214 participants provided 419 responses to this multiple-choice question, for which up to three responses were permitted per participant.

The majority of respondents indicated that the organization's regular budget was their main funding source for digitization project(s) or program(s) (36%). Public funding sources (22%) and the organization's special, non-regular project funds (19%) were the second and third most selected options. The results are encouraging in that they suggest that digitization is a regularly-funded activity in many organizations. On the other hand, the results also show that 63.5% of digitization funding is sourced from outside organizations' regular budgets.

Q27. What are the principal funding source(s) for digitization project(s) or program(s) in your organization?

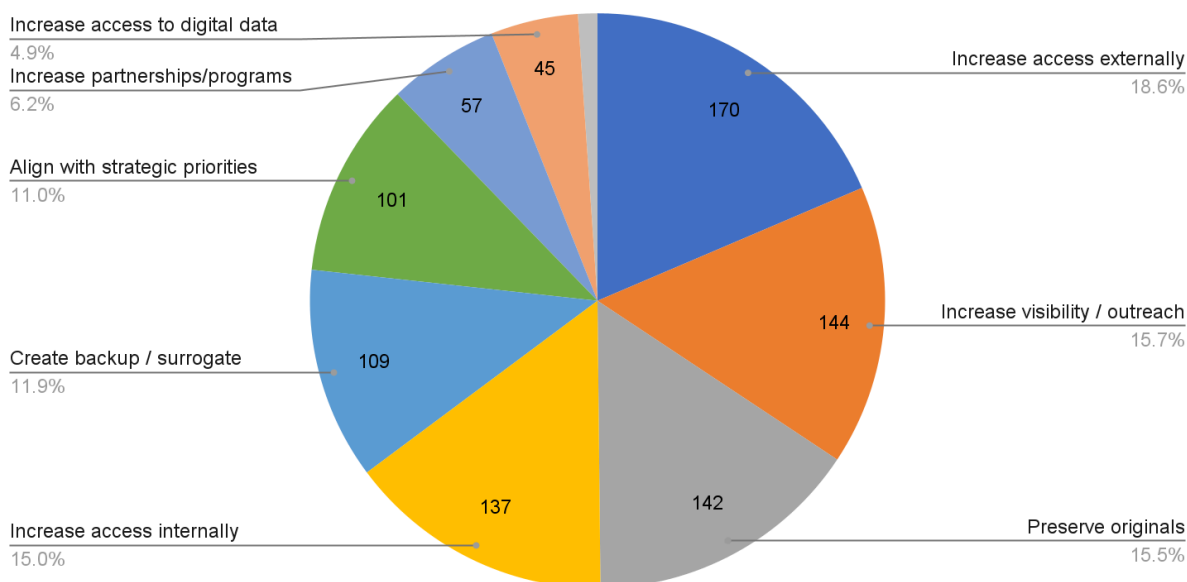


28) What are the primary motivators for digitization in your organization? Select all that apply.

Of 217 submissions included in the overall analysis, 214 participants provided 419 responses to this multiple-choice question, for which participants were asked to select all answers that apply.

Participants responded that the top motivator for digitization was “To increase access to holdings/collections for external users” (18.6%). The 2nd, 3rd and 4th top motivators were almost equally ranked in the following order: “To increase visibility of collections/outreach opportunities” (15.7%); “To preserve originals” (15.5%); and “To increase access to holdings/collections for internal users” (15%). Access for external users was ranked higher as a motivator than access for internal users. The top motivators show a focus on core archival functions, namely on access and preservation. Similarly, the top motivators speak to widely accepted benefits of digitization, highlighting again both access and preservation. A significant portion of respondents stated that digitization “...align[s] with the strategic priorities of the organization,” including digital transformation processes for institutions (11%), ranking higher than motivators related to partnerships/programs (6.2%). The “Other” responses (1.1%) are listed below the pie chart.

Q28. What are the primary motivators for digitization in your organization? Select all that apply.



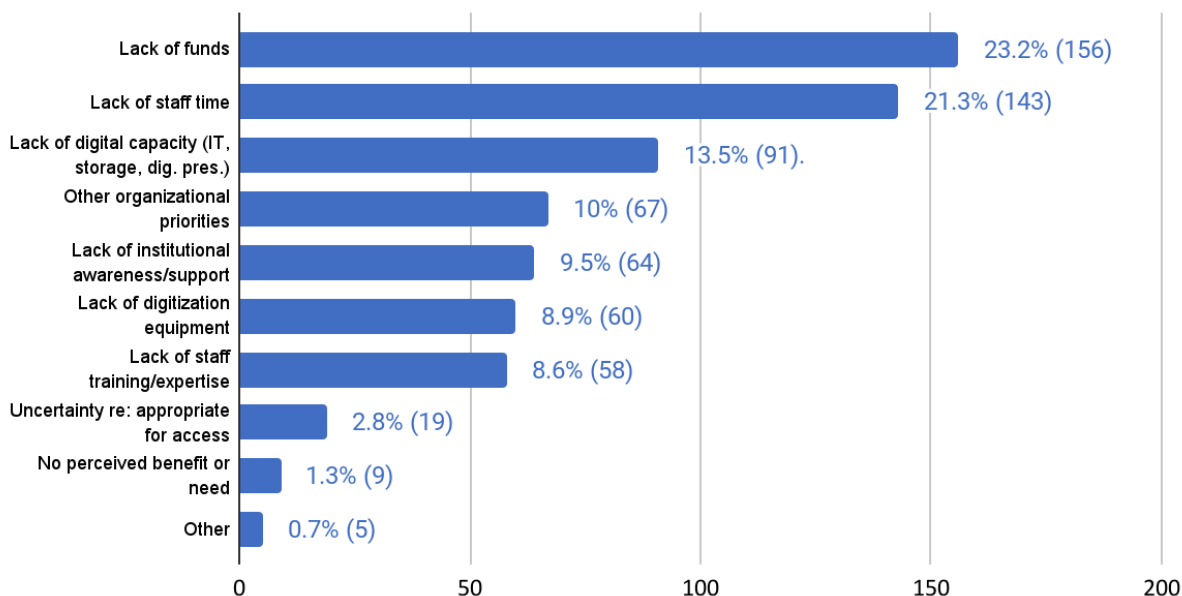
Q28. What are the primary motivators for digitization in your organization? Select all that apply.	
Other responses	Count
To enable the exercise of cultural rights	1
To capture content from obsolescent / deteriorating formats	1
To collect and preserve digital (online) legal deposit	1
For profit	1
To reduce the amount of physical storage space required	1
Preservation, especially of magnetic tape media that are quickly becoming obsolete	1
To facilitate online teaching	1
For upcoming exhibits or reference requests	1
Revival of cultural heritage	1
Funded projects or patron requests	1
Total	10

29) What are the primary barriers to digitization in your organization? Select all that apply.

Of 217 submissions included in the overall analysis, 217 participants provided 672 responses to this multiple-choice question, for which participants were asked to select all answers that apply.

The top two barriers to digitization by a significant margin were “Lack of funds” (23.2%) and “Lack of staff time to carry out digitization” (21.3%), amounting to nearly half of total responses (44.5% combined). The responses highlight that the main obstacles to digitisation are a lack of resources, both financial and human. The third most cited barrier to digitization was “Lack of digital capacity (IT infrastructure, storage, digital preservation software)” (13.5%). The next most frequently selected barriers, in 5th and 6th places, rank almost equally and are closely related: “Other organizational priorities” (10%) and “Lack of institutional awareness/support” (9.5%).

Q29. What are the primary barriers to digitization in your organization?

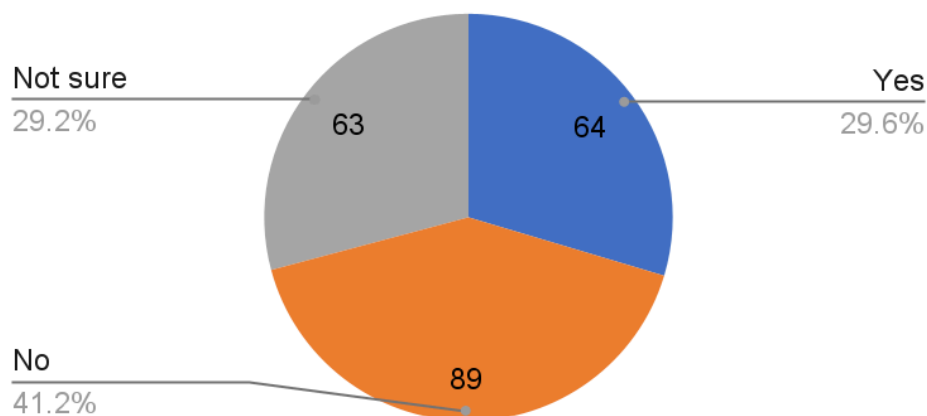


30) Has your organization identified any potential risks or drawbacks of digitization?

Of 217 submissions included in the overall analysis, this question received 216 responses.

The majority of respondents stated that their organizations have not identified any potential risks or drawbacks of digitization (41.2%), while 29.6% affirmed that their organizations have identified risks, and an almost equal number, 29.2%, responded that they are not sure.

Q30. Has your organization identified any potential risks or drawbacks of digitization?

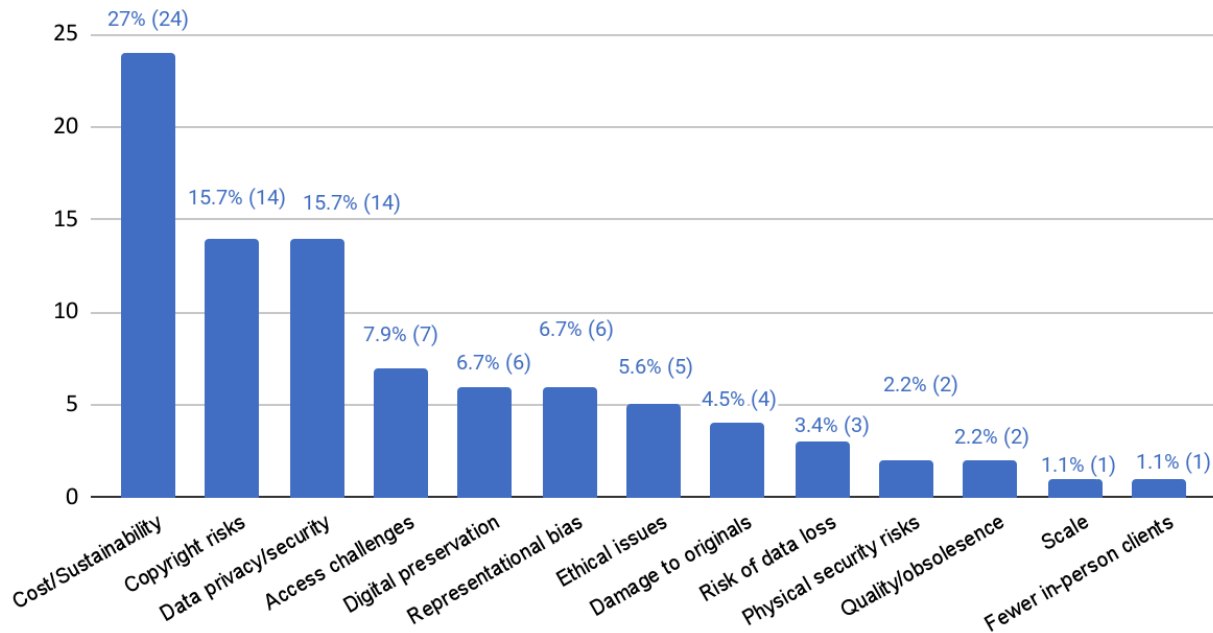


30(b) If yes, please elaborate.

This question was conditional upon a “Yes” answer to question no. 30, which received 64 “Yes” responses. 64 participants provided a total of 89 responses to this free-form question.

When asked to elaborate on the potential risks and drawbacks of digitization identified by their organizations, participants referred most often to the issue of sustainability, including costs and maintenance requirements for digitized resources (27%). They described issues such as the cost of storage, including potential failure of storage mediums; IT requirements, staff and equipment requirements, and the impact on climate change. One respondent noted that digitized files require more maintenance than physical originals. Tied for second place were copyright risks related to digitizing and publishing resources online (15.7%) and data privacy and security risks, related to breaches of potentially sensitive, private, or restricted information (15.7%). In third place was access challenges (7.9%), including the lack of an access platform for digitized content, bandwidth issues, and access barriers such as the digital divide. Tied for fourth place was the lack of a digital preservation infrastructure (6.7%) and representational bias when selecting and making digitized resources available online (6.7%). For the latter, several respondents noted issues around the gaps and silences in archives when selecting materials to digitize, or the concern that “digitization can play into the public’s perception that what is available online is all that there is to see in our holdings.” Some respondents referred to the lack of context when publishing items online.

Q30(b) If yes, please elaborate (potential risks and drawbacks of digitization).

