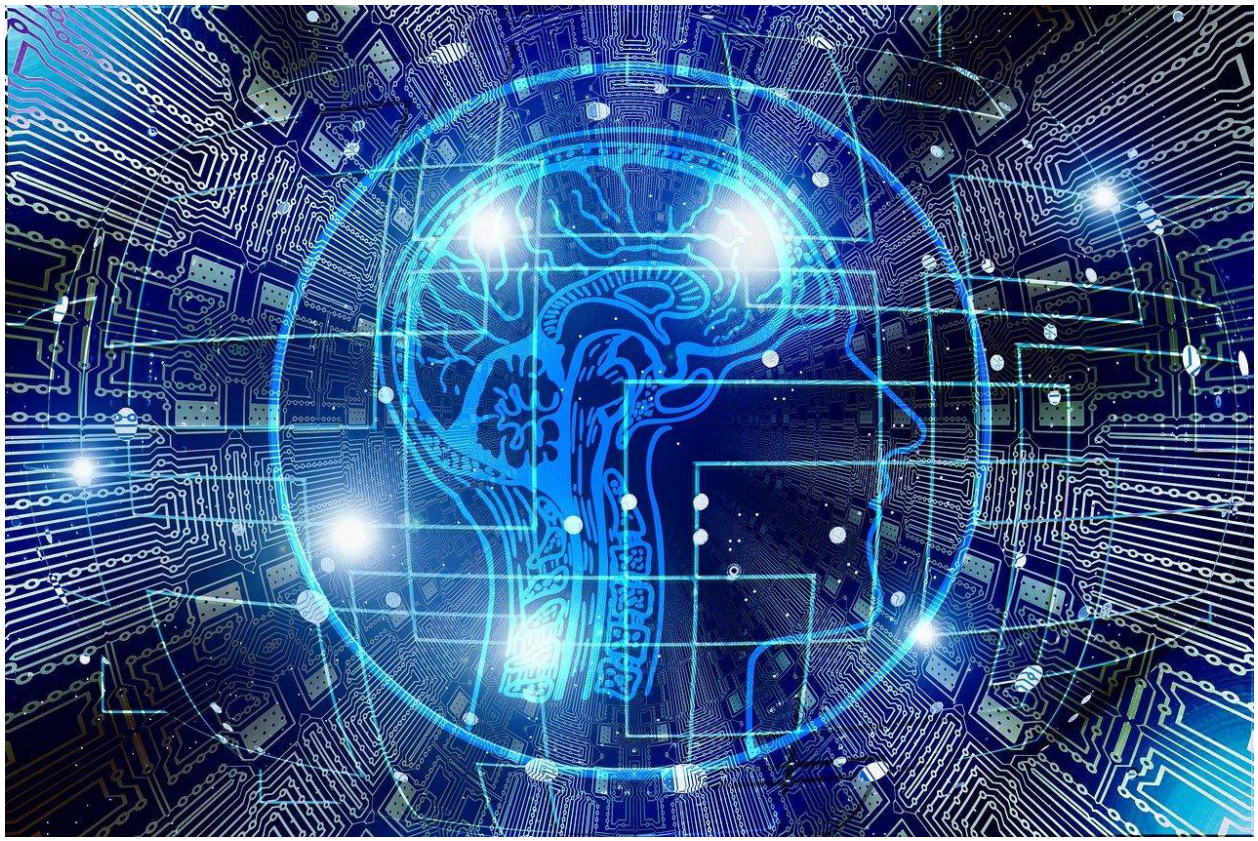
Employing AI for Retention & Disposition in Digital Information and Recordkeeping Systems



An ITrust^{AI} User Survey Report

June 30, 2023

Table of Contents

Introduction

AI tools and techniques are exerting a pervasive influence on society and the workplace. Individuals engage with personal fitness devices and virtual assistants daily. Simple solutions such as chatbots and recommendation systems are being employed by businesses to engage with the public. More advanced AI solutions enable self-driving automobiles and detect insurance fraud. Consequently, information professionals must understand the benefits of employing artificial intelligence to assist them with their recordkeeping tasks. They must also understand how the use of AI by others brings about very different kinds of records that must also be retained. To accomplish these goals, a team of researchers engaged in a study entitled *Employing AI for Retention & Disposition in Digital Information and Recordkeeping Systems*.