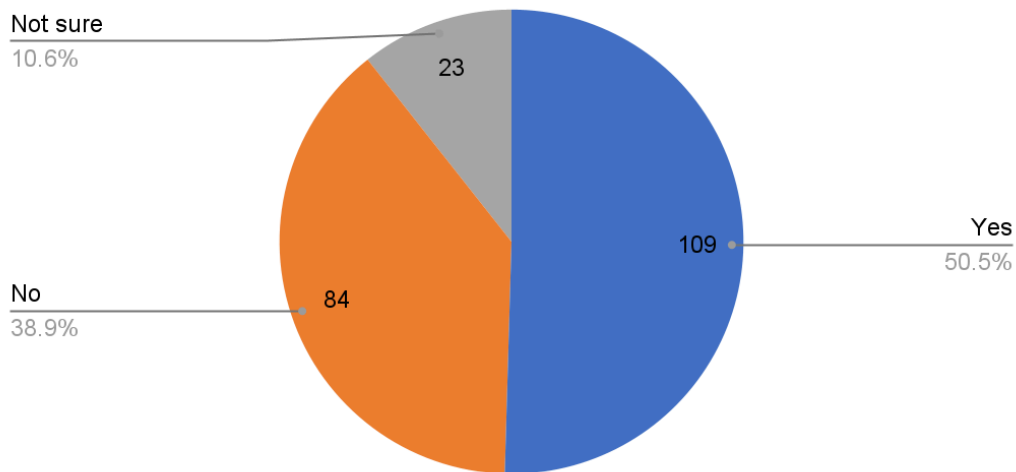


31) Is your Information Technology (IT) unit or staff involved in your organization's digitization project(s) or program(s)?

Of 217 submissions included in the overall analysis, this question received 216 responses.

50.5% of respondents affirmed that their organization's IT unit is involved in digitization projects or programs, while 38.9% stated that the IT department is not involved. 10.6% stated that they are not sure.

Q31. Is your Information Technology (IT) unit or staff involved in your organization's digitization project(s) or program(s)?

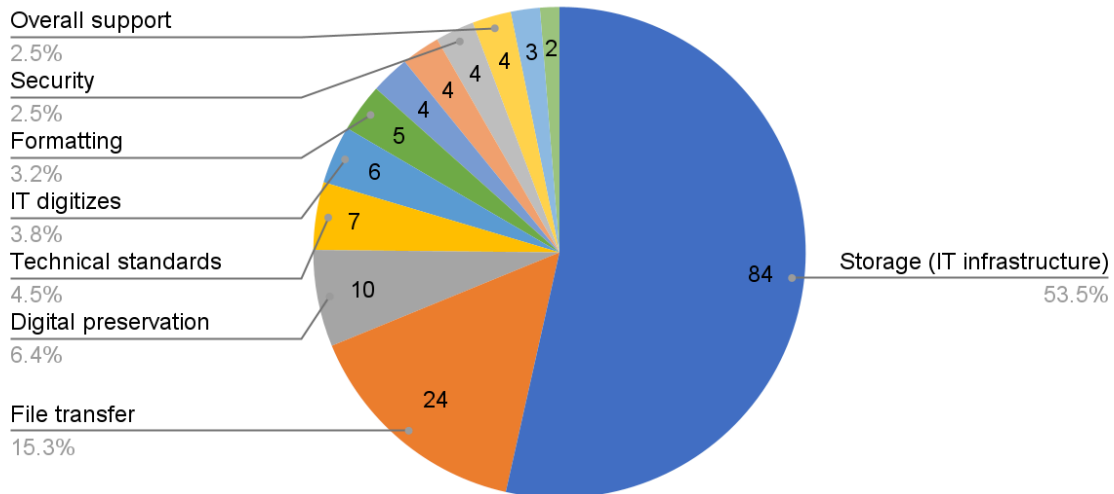


31(b) If yes, what is/was the role of the IT unit/staff in the digitization project(s) or program(s)? Please describe briefly (e.g., storage infrastructure, file transfer, support applications, formatting, etc.).

This question was conditional upon a "Yes" answer to question no. 31, which received 109 "Yes" responses. Of the 109 possible responses, 100 participants provided a total of 157 responses to this free-form question.

When asked to describe the role of IT in their organization's digitization projects or programs, the majority of respondents (53.5%) noted that IT provided storage infrastructure, including cloud services, e-repository, back-up services, and servers. This was followed by file transfer support, including data migration (15.3%). Digital preservation, including OAIS implementation and tasks supporting long-term maintenance and preservation of digital files, such as fixity checks, came in third (6.4%). This was followed by support related to setting technical standards for digitization, such as consultation on file formats and resolutions (4.5%). Perhaps surprisingly, 3.8% of respondents stated that IT undertakes the digitization work itself. Another unexpected result is that IT security services ranked very low, at only 2.5%.

Q31(b) If yes, what is/was the role of the IT unit/staff in the digitization project(s) or program(s)? Please describe briefly (e.g., storage infrastructure, file transfer, support applications, formatting, etc.).

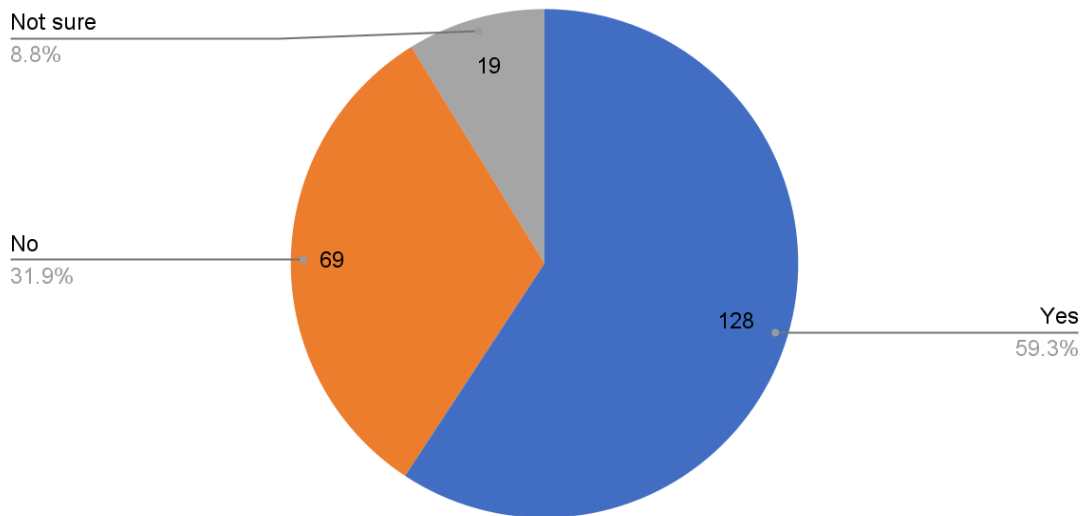


32) Are units/staff other than IT involved in your organization's digitization project(s) or program(s)?

Of 217 submissions included in the overall analysis, this question received 216 responses.

More than half of respondents reported that other units/staff besides Information Technology (IT) are involved in their organization's digitization projects or programs (59.3%). 31.9% responded "No" to this question while 8.8% responded "Not sure."

Q32. Are units/staff other than IT involved in your organization's digitization project (s) or program(s)?



32(b) If yes, please specify which units/staff and how.

This question was conditional upon a “Yes” answer to question no. 32, which received 128 “Yes” responses. Of these, 118 participants provided a total of 230 answers to this free-form question.

Respondents indicated that archivists and archives units (17%) and librarians and library units (13%) were most involved in digitization programs and projects, for a combined total of 30%. The functions provided by library and archives units include selection of materials, description, metadata creation, quality control, digital preservation, audiovisual archiving, storage, and in one case, performing the digitization/scanning. Administration and management (7%), collections management units (6.5%), conservation (5.7%), and curatorial/museum staff (5.7%) were the next most frequently mentioned units. Digitization units, including photography and imaging staff, received only 5.2% of responses. This suggests that many organizations do not have dedicated digitization staff or units.

Q32(b) If yes, please specify which units/staff and how.		
Units/Staff other than IT Involved in Digitization	Count	%
Archives (including preservation, documentation)	39	17.0%
Library (including cataloguing, special collections)	30	13.0%
Administration/Management (including inter-departmental task force, upper management/leadership, human resources, finance)	16	7.0%
Collections management unit	15	6.5%

Conservation	13	5.7%
Curatorial/Museum	13	5.7%
Digitization unit (including photo/imaging)	12	5.2%
Legal/Copyright	11	4.8%
Communications/Marketing/Outreach (including editors)	9	3.9%
Fundraising/Partnerships (including donors themselves)	6	2.6%
Academic	5	2.2%
Audiovisual unit (including AV engineering)	5	2.2%
Information management	5	2.2%
Metadata	5	2.2%
Records owners/creators	5	2.2%
Digital unit (including innovation lab, digital initiatives)	4	1.7%
Education	4	1.7%
Physical logistics unit(s) (facility, equipment)	4	1.7%
Supplementary staff (P/T, volunteers, student assistants)	4	1.7%
Access unit	3	1.3%
Database/Repository/DAMs	3	1.3%
Registry	3	1.3%
Research (including historian)	3	1.3%
External vendor	2	0.9%
Procurement	2	0.9%
Small team/Lone digitizer	2	0.9%
Ethics/Accountability	2	0.9%
Post-production	1	0.4%
Program unit	1	0.4%
Security	1	0.4%
Visual resources	1	0.4%
Translation	1	0.4%
Total	230	100.0%

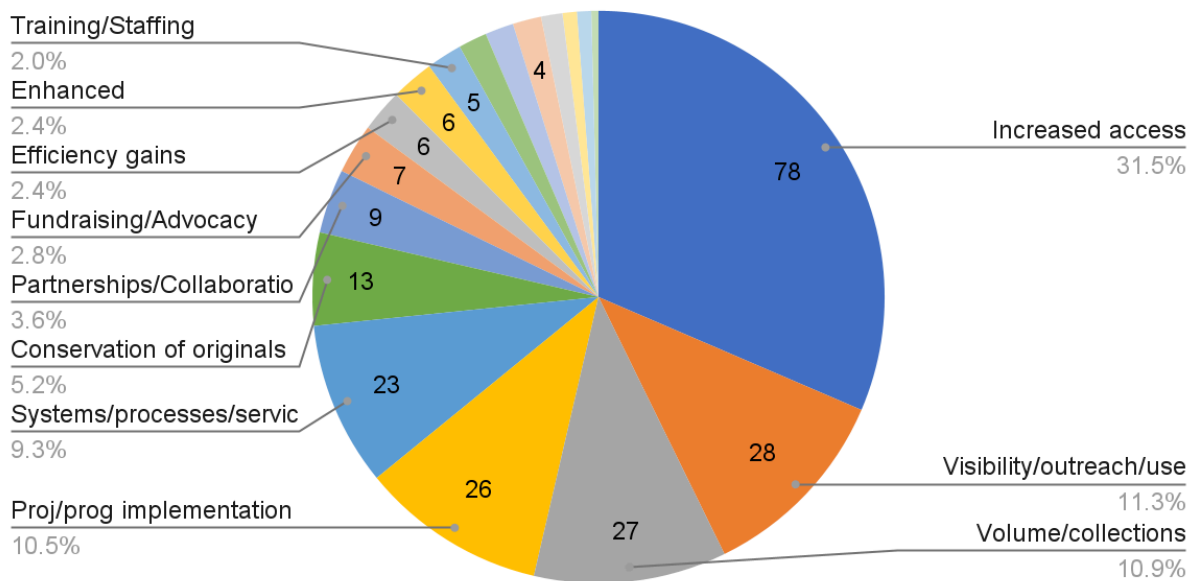
33) Please briefly describe what you consider to be the greatest success(es) of the digitization project(s) or program(s) you have been involved in.

Of 217 submissions included in the overall analysis, 173 participants provided 230 responses to this free-form question.

Respondents answered that they consider increased access to collections (31.5%) to be the greatest success of the digitization project(s) or program(s) they have been involved in. This

was followed by four types of successes that received almost an equal number of mentions: increased visibility/outreach/use of the collections (11.3%); digitization itself in terms of either the volume/scale of the digitization effort, or the nature of collections digitized (10.9%); successful implementation of the digitization program or project (10.5%); and the creation of systems, processes, or services during or as a result of digitization (9.3%). The physical conservation of originals as a result of digitization (5.2%), creation of partnerships and collaboration (3.6%), and successful fundraising and advocacy for digitization followed these responses (2.8%).

Q33. Please briefly describe what you consider to be the greatest success(es) of the digitization project(s) or program(s) you have been involved in.



Q33. Please briefly describe what you consider to be the greatest success(es) of the digitization project(s) or program(s) you have been involved in.			
	Selected Descriptions	Count	%
Increased access	Access to community/family members (x 2); Access due to negotiation of licensing contracts; Freedom of information, transparency; Positive public reaction to access; Worldwide access.	78	31.5%
Increased visibility/outreach/use	Increased use in communications - media, social media (x 3); Exhibitions (x 2); Film screenings; First-time use by researchers; Website enhancements.	28	11.3%

Digitization itself - volume/collections digitized	Digitized 100%/almost 100% of collection or archive (x 7); Mass digitization (x 2); Nation-wide program; Oral histories; Priority collections digitized.	27	10.9%
Successful project/program implementation	Project documentation (x 3); Project completion (x 3); Compliance with standards/national digitization rules (x 2); Continuity of digitization program (x 2); Good vendor relations (x 2); Model/precedent for others (x 2); Citizen archivists; Good project management - planning, specifications, selection process, workflows/streamlined operations, meeting deadlines, realistic goals; Hybrid in-house/vendor digitization; In-house digitization.	26	10.5%
Creation of systems/processes/services	Digital library/repository (x 5); On-demand digitization (x 4); Federated platform (x 3); Built open source tools (x 2); New backup storage (x 2); Streamlined processes (x 2); Born-digital; Integrated into running of institution; New equipment; Public-facing web portal; Procurement management system.	23	9.3%
Conservation of originals	Conservation in a country at risk with no stable power supply.	13	5.2%
Partnerships/Collaboration	Collaboration across units increased capacity for film digitization; Public-private partnership.	9	3.6%
Fundraising/Advocacy	Funds raised; Increased awareness of importance of digitization in own institution.	7	2.8%
Efficiency gains	Cost-savings of using in-house tools and expertise; Enable online teaching to continue; Modernization; Paperless; Reduce physical storage after destruction.	6	2.4%
Enhanced searchability/analysis of digital collections	OCR (x 2); Collections as data; Use of AI.	6	2.4%
Training/Staffing	Expertise gained; Staff retention/hiring; Training.	5	2.0%
Increased control over collections	Cataloguing; Condition assessment; Inventory; New knowledge of collections.	4	1.6%
Other	Still in early stages (x 2); Not involved in digitization; Still awaiting funding.	4	1.6%

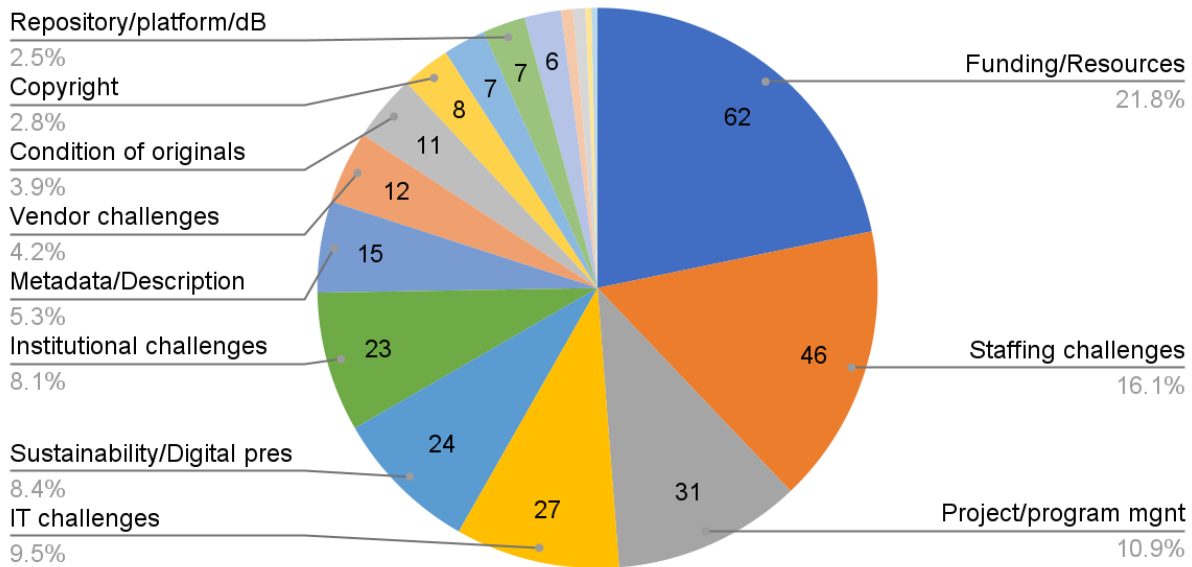
Quality of digitization		4	1.6%
Digital preservation		3	1.2%
Accessibility	Gender expansive and transgender friendly metadata; WCAG (Web Content Accessibility Guidelines) compliant.	2	0.8%
Ethical value	Returning photographs to communities; Supporting diversity.	2	0.8%
Security		1	0.4%
Total		248	100.0%

34) Please briefly describe what you consider to be the greatest challenges of the digitization project(s) or program(s) you have been involved in.

Of 217 submissions included in the overall analysis, 174 participants provided 285 responses to this free-form question.

The greatest challenge experienced by respondents with regard to digitization projects or programs is inadequate funding and resources, including equipment, space, and lack of time (21.8%). Some respondents lamented the lack of a regular or continuous funding source and the need to rely on grants. A significant number of respondents noted the lack of time and the amount of time needed for digitization programs/projects. The second most mentioned challenge is staffing (16.1%), which could be considered a type of resource challenge. The high number of responses mentioning staffing, however, merited its own category. Staffing challenges related principally to staff shortages, staff retention and turnover, lack of expertise or training, and work overload. The next four challenges were rated similarly: challenges related to various aspects of Program/Project management, such as the selection process, quality control, migration, coordination and planning, among others (10.9%); IT challenges, including storage and IT infrastructure challenges (9.5%); Digital preservation and sustainability challenges (8.4%); and Institutional challenges (8.1%). Many respondents stated a lack of institutional support for digitization, including from senior management, as well as conflicting priorities. Significantly, some respondents noted a lack of understanding within their institutions of the resource-intensive nature of digitization. The responses point in general to the fact that digitization is time-consuming and requires resources, staff, and expertise for its realization. Additionally, while it is positive that sustainability and long-term preservation are recognized as an important aspect of digitization, it is clear from the responses that these are not always planned for within institutions that digitize their collections.

Q34. Please briefly describe what you consider to be the greatest challenges of the digitization project(s) or program(s) you have been involved in.



Q34. Please briefly describe what you consider to be the greatest challenges of the digitization project(s) or program(s) you have been involved in.			
Greatest Challenges	Selected Descriptions	Count	%
Funding/Resource challenges	Lack of time/underestimation of time (x 16); Lack of equipment/breakdown (x 6); Lack of regular/long-term funding (x 4); Obtaining funds, e.g., to hire external vendors (x 3); Space (x 2); Grant paperwork.	62	21.8%
Staffing challenges	Lack of expertise, knowledge, training, e.g. curatorial staff don't know what to do with digitized assets, knowledge of legacy equipment, digitization training (x 11); Lack of staff/lone staffer (x 8); Staff retention/turnover (x 7); Work overload (x 3); Planning for and maintaining staff.	46	16.1%

Project/program management	Selection (x 8); Coordination/Planning (x 5); Preparation; Quality control including finding time for it (x 5); Migration (x 3); Workflows issues, e.g. complex workflows for checksums, inefficient legacy workflows (x 3); Ad-hoc digitization/Lack of overall strategy (x 2); Following policies/procedures (x 2); Digitization before cataloguing; Documentation; ; File naming conventions; Managing multiple projects; Remote processing; Switching from vendor to in-house; Writing requirements.	31	10.9%
IT challenges	Storage - cost, space, location of server, backups (x 11); Lack of IT infrastructure (x 5); Handling large volumes on old equipment; Keeping hardware/software updated or working, e.g., digitization software (x 3); Lack of IT support/staff (x 3); Internet speed and bandwidth.	27	9.5%
Digital preservation/Sustainability	Adapting to evolving file formats (x 3); Implementing best practices e.g., OAIS (x 2); Lack of long-term planning (x 2); Long-term maintenance of digital assets (x 2); Archiving emails; Fixity checks; Trusted digital repository.	24	8.4%
Institutional challenges	Lack of understanding/buy-in from senior management, e.g., that it is labour-intensive and time-consuming (x 8); Conflicting priorities (x 7); Lack of support (x 6); Policy.	23	8.1%
Metadata/Description challenges	Creation of metadata esp. for large volume (x 2); Indexing and structuring records (x 2); Multilingual content (x 2); Lack of intellectual control over collections; Linking digital files to catalogue; Moving to item-level description; Time-consuming to create it; Transcription for AV; Understanding need for good metadata and implementing it.	15	5.3%
Vendor challenges	Need to outsource, e.g., for AV digitization (x 2); Vendor education/training, e.g., ensuring detail and care required (x 2); Changing requirements from clients; Errors made in file naming; Institution does not trust vendor; Multiple vendors; Procurement;	12	4.2%

	Promises made vs. reality of what they deliver; Takes longer; Risk of collections leaving premises.		
Condition of originals	Degradation/Technological obsolescence of physical carriers (x 5); Blurry originals; Content exists in multiple formats and copies; Film sound elements; Fragile films; Handling originals; To bake or not bake videos.	11	3.9%
Copyright		8	2.8%
Access and accessibility	Lack of access platform (x 2); Not everything scanned can be public (x 2); Access to information requests; Facilitating access; Getting health professionals onboard viewing digital instead of paper documents; HTR and remediated documents for accessibility; Researchers thinking what is online is all there is.	7	2.5%
Repository/platform/database challenges	Lack of access platform (x 2); Building reliable storage and access system; Creation of apps; Lack of understanding of DAM requirements; Save digital library.	7	2.5%
Scale of digitization	Overwhelming volume still needed to digitize; Size of collection in multiple distant locations.	6	2.1%
Other	Not involved in digitization (x 2).	2	0.7%
Outreach		2	0.7%
Data loss		1	0.4%
Security concerns	Digitizing classified materials via vendor because cannot digitize in-house.	1	0.4%
Total		285	100.0%

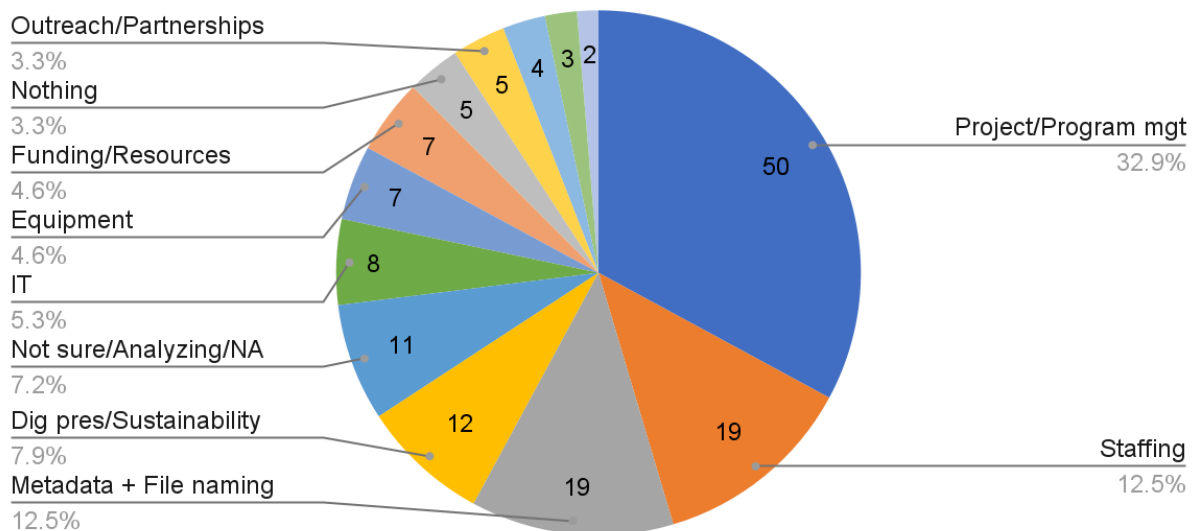
35) What, if anything, would you do differently in your next digitization project(s) or program(s)? Please briefly describe.

Of 217 submissions included in the overall analysis, 136 participants provided 152 responses to this free-form question.

Participants most often expressed that they would change various aspects of project/program management in their next digitization efforts (32.9%). Project/Program management includes planning and timeline, scope, selection of materials to be digitized, policy, and strategic or methodological approaches. Respondents stated that they would inventory or understand collections better before digitization; build in more time for different phases of digitization; have more defined policies for selection of materials to be digitized; and understand required

resources and establish clearer tasks and outcomes, among other aspects. Next to that, participants stated most often that they would change their approach to Staffing (12.5%), by hiring more dedicated staff and providing more training for staff, for example. Metadata and File naming (12.5%) received the same number of responses as staffing. Respondents stated they would, for instance, change their approach to ensure that metadata is correct, have clearer requirements for metadata, or that they would introduce more structured metadata. On the lower end of the scale, some respondents stated they would change nothing (3.3%), suggesting both that there are some individuals who are completely satisfied with their digitization projects and programs, but also that such instances are relatively rare. Few respondents stated they would change the Funding/Resources at their disposal (4.6%), compared to those who noted in the previous question that Funding/Resources are a significant challenge to digitization (21.8%). This result suggests perhaps that respondents do not feel they are in control of this aspect of digitization.