About the Study

This document reports on the survey phase of the study. Additional data were collected through literature reviews, an inventory of AI-enabled products and services available to manage digital content, interviews with practitioners and vendors, and product demonstrations. This survey gathered insights from the users of software and systems that manage digital content to understand if and how AI can be employed in trusted digital recordkeeping repositories or other electronic storage solutions (e.g., records management systems). This survey was conducted as part of the InterPARES Trust AI research agenda (https://interparestrustai.org), which is described at the end of this report.

Survey Demographics (Target Population)

The primary target population for this survey was information professionals working in archives, records management, and information governance positions. The secondary target population was professionals working in related fields such as security, privacy, law, and finance. The results indicate that the survey reached both its primary and secondary intended populations.

Process (Methodology)

The survey instrument was created using Qualtrics survey software. It was open approximately 5 weeks, from March 7, 2023, through April 15, 2023. To reach our intended population, requests were made to representatives of professional associations (as listed in Appendix 2) in the fields of records management, archives, information governance, etc. to share the link to

the survey with their members and to encourage their participation. Additional invitations to participate by individuals seeking their participation, as well as by posting the invitation to relevant professional association listservs, forums, and community chat areas.

Survey Design, Results, and Data Analysis

The survey consists of 4 sections and a total of 36 questions: Demographics (8 questions); Data, Information, and Records Systems (11 questions); Retention and Disposition (8 questions); and Artificial Intelligence (9 questions). The sections were designed to help us understand who we were reaching; the kinds of systems in use in their workplace; the extent to which they are accomplishing retention and disposition goals; and if and how they are employing (or intend to employ) AI-enabled technologies to help them manage their digital content. The respondents were provided with the option to include their name and email address if they were willing to be contacted regarding their responses; 63 respondents provided contact information.

Since none of the questions required answers, not all participants answered all questions. Although there were more than 400 responses (only a small number of which appear to be bots), not all respondents completed the entire survey. The team decided to analyze only those 214 responses that were considered 100% complete by the system; however, not all 214 responded to each question since the presentation of some questions was based on a certain response to the previous question.

The data was analyzed by four of the researchers, this report was created, and the remaining members of the team were involved in reviewing and editing the report before it was published.

Survey Results

Part 1: Survey Demographics

Of 324 valid responses to the survey, a total of 214 were considered 100% complete. To get the best picture of the typical respondent, it was decided to use the responses of only those 214 individuals. Because answers to the questions were not mandatory, the number of responses to each question varies slightly. A total of 8 demographic questions were posted to provide context to the answers provided by the respondents.

1. In what industry do you work?

The primary target population comprised individuals working in archives, records and information management, and information governance. The secondary target population is individuals working in related domains including privacy, security, legal services, and information technology. The responses to this question shown in figure 1 revealed that the survey has reached its intended target populations.

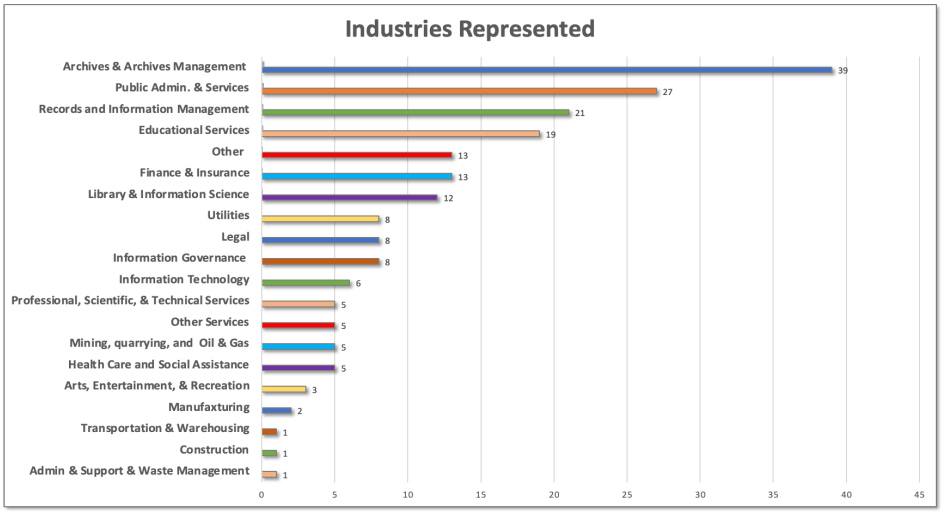


Figure 1. Industries represented by respondents to question 1.