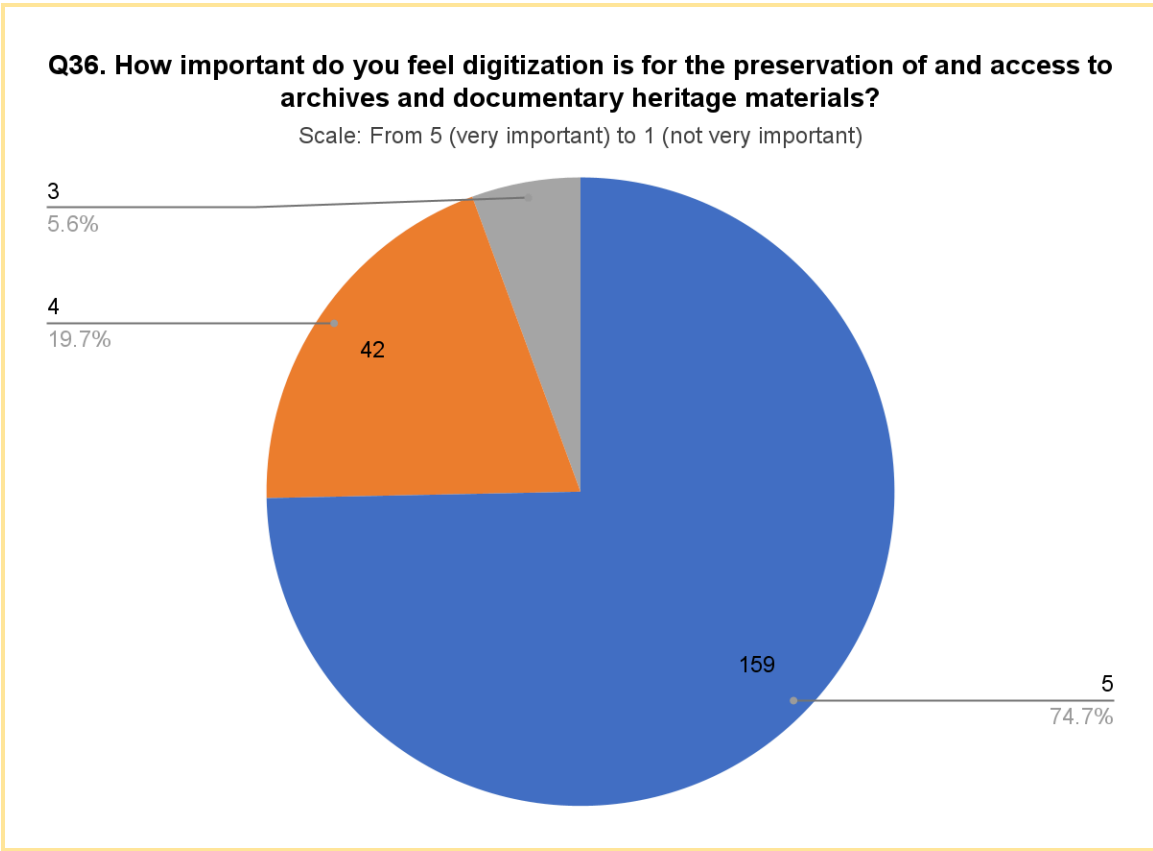
Q35. What, if anything, would you do differently in your next digitization project(s) or program(s)? Please briefly describe.



36) How important do you feel digitization is for the preservation of and access to archives and documentary heritage materials?

Of 217 submissions included in the overall analysis, this question received 213 responses.

Respondents overwhelmingly responded that digitization is “5 (very important)” for the preservation of and access to archives and documentary heritage materials (74.7%). The remaining quarter of respondents ranked digitization as “4” and “3” on the provided Lickert scale, with no responses lower than “3.”



[Back to Table of Contents](#)

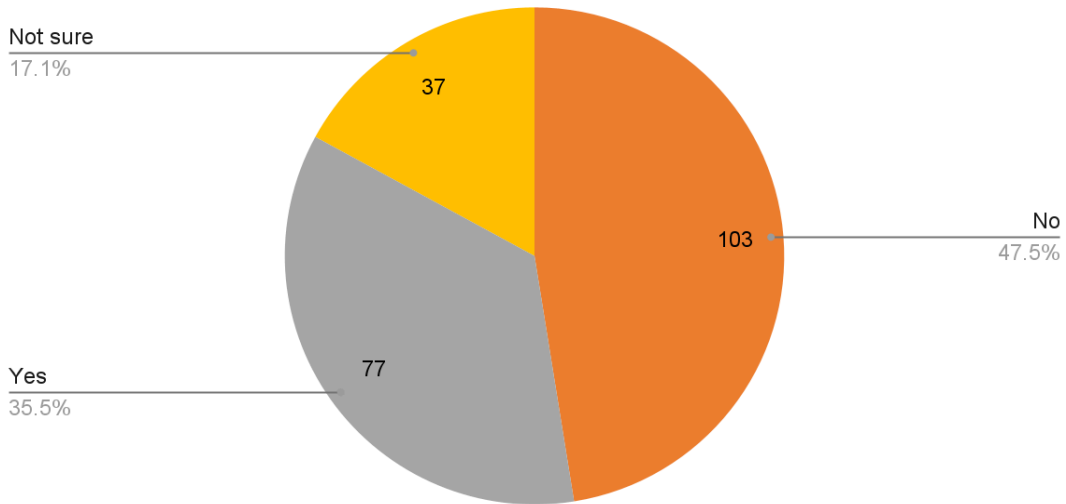
5.4. Section 4: Digitization and AI

37) Has your organization considered using AI models or tools during digitization project(s) or program(s)?

Of 217 submissions included in the overall analysis, this question received 217 responses.

Just under half of respondents' organizations have not considered using AI models or tools during digitization projects (47.5%), while a significant number have (35.5%). Another portion of respondents were not sure (17.1%).

Q37. Has your organization considered using AI models or tools during digitization project(s) or program(s)?

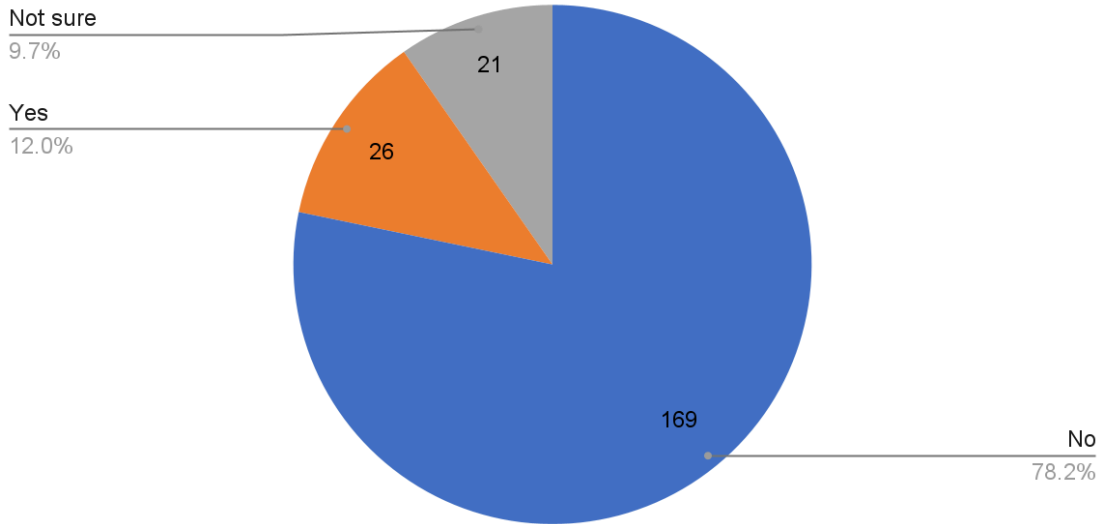


38) Has your organization used in-house, custom-developed AI models or tools during digitization project(s) or program(s)?

Of 217 submissions included in the overall analysis, this question received 217 responses.

A clear majority of respondents stated that their organization has not used in-house, custom-developed AI models or tools during digitization projects or programs (78.2%). A small percentage of respondents affirmed that their organization has used them (12%), while 9.7% were not sure.

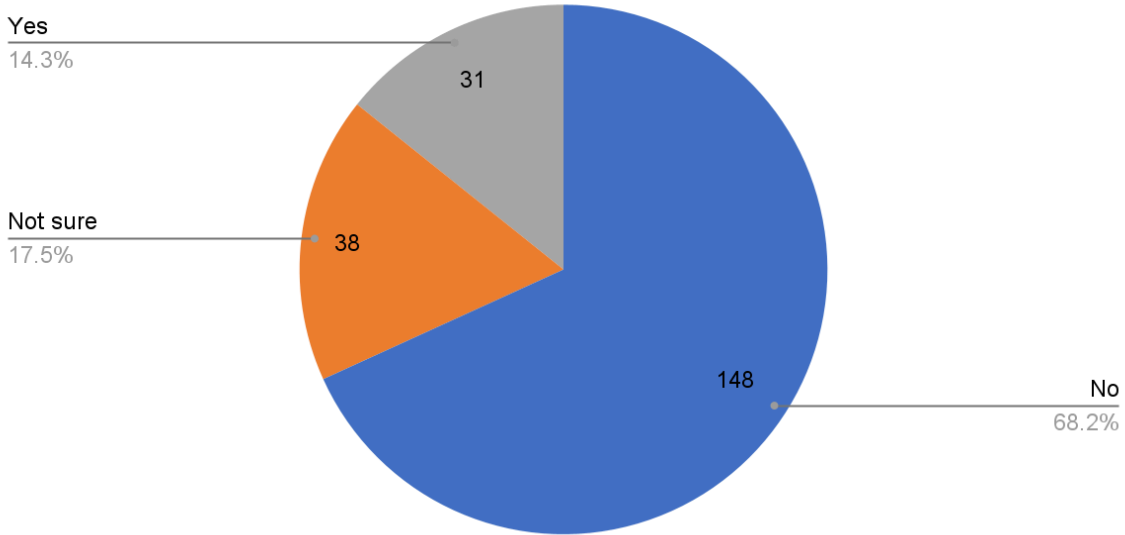
Q38) Has your organization used in-house, custom-developed AI models or tools during digitization project(s) or program(s)?



39) Has your organization used off-the-shelf - ready-made as opposed to in-house, custom-developed - AI tools during digitization project(s) or program(s) (e.g., open source AI models, Microsoft Azure Facial Recognition, Google Cloud Vision AI, etc.)?
Of 217 submissions included in the overall analysis, this question received 217 responses.

In comparison to question 38, a smaller majority of respondents reported that their organizations do *not* use off-the-shelf AI tools (68.2%), while slightly more respondents stated that their organizations *do* use off-the-shelf AI products (14.3%). A significantly larger portion of respondents as compared to question 38 stated that they do not know if their organization uses off-the-shelf AI products (17.5%). This could be related to the fact that in some cases, tools are powered by AI without the client being aware of it.

Q39. Has your organization used off-the-shelf - ready-made as opposed to in-house, custom-developed - AI tools during digitization project(s) or program(s)?



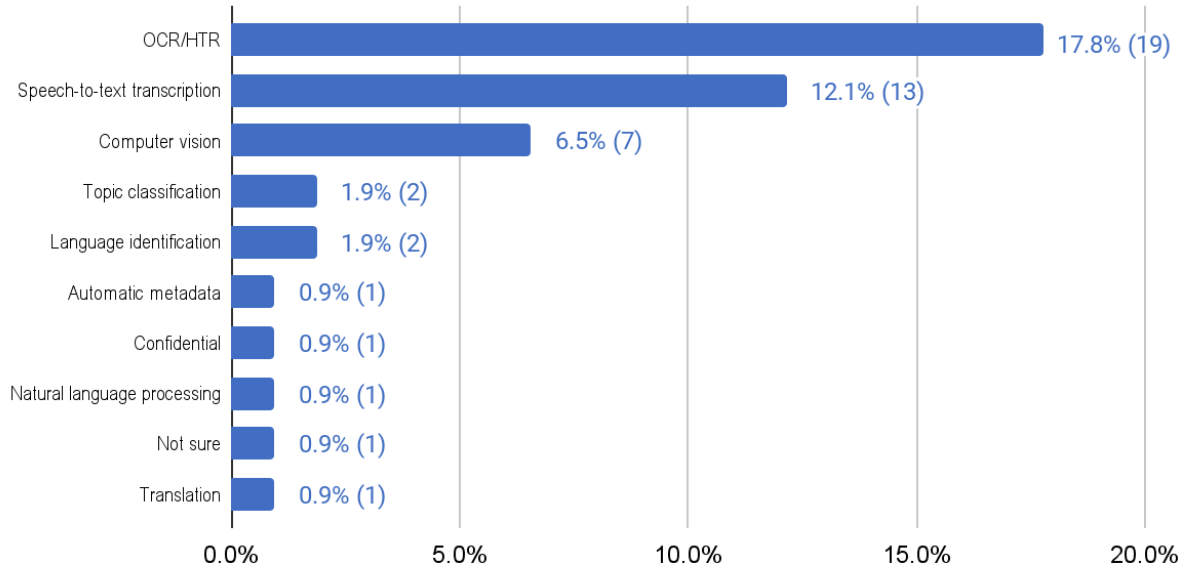
40) Please briefly describe the AI models/tools your organization has used, including for what digitization-related activities in general.

This question was conditional upon a “Yes” answer to either question no. 38 or no. 39, which received a total of 57 “Yes” responses. Of these, 36 participants provided a total of 107 answers to this free-form question.

Participants responded with answers that fell into three distinct categories: the Type of Machine Learning, the Name of the AI Tool itself (proprietary or open source), or the Digitization Activity Assisted by AI. We therefore created a separate table for each of the three categories of responses. For the Type of Machine Learning, participants indicated most often that their organizations use OCR/HTR (17.8%), Speech-to-text transcription (12.1%), and Computer vision (6.5%) for digitization-related activities. For the Name of the AI Tool, respondents most frequently named Open AI’s ChatGPT and Tesseract (3.7% each), followed by Transkribus (2.8%), and Adobe Acrobat, Google, or an in-house tool (1.9% each). Regarding the Digitization Activity Assisted by AI, participants most often described using AI for metadata creation, image processing, writing code, or searching/filtering metadata (1.9% each).

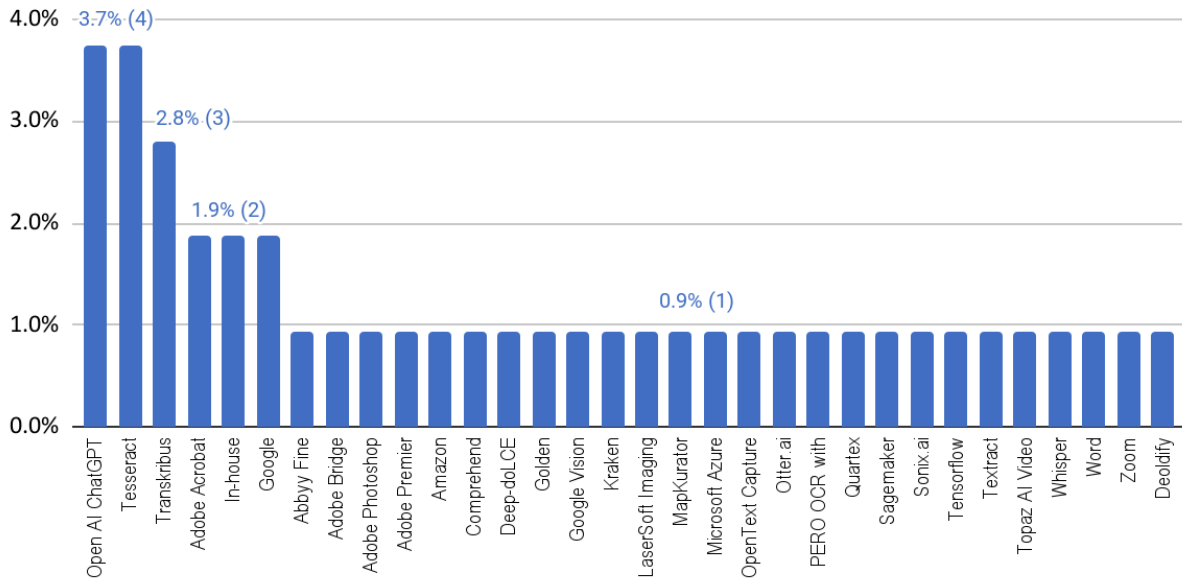
Q40. Please briefly describe the AI models/tools your organization has used, including for what digitization-related activities in general.

Category: By Type of Machine Learning



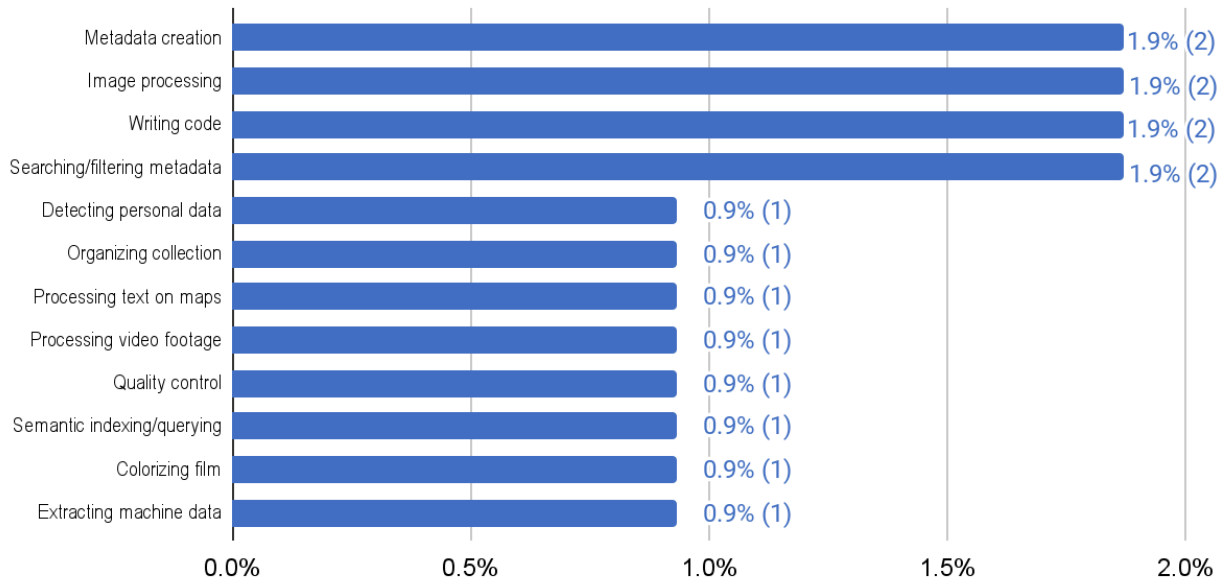
40. Please briefly describe the AI models/tools your organization has used, including for what digitization-related activities in general.

Category: By Name of AI Tool



40. Please briefly describe the AI models/tools your organization has used, including for what digitization-related activities in general.

Category: By Digitization Activity Assisted by AI



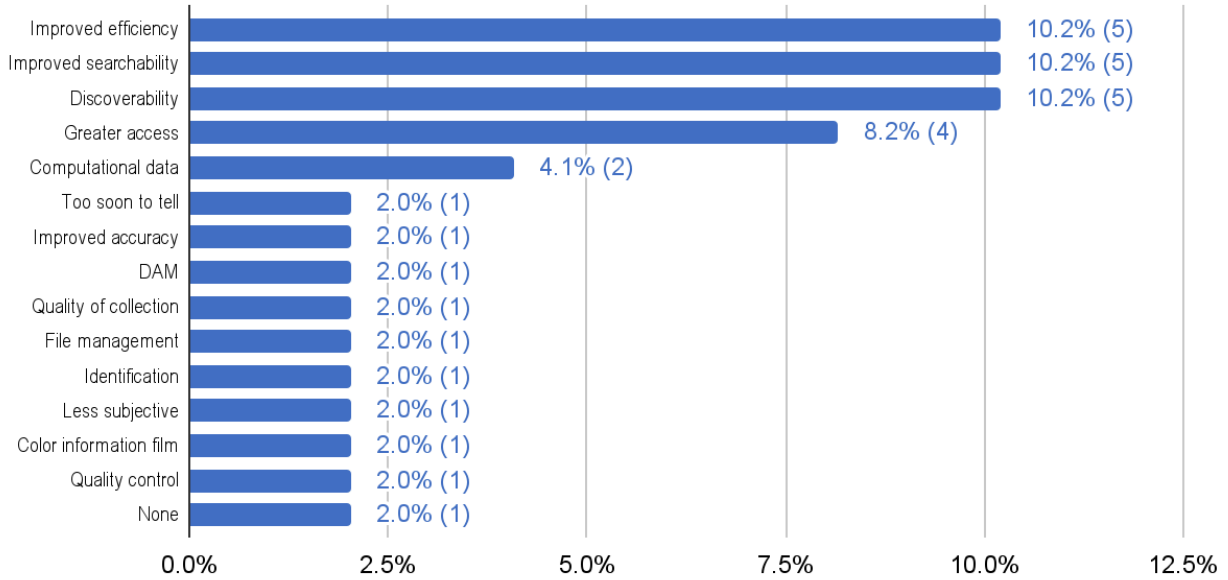
41) What aspects of the AI tool(s) used helped to achieve the digitization objectives/processes? Please briefly describe.

This question was conditional upon a “Yes” answer to either question no. 38 or no. 39, which received a total of 57 “Yes” responses. Of these, 31 participants provided a total of 49 answers to this free-form question.

Participants responded with answers that fell into two distinct categories: Positive Results/Impact of using AI tools for digitization, or Type of Machine Learning with Positive Impact. We therefore created a separate table for each of the two types of responses. Regarding the Positive Results/Impact of using AI tools, these included improving efficiency, searchability, and discoverability (10.2% each). AI tools also facilitated greater access (8.2%) and the creation of computational data (4.1%). Regarding the Type of Machine Learning with Positive Impact, respondents mentioned that transcription/speech-to-text was most positively impactful (14.3%), followed by OCR/HTR (12.2%), automatic metadata (4.1%), computer vision (4.1%), and finally neural networks (2%).

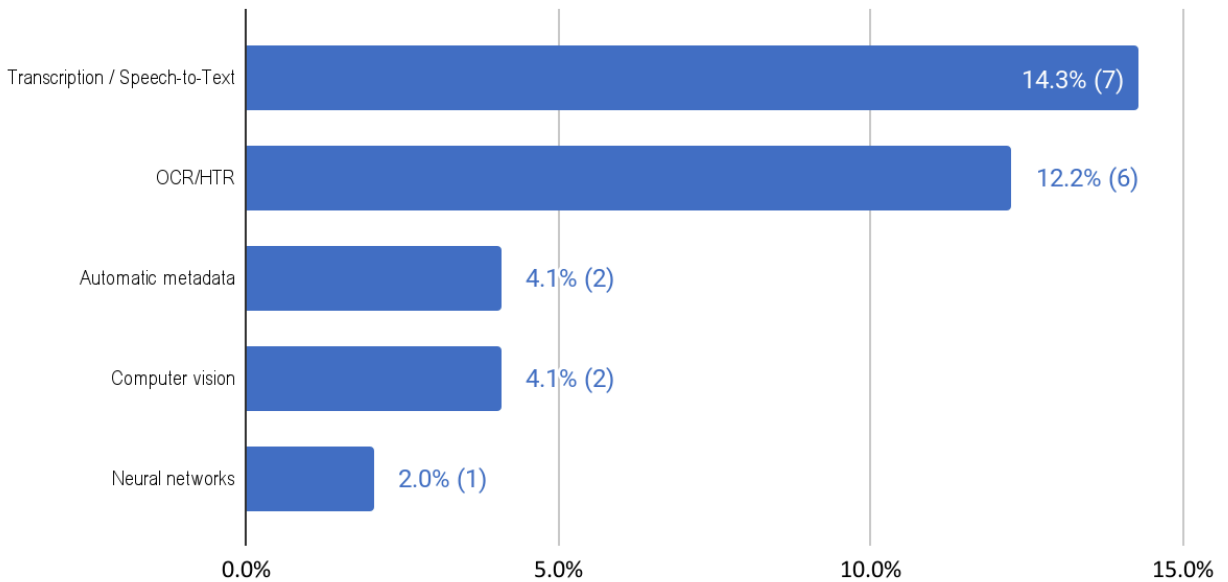
**41. What aspects of the AI tool(s) used helped to achieve the digitization objectives/processes?
Please briefly describe.**

Category: Positive Results/Impact of AI Use



**41. What aspects of the AI tool(s) used helped to achieve the digitization objectives/processes?
Please briefly describe.**

Category: Type of Machine Learning with Positive Impact

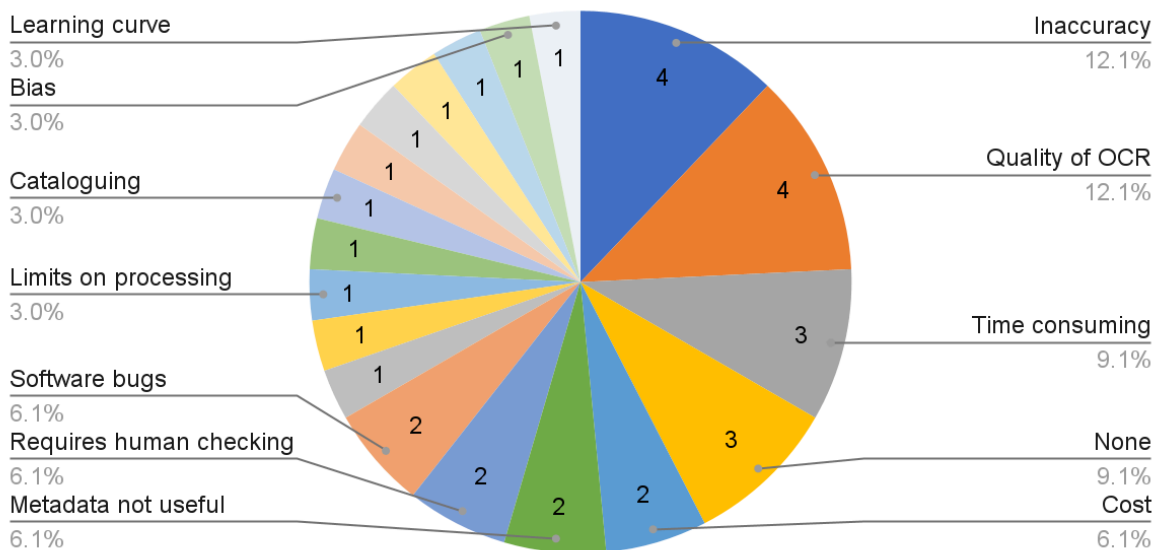


42) What aspects of the AI tool(s) used restricted or negatively affected the digitization objectives/processes? Please briefly describe.