2. How many people are employed in your organization?

There was almost equal representation from organizations of less than 100 employees, between 1,000 and 5,000 employees, and more than 5,000 employees, as illustrated in figure 2.

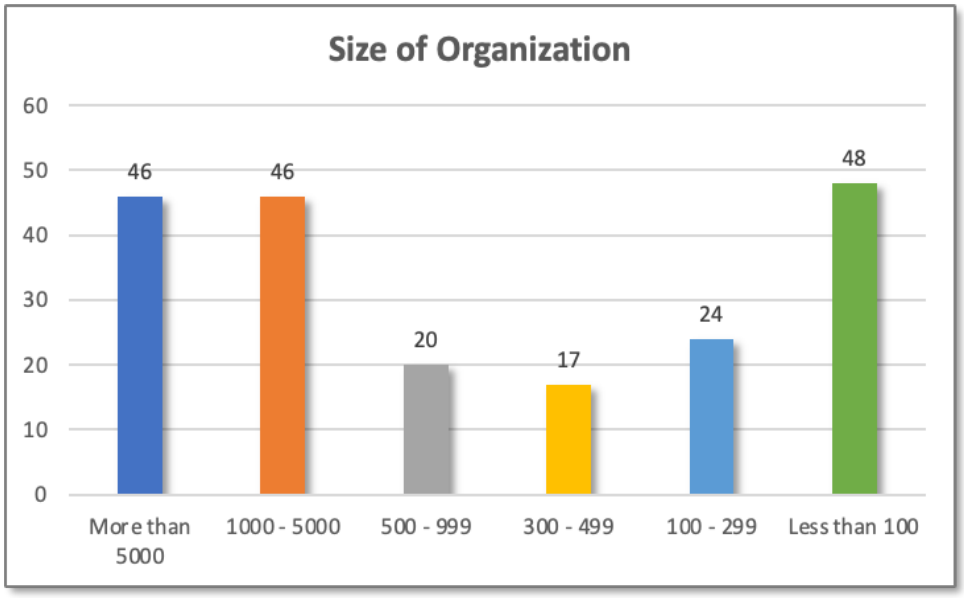


Figure 2. Number of employees in the organization.

3. In which country is your organization's headquarters located?

Because we target professional associations primarily based in the U.S.A. and Canada, it is not surprising that most respondents came from those two countries, as shown in figure 3. However, respondents from Mexico and the United Kingdom were also in the double digits.

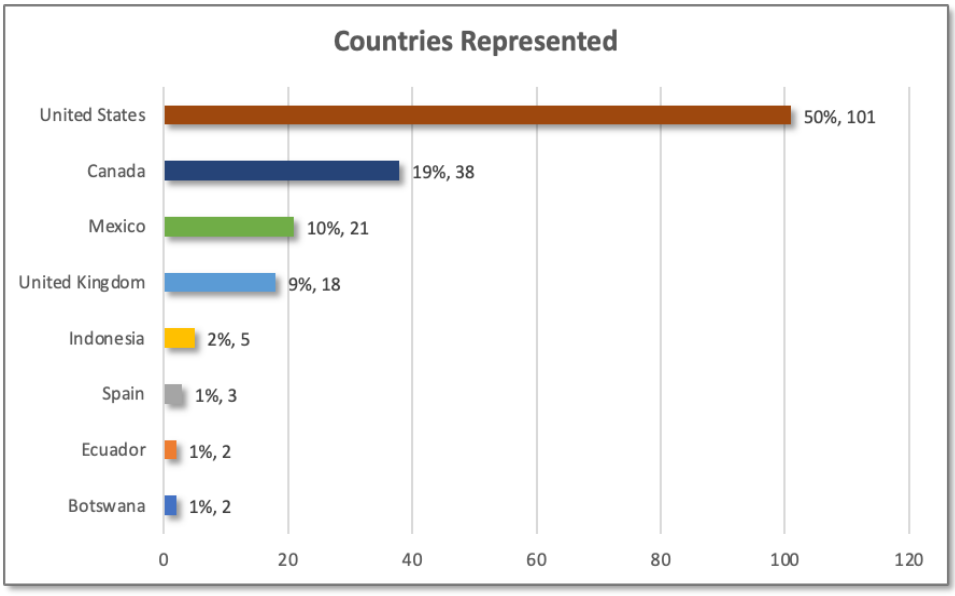


Figure 3. Countries represented by at least 2 respondents to question 3.

In addition to those countries shown in figure 3, one respondent came from each of the following countries: Argentina, Australia, Brazil, Bulgaria, France, Ghana, Italy, Netherlands, Saudi Arabia, Scotland (UK), Slovenia.

4. How would best describe your current specialization? (Select all that apply.)

In this case, specialization reflects the domains for which the respondents claim expertise and interest. It does not describe their current position at work. As shown in figure 4, records management was the most frequent choice, followed by archives. Two dual specializations—archives/records management and information governance/records management—were indicated by the same number of respondents. The next combination indicated by respondents was a mix of archives, information governance, and records management.

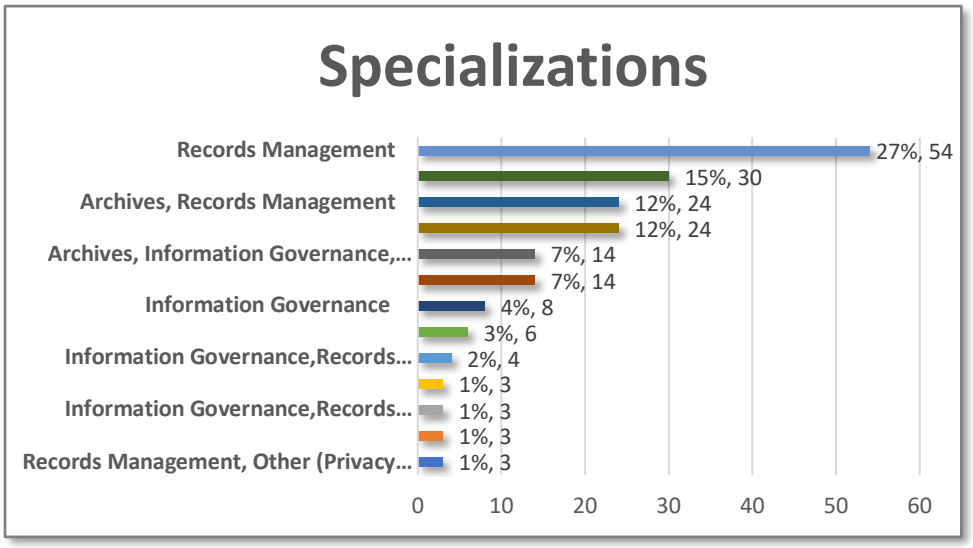


Figure 4. Areas in which respondents indicated they have specialized knowledge.

Of the 203 responses to this question, 95 (47%) indicated a single area of specialization from the options provided: 27% selected records management, 15% archives, 4% information governance, and 1.5% information security. A combination of specializations was selected by 94 (46%), such as archives and information governance or records management, privacy management, FOI, and quality assurance. Fourteen (7%) respondents indicated "other" and provided additional details, as shown in table 1.

Table 1. Responses by 14 respondents selecting "other."

Access and Privacy	Digital Transformation
Administration	Information
Archives and Records Management	Information Technology

Collections Management	Labores administrativas (Administrative Work)
Compliance	Library
Corporate Lawyer	Museum with Archives
Digital Ethics and Knowledge Management	Public Records Disclosure and Records Information Management/Information Governance

5. What is your role in the organization?

Regardless of the specialization indicated by respondents, they occupy specific positions and play certain roles within their organizations. Again, the most often selected choice was Records Manager, followed by Archivist, and Information Governance Professional, as shown in figure 5.