This question was conditional upon a “Yes” answer to either question no. 38 or no. 39, which received a total of 57 “Yes” responses. Of these, 28 participants provided a total of 33 answers to this free-form question.

Participants mentioned various aspects of AI that negatively affected digitization, including: inaccuracy (12.1%), the quality of OCR (12.1%), the time required to implement AI (9.1%), the cost of AI (6.1%), the fact that the metadata generated is not useful (6.1%), that AI requires human checking (6.1%), and software bugs (6.1%). Interestingly, a comparatively significant number of participants reported no negative effects of the AI tool(s) on digitization (9.1%).

42. What aspects of the AI tool(s) used restricted or negatively affected the digitization objectives/processes? Please briefly describe.



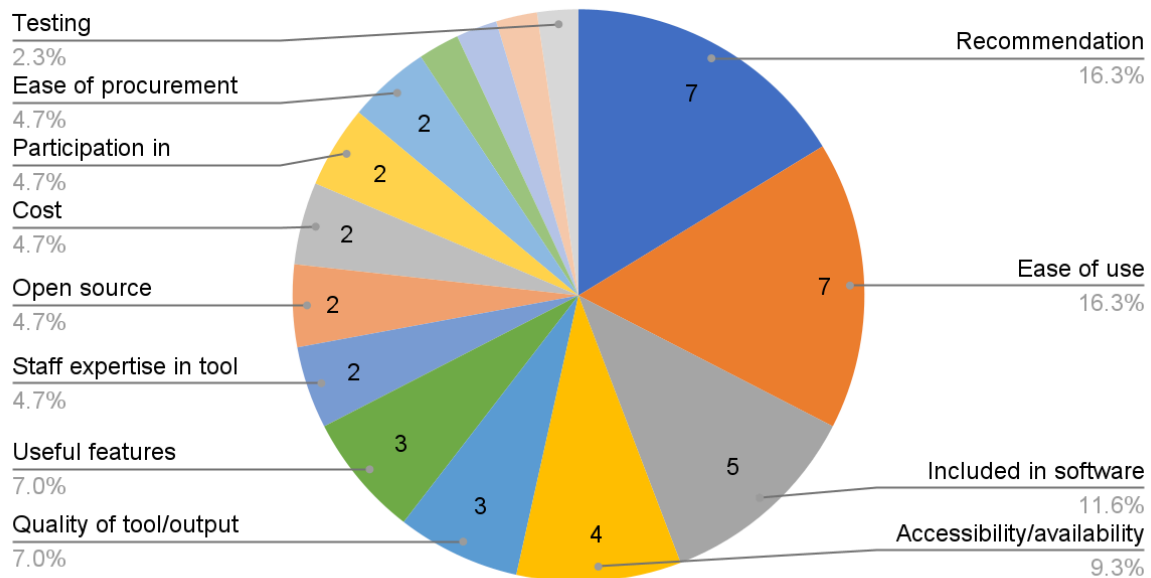
43) Please describe how the AI tool(s) used were selected (e.g. quality of the data used to train the model(s), ease of use, ease of procurement, availability of information about the capabilities and limitations of the tool, recommended by archival/IT colleagues, etc.).

This question was conditional upon a “Yes” answer to either question no. 38 or no. 39, which received a total of 57 “Yes” responses. Of these, 29 participants provided a total of 45 answers to this free-form question.

When asked how the AI tools used were selected, an equal number of respondents stated that the AI tools were either recommended or that they were selected based on ease of use (16.3%

each). The third-most cited reason was simply that the AI tools were included in software (11.6%). Following this were the accessibility/availability of the tool (9.3%), the quality of the tool or its output (7%), and the useful features of the tool (7%).

43. Please describe how the AI tool(s) used were selected.



44) Please list any (other) off-the-shelf AI tools that you are aware of/can recall, even if your organization has not used them.

Of 217 submissions included in the overall analysis, 79 participants provided 130 responses to this free-form question.

The top response to this question was some form of “Don’t know” or “Not applicable” (17.7%). Following that, respondents most often mentioned Open AI’s ChatGPT (11.5%), the HTR tool Transkribus (7.7%), Google Bard, the chatbot now known as Gemini (3.8%), the OCR tool ABBY FineReader (2.3%), Bing (without specifying which AI tool) (2.3%), and the text-to-image tool DALL-E by OpenAI (2.3%). The below table lists all tools mentioned, providing a snapshot of the AI landscape at the time the survey was active.

Q44. Please list any (other) off-the-shelf AI tools that you are aware of/can recall, even if your organization has not used them.		
Off-the-shelf tool	Count	%
None/Don't know/NA	23	17.7%
Open AI ChatGPT	15	11.5%
Transkribus	10	7.7%

Google Bard	5	3.8%
ABBYY FineReader	3	2.3%
Bing	3	2.3%
DALL-E	3	2.3%
Deep-L	2	1.5%
Microsoft Word	2	1.5%
Midjourney	2	1.5%
OCR/HTR	2	1.5%
Speech-to-text	2	1.5%
Stable Diffusion	2	1.5%
Tesseract	2	1.5%
Too many to list	2	1.5%
Amazon Transcribe	1	0.8%
Amazon Textract	1	0.8%
Annif AI	1	0.8%
BERT	1	0.8%
Chatbot	1	0.8%
Comprehend	1	0.8%
Confidential	1	0.8%
Copilot	1	0.8%
Crayon	1	0.8%
Davinci Resolve	1	0.8%
Expert System's Cogito	1	0.8%
Express Scribe	1	0.8%
Face recognition software	1	0.8%
FILMIC Tech	1	0.8%
Generative AI	1	0.8%
Google Document Recovery	1	0.8%
Google Lens	1	0.8%
HAL 9000	1	0.8%
HS-Art Diamant	1	0.8%
Futuretools.io	1	0.8%
IBM Watson	1	0.8%
Image creation	1	0.8%
Internet Archive OCR	1	0.8%
iWave	1	0.8%

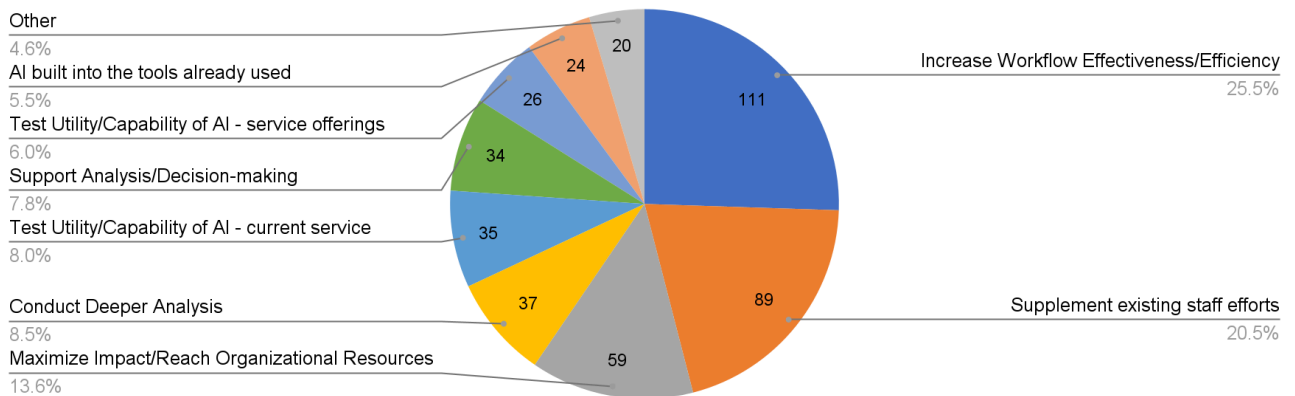
KAITOS software from EXON company	1	0.8%
Lensa	1	0.8%
Libchat	1	0.8%
LINDAT/CLARIAH-CZ tools	1	0.8%
MALLET	1	0.8%
Microsoft Azure	1	0.8%
Microsoft Purview	1	0.8%
Microsoft transcription	1	0.8%
MSU Deinterlacer	1	0.8%
NovelAI	1	0.8%
Open AI Whisper	1	0.8%
oTranscribe	1	0.8%
Otter.ai	1	0.8%
Partnership with other institution	1	0.8%
PimEyes	1	0.8%
Pixop	1	0.8%
Project Sheeko	1	0.8%
Restoration only	1	0.8%
Rev.com	1	0.8%
Sagemaker	1	0.8%
Temi	1	0.8%
Text recognition software	1	0.8%
Topazlabs	1	0.8%
Translation	1	0.8%
Trevi	1	0.8%
Video transcription	1	0.8%
Virtual Bench	1	0.8%
YouTube	1	0.8%
Total	130	100.0%

45) What are the primary motivators for using AI in your organization's digitization project(s) or program(s)? Select all that apply.

Of 217 submissions included in the overall analysis, 168 participants provided 435 responses to this multiple-choice question.

Participants reported that the top motivator for using AI in their organizations is to “Increase effectiveness and efficiency of workflows” (25.5%), followed by “Supplement existing staff efforts” (20.5%) and “Maximize the impact and/or reach of organizational resources” (13.6%). The next most cited reasons included “Conduct a deeper analysis than humanly possible” (8.5%), “Test the utility/capability of AI with regards to current service offering” (8%), and “Support analysis and decision-making” (7.8%). The top three answers point to the desire of organizations to use AI to increase efficiency and support human staff in their work.

Q45. What are the primary motivators for using AI in your organization's digitization project(s) or program(s)?
Select all that apply.

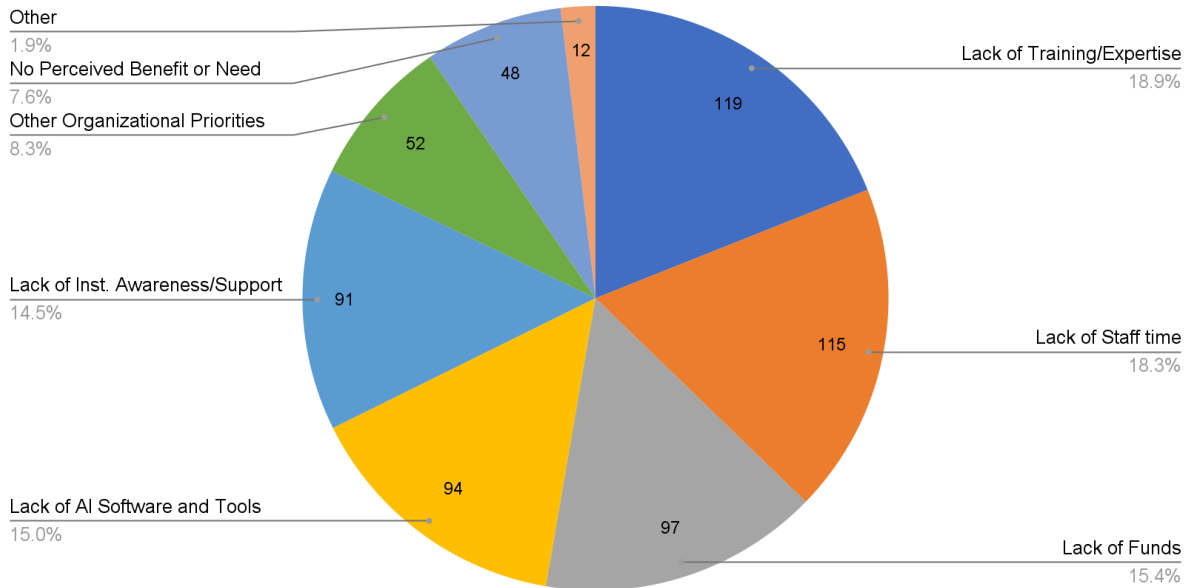


46) What are the primary barriers to using AI in your organization's digitization project(s) or program(s)? Select all that apply.

Of 217 submissions included in the overall analysis, 201 participants provided 628 responses to this multiple-choice question.

Respondents indicated that the top two primary barriers to using AI in their organizations are “Lack of training/expertise to use AI” (18.8%) and “Lack of staff time to integrate/use AI” (18.3%). These were followed by “Lack of funds” (15.4%), “Lack of AI software and tools” (15%), and “Lack of institutional awareness/support” (14.5%).

Q46. What are the primary barriers to using AI in your organization's digitization project(s) or program (s)? Select all that apply.