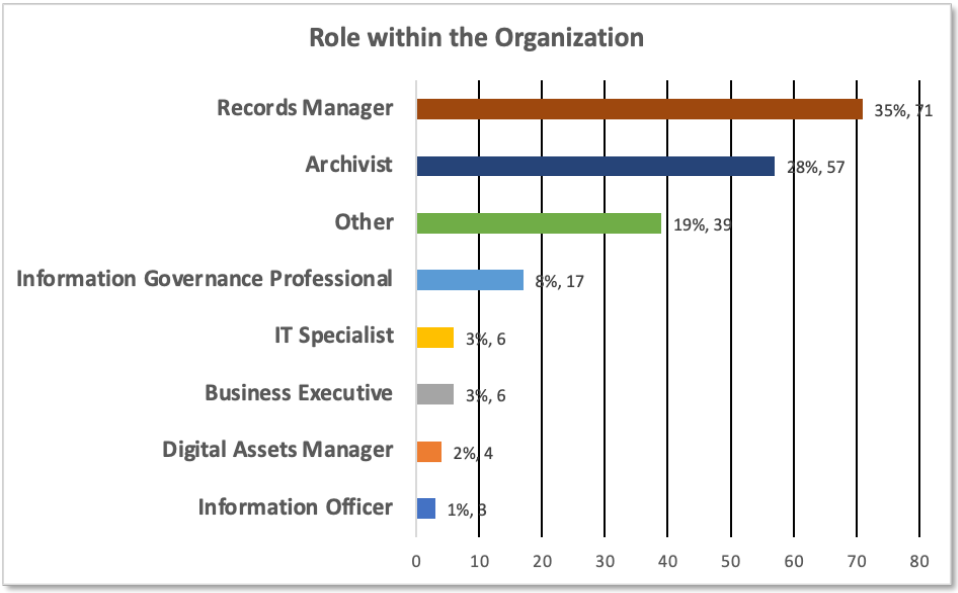


Figure 5. Positions respondents hold within organizations.

6. How many years have you been in your present position?

Of the 158 respondents that answered this question, 71 were Records Managers; 57 Archivists; 17 Information Governance Professionals; 4 Digital Asset Managers; and 3 Information Officers. As shown in figure 6, the category selected most often by Archivists, Information Governance Professionals, Digital Asset Managers, and Information Officers was 1-3 years; while the category selected most often by Records Managers and Information Technology Specialists was 4-10 years. The only positions held by individuals working more than 25 years were Records Managers, Archivists, and Information Governance Professionals.

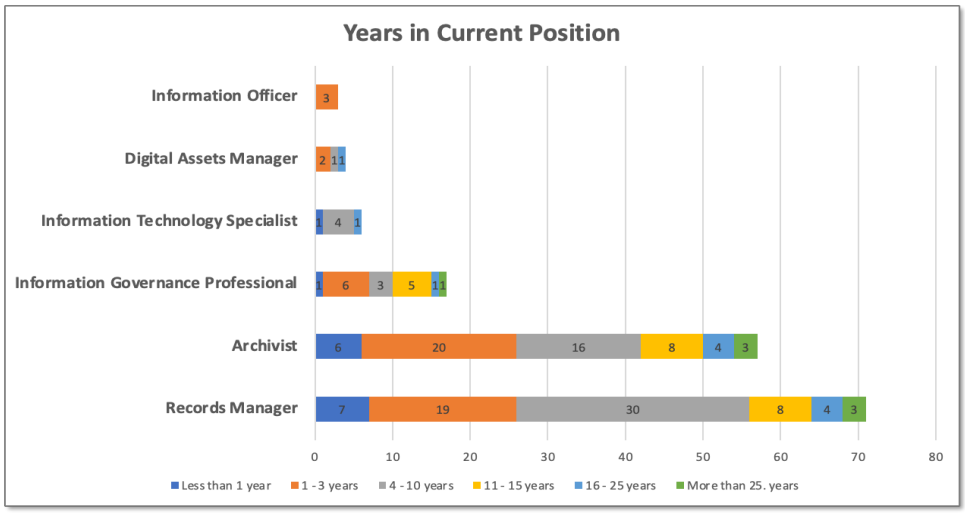


Figure 6. The number of years in which respondents held their current positions. The data labels show the number of individuals who selected those options.