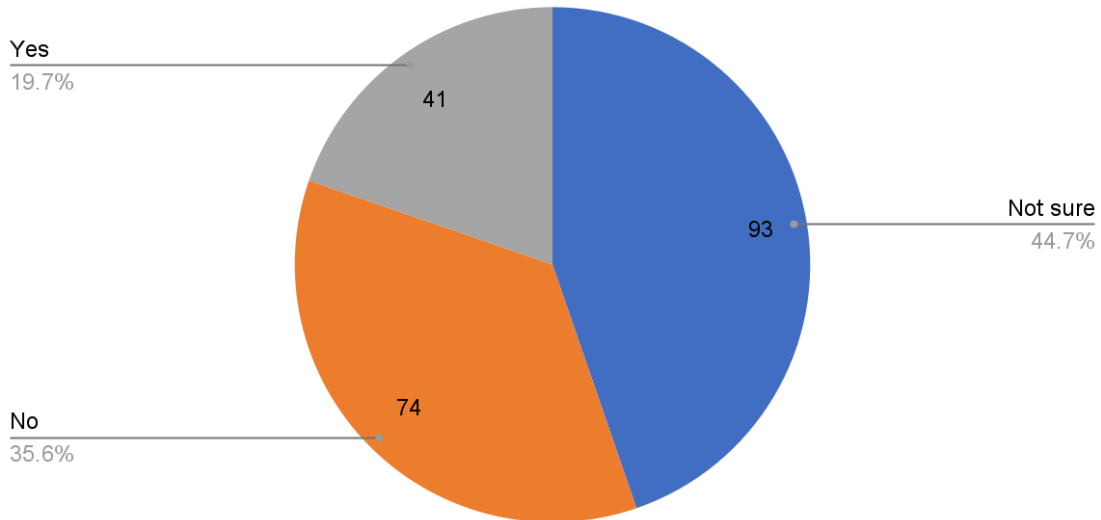


47(a) Has your organization identified any potential risks or drawbacks of using AI in digitization?

Of 217 submissions included in the overall analysis, this question received 208 responses.

When asked whether their organizations had identified any potential risks or drawbacks of using AI in digitization, almost half of respondents answered “Not sure” (44.7%). On the other hand, a considerable percentage of respondents answered “No” (35.6%), while fewer respondents answered “Yes” (19.7%).

Q47. Has your organization identified any potential risks or drawbacks of using AI in digitization?

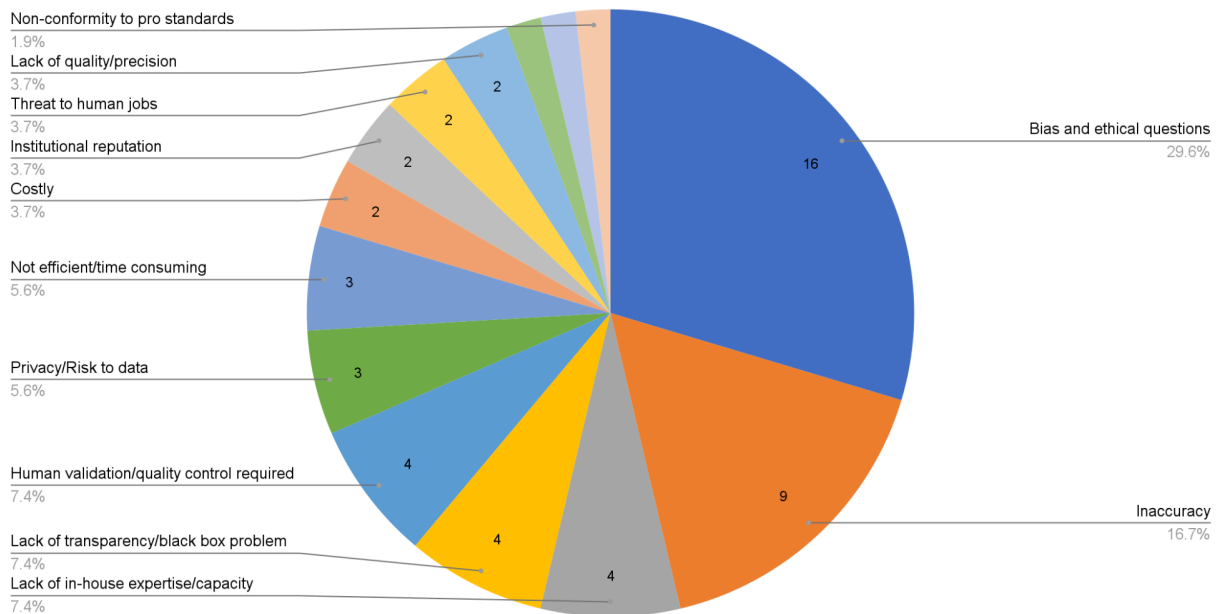


47(b) If yes, please elaborate (on potential risks or drawbacks of using AI in digitization).

This question was conditional upon a “Yes” answer to question no. 47(a), which received a total of 41 “Yes” responses. Of these, 36 participants provided a total of 54 answers to this free-form question.

Of the respondents who indicated that their organizations had identified potential risks or drawbacks of using AI in digitization, a significant margin reported that bias and ethical questions was a concern (29.6%). The next most reported concern was inaccuracy (16.7%), followed by an equal number of responses for lack of in-house expertise/capacity, lack of transparency, and the need for human validation/quality control when using AI (7.4% each).

47(b) Has your organization identified any potential risks or drawbacks of using AI in digitization? If yes, please elaborate.

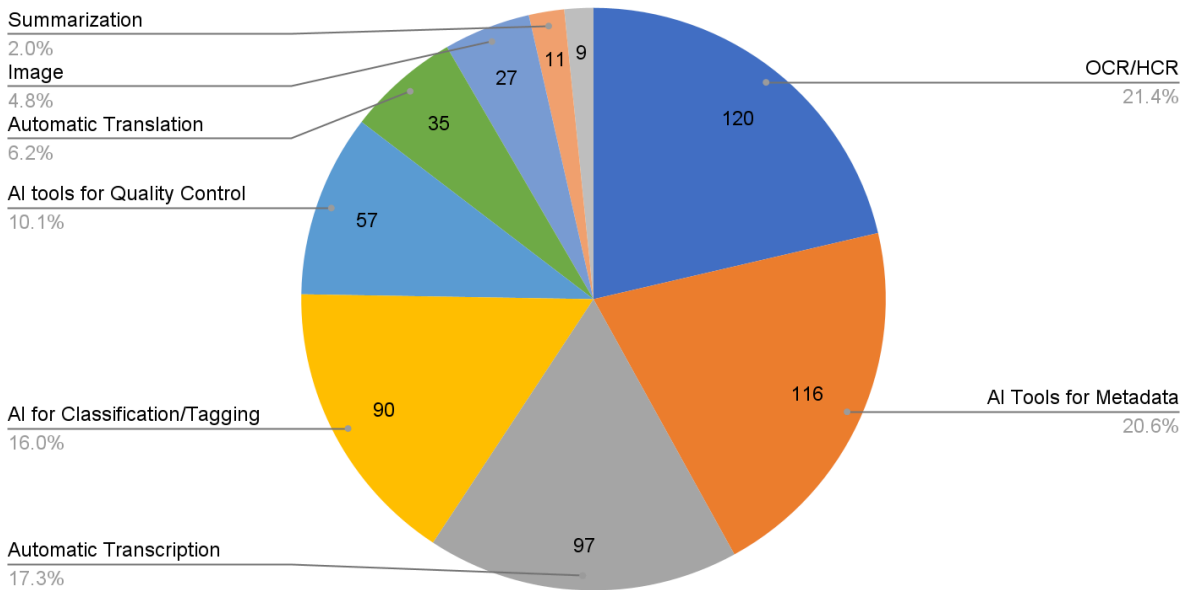


48(a) In your view, what are three types of AI tools that potentially address the greatest need or offer the greatest benefit during digitization? The purpose of this question is to explore how we can make a meaningful impact in our study. Please select a maximum of three responses from the list below.

Of 217 submissions considered in the overall analysis, 203 participants provided 562 responses to this multiple-choice question, for which up to three responses were permitted per participant.

The AI tool that participants thought could potentially address the greatest need or offer the greatest benefit during digitization was “OCR/HTR” (21.4%), followed by “AI tools for metadata management” (20.6%), “Automatic transcription of audio/moving image” (17.3%), and “AI tools for classification or tagging” (16%).

Q48. In your view, what are three types of AI tools that potentially address the greatest need or offer the greatest benefit during digitization?



48(b) Please elaborate on your choices here.

Of 217 submissions included in the overall analysis, 91 respondents provided 129 answers to this freeform question.

Respondents' answers to this question were classified according to the type of comments they made to elaborate on their choices in question 48(a). The top type of comments pertained to how AI could help save staff time and support staff in their current work, or how certain digitization tasks are time-consuming (29.5%). In this context, participants frequently remarked on the time-consuming nature of various aspects of metadata creation and management, transcription, OCR/HTR, quality control, and translation. Some participants complained of the drudgery and tedium of certain tasks, such as data entry, and stated that AI could help with repetitive tasks. The next most frequent type of comments related to how AI could help to enhance the searchability, access, and discoverability of digital collections (14.7%). One person stated that AI could help with improving the accessibility of collections at the item level. Following this, respondents most frequently replied with their concerns or criticism of AI (12.4%), such as that AI still requires human review or intervention, or that there are concerns with the quality and accuracy of AI output. A few respondents stated that AI could be used to describe existing characteristics (such as in transcription or OCR/HTR), but should not be used for the creation of metadata, or for image enhancement, for instance. One person cited the hidden costs of unpaid human labour behind the development of AI, such as the use of human translation for machine learning.

Q48(b) Please elaborate on your choices here.			
Elaboration of choices	Description	Count	%
Save time and resources/Support staff	Metadata (x 14) - e.g., creation, classification, tagging, photos, decolonisation, geotagging; Transcription (x 9) - e.g., of AV, handwritten documents; OCR/HTR (x 4); Quality control (x 4) - including cropping and deskewing; Translation (x 4); Alleviate drudgery/tedium (x 3) - e.g., data entry, transcription; Streamline processes (x 3); Enhance human capabilities for images; Image recognition; Repetitive tasks; Tasks impossible without AI.	38	29.5%
Increase discoverability/access/searchability	Metadata including tagging, classification (x 7); OCR/HTR (x 7); Transcription (x 5); Data extraction/computational data (x 3); Translation (x 2); Improve accessibility at item level.	19	14.7%
Concerns/Criticism of AI	Require human review/humans do it fine or better (x 6) - e.g., taxonomies/ standards-based metadata, classification, tagging, transcription; Accuracy/quality concerns (x 5); Metadata concerns (x 4) - e.g., would not use for description/subject analysis/tagging; OCR (x 4); AI should not be used to change or enhance content (x 3); Bias/discrimination (x 2); No reliable existing tools (x 2); Transcription (x 2); AI does not know ethics of film restoration; AI obscures truth and authenticity - e.g., image restoration/reconstruction; Hidden unpaid human labour behind AI, e.g., translation; Privacy issues.	16	12.4%

Description of content or tasks for AI	OCR/HTR (x 9) - e.g., for handwritten content, blind people, non-Latin scripts, German and Gothic script, Cyrillic; Metadata (x 7) - e.g., structured metadata, AV indexing, alt text for photos; Quality control (x 5) - e.g., cropping and rotation, for large-scale projects; AI for accessibility, visual- or hearing-impaired (x 3); Transcriptions (x 3) - e.g., for AV, hard to understand dialogue; Facial recognition (x 2); Translation (x 2) - e.g., for AV, from/between colonizing languages to African languages.	15	11.6%
Reiteration of answer with no further description	OCR/HTR (x 3); Metadata management/creation (x2); Choices based on needs for archival processing; Image reconstruction/restoration; Quality control; Summarization; Transcription.	14	10.9%
Other/NA	Unsure (x 2); Confidential; No time to develop AI; Not yet using AI; Nothing to add; Would like to know pros and cons of AI; Would like to test.	11	8.5%
AI for tasks not provided in multiple choice options	AI could help with staff disputes about classification, metadata, quality control; Big data extraction and analysis; De-duplication of catalogue records; De-interlacing or scaling of interlaced SD-Video; Digital newspaper processing - e.g., improved OCR, de-skewing & cropping, article segmentation, image identification; Image enhancement.	8	6.2%
Current/Future uses of AI	OCR (x 2); AI for access restriction decisions on born-digital records; Have seen impact the AI tools have; Hope to use HTR; Need AI for metadata but have not received support thus far.	5	3.9%

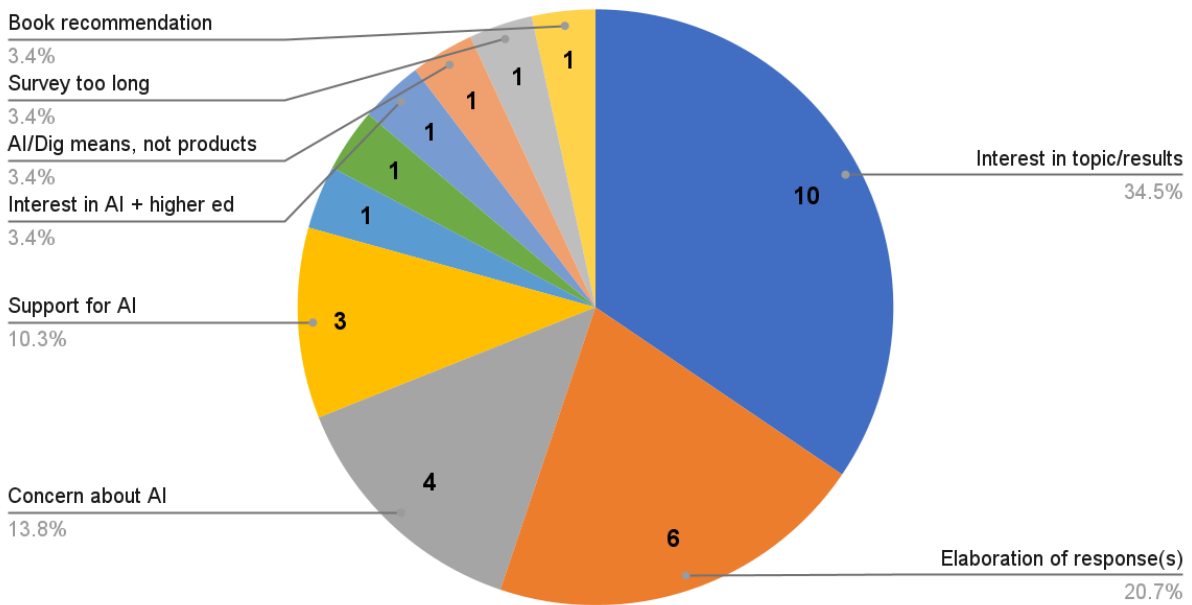
Policy/Regulation	Federal translation requirements; Mandated to transcribe AV; Transcription.	3	2.3%
Total		129	100.0%

49) Please provide any other comments here.

Of 217 submissions included in the overall analysis, this question received 29 responses.

At the end of the survey, participants were given the opportunity to provide any other comments. Most often, participants expressed their interest in the topic of AI and digitization, and/or stated that they would like to be informed of the survey results (34.5%). Other types of responses included elaborations of previous responses (20.7%), expressions of concern about AI (13.8%), and expressions of support for AI (10.3%).

49. Please provide any other comments here.

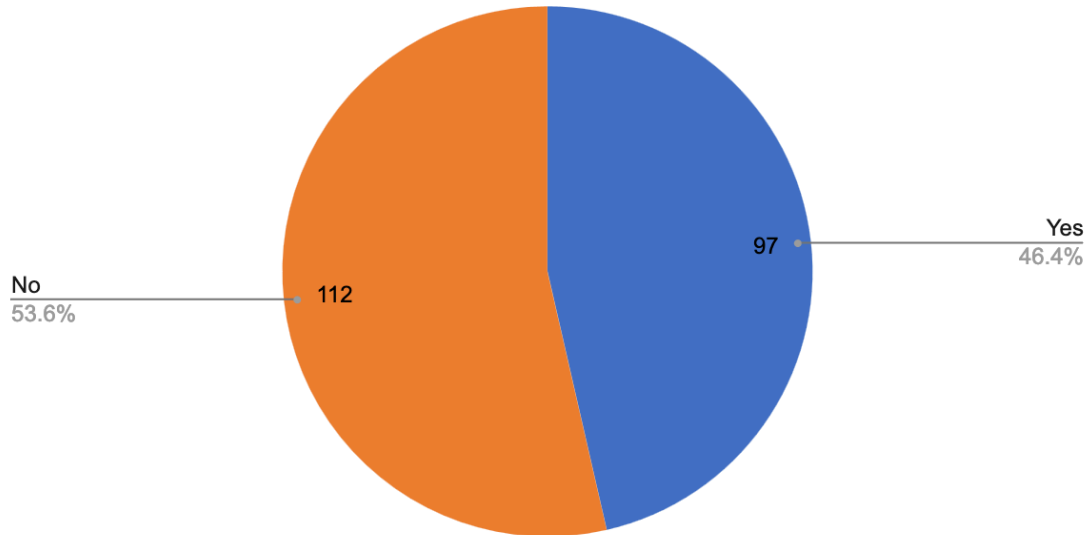


50) Would you agree to being contacted by the study team for a follow-up interview? Note that responding “Yes” does not imply your consent to being interviewed, only your consent to being contacted by the study team. Even if you provide your contact information, your responses will still be treated anonymously.

Of 217 submissions included in the overall analysis, this question received 209 responses.

A total of 97 respondents accepted to be contacted for a follow-up interview.

Q50. Would you agree to being contacted by the study team for a follow-up interview?



[Back to Table of Contents](#)

6. Conclusion

The findings of this survey highlight the importance and value of digitization for enhancing access to and safeguarding archival and documentary heritage collections. While digitization is a resource-intensive activity that is actively practiced within organizations - often at large scale - staff are often constrained by challenges such as limited or unstable resources, limited awareness of legal frameworks, lack of internal digitization policies, and inadequate institutional support. This situation may reflect the fact that in many institutions, digitization is not part of core strategic objectives, and digitization may not necessarily be recognized as part of an organization's digital strategy. Addressing barriers such as funding, staffing, and digital infrastructure will be critical for organizations to sustainably manage and realize the potential of digitization. Furthermore, the risks of digitization - ranging from sustainability concerns to copyright and data privacy issues - underscore the importance of long-term planning in digitization initiatives.

At the time of the survey, organizations were still at an early stage of uptake and risk

analysis in their employment of AI tools when digitizing documentary heritage materials. Yet a number of organizations are considering using AI for digitization, motivated by the potential for AI to alleviate and enhance human labour. Among participants whose organizations had used AI tools for digitization, respondents cited both benefits and limitations of using AI tools. Benefits were related to improved efficiency and access afforded by AI tools, while limitations consisted of concerns related to quality and inaccuracy as well as inefficiencies when implementing AI. Some respondents also expressed concerns related to the risks of using AI for digitization, particularly in terms of accuracy and quality of AI-produced results, and ethical issues related to privacy and access and the development of machine learning.

Digitization is an activity that will continue to be widely practiced and to have a significant long-term impact within institutions and for users of digitized collections. AI tools have the potential to support digitization needs and gaps, particularly when it comes to assisting with existing digitization activities related to digitization workflows, automating labour-intensive tasks, and enhancing access, searchability, and discoverability of collections. AI tools such as OCR/HTR, speech-to-text transcription, metadata creation, classification/tagging, image processing, and translation, among other functions, are and will continue to be increasingly used for digitization purposes.

While the potential benefits of AI for digitization purposes may be significant, it is critical for organizations to systematically assess and understand the risks of using AI for digitization, including ways to mitigate the risks. We acknowledge that many organizations may not have either the resources, the expertise, or the will to take on such risk assessments - individual staff may, for example, be faced with organizational hierarchies with uncritical approaches to AI, or AI tools integrated into software they are already using. For these reasons, it will also become particularly important for standard-setters and organizations at a more mature stage of AI employment to develop and share best practices at the intersection of AI and digitization. The research findings demonstrate a need for robust discourse and published case studies that address the benefits, challenges and risks of AI tool integration in the digitization of documentary heritage materials, with the goal of normalizing critical approaches in the archival profession and beyond towards the use of AI tools as part of digitization planning.

[Back to Table of Contents](#)

Annex A Dissemination List

North America

Name of Group / Organization	Listserv
The Archival Education and Research Initiative (AERI)	AERI@LISTS.UCLA.EDU
ARSC Library and Archives Discussion List	ARSCLIB@LISTSERV.LOC.GOV
Art Libraries Society of North America	LISTSERV@LSV.ARLISNA.ORG
Association of Moving Image Archivists	AMIA-L@LSV.UKY.EDU
Audiovisual Archives Interest Group	AVAIL@SI-LISTSERV.SI.EDU
Canadian Council of Archives	ARCAN-L@MAILMAN.SRV.UALBERTA.CA
Canadian Map & GIS Libraries and Archives	CARTA-L@LISTSERV.UTORONTO.CA
Digital Library Forum (DLF) Announcements	DLF-ANNOUNCE@LISTS.CLIR.ORG
Digital Library Forum (DLF) Levels of Preservation WG	NDSA-LEVELS@LISTS.CLIR.ORG
Digital Library Forum (DLF) Museums Cohort	DLF-MUSEUMS-COHORT@LISTS.CLIR.ORG
Digital Library Forum (DLF) Project Managers	DLF-PM-GROUP@LISTS.CLIR.ORG
Digital Library Forum (DLF) Technology Strategy for Archives	DLF-TS4A@LISTS.CLIR.ORG
Electronic Records Listserv	ERECS-L@LISTSERV.ALBANY.EDU
Human Rights Archives and Archivists	HR_ARCHIVES-L@LISTSERV.UCONN.EDU
Library and Information Science in Canada	BIBCANLIB-L@LISTSERV.LAC-BAC.GC.CA
Library Reference Issues Discussion List	LIBREF-L@LISTSERV.KENT.EDU
National Digital Stewardship Alliance (NDSA)	NDSA-ALL@LISTS.CLIR.ORG
Society of American Archivists (multiple sections)	WWW2.ARCHIVISTS.ORG/LISTSERVS
The Museum System (TMS) Users	TMSUSERS@SI-LISTSERV.SI.EDU
Western Archivists	WEST_ARCH@LISTS.BERKELEY.EDU

Europe

Name of Group / Organization	Listserv
Association for Historical and Fine Art Photography (AHFAP)	AHFAP@JISCMail.AC.UK
Archivists, Conservators, and Records Managers	ARCHIVES-NRA@JISCMail.AC.UK
Cultural Heritage Datasets	CULTURAL-HERITAGE-DATASETS@JISCMail.AC.UK
Forum de discussion de l'Association des archivistes français	ARCHIVES-FR@LISTES.ARCHIVISTES.ORG
Health Archives Records Group	HEALTH-ARCHIVES-RECORDS-GROUP@JISCMail.AC.UK
DigiPres	DIGITAL-PRESERVATION@JISCMail.AC.UK
Library and Information Professionals	LIS-PROFESSION@JISCMail.AC.UK
Library - Technology List	LIBRARY-TECHNOLOGY@JISCMail.AC.UK
LIS-LINK (General LIS News and Discussion)	LIS-LINK@JISCMail.AC.UK
LIS-MEDICAL (UK Medical/Health Care Library Community)	LIS-MEDICAL@JISCMail.AC.UK
LIS-MIDDLE-EAST (Middle Eastern & Islamic Collections)	LIS-MIDDLE-EAST@JISCMail.AC.UK
Libraries in Croatia	Members-only list.
Member List - Croatian Archivists' Society	No formal member list.
Mass Digitization of Medieval Manuscripts	MASS-DIGITIZATION@JISCMail.AC.UK
Museum Librarians and Archivists Group	MLAG@JISCMail.AC.UK
Museums Association (London Members)	MA-ENGLANDLONDON@JISCMail.AC.UK
Museums Association (Midlands Members)	MA-ENGLANDMIDLANDS@JISCMail.AC.UK
Museums Association (Northern England Members)	MA-ENGLANDNORTH@JISCMail.AC.UK
Museums Association (Scotland Members)	MA-SCOTLAND@JISCMail.AC.UK
Museums, Galleries and Heritage Research	MUSEUMSGALLERIESHERITAGERESEARCH@JISCMail.AC.UK
The Galleries, Libraries, Archives and Museums' GLAM Labs	GLAMLABS@JISCMail.AC.UK

Southeast Asia & Pacific

Name of Group / Organization	Listserv
Archives and Records Association of New Zealand	NZRECORDS@LISTS.VUW.AC.NZ
Australia and New Zealand Library and Archives Group	LIS-ANZLAG@JISMAIL.AC.UK
SouthEast Asia-Pacific AudioVisual Archives Association (SEAPAVAA)	SEAPAVAA@GROUPS.IO

Middle East

Name of Group / Organization	Listserv
Librarians & Library Staff in Qatar	QLIB@LISTSERV.TAMU.EDU

International

Name of Group / Organization	Listserv
CODE4LIB	CODE4LIB@LISTS.CLIR.ORG
International Association of Sound and Audiovisual Archives (IASA)	LIST@IASA-WEB.ORG
International Council of Museums (ICOM) Discussion Group	ICOM-L@HOME.EASE.LSOFT.COM
International Council on Archives (ICA)	ICA-L-REQUEST@MAILMAN.SRV.UALBERTA.CA
ICA Section of International Organisations (SIO-ICA)	SIO@ICA.ORG
International Federation of Library Associations and Institutions (IFLA)	IFLA-L@IFLALISTS.ORG
InterPARES Trust AI Researchers	ITRUST-ALL-RESEARCHERS@LISTS.UBC.CA
The Preservation and Archiving Special Interest Group (PASIG)	PASIG-DISCUSS@LISTS.STANFORD.EDU

[Back to Table of Contents](#)