7. Within what department does your position reside?

Reporting departments are shown in figure 7 for the top three roles only: Records Manager, Archivist, and Information Governance Professional (IG). The numbers reflect the number of respondents and not percentages. Archivists were most likely to work in a department that reflected their roles—20 of the 57 work in Archives. Records Managers were most often dispersed across departments with diverse titles such as Knowledge Management, Public Works, People and Culture, and Accounting and Finance. An equal number of Records Managers (12) worked in Information Technology and Administration. Information Governance Professionals were most often in Legal.

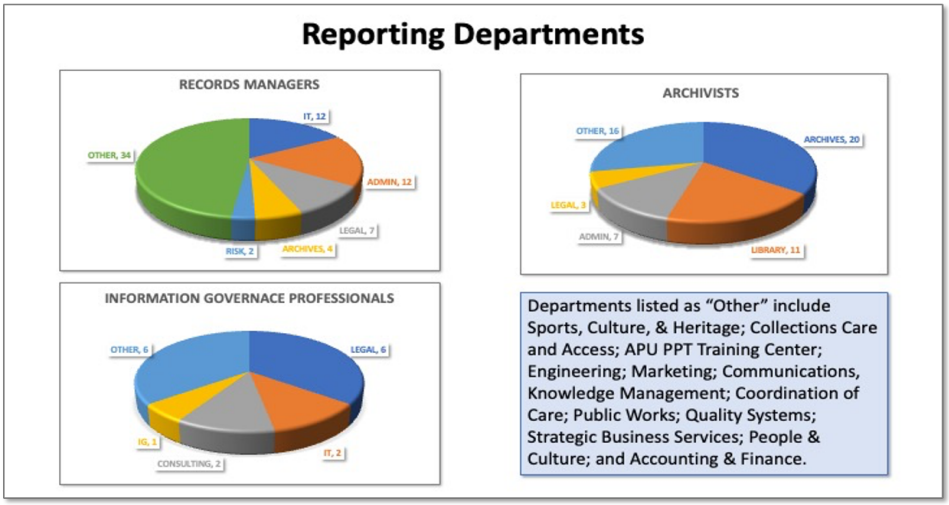


Figure 7. Departments to which the respondents reported.

8. How many employees are in this department?

As shown in figure 8, of the 191 responses to this question, 163 (85%) work in departments of 1 to 30 employees. Only 28 (15%) work in departments of more than 50 employees.

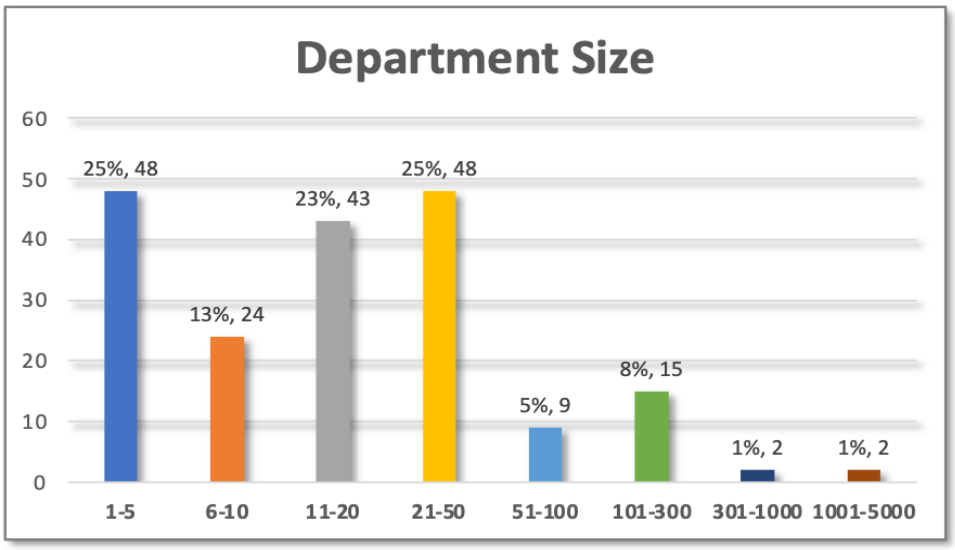


Figure 8. Size of departments within which respondents work by number of employees.

The size of the departments reflects total employees, not those tasked with archives, records management, or information governance duties. A few comments made by survey respondents related to size are:

- One person in the department of 30 is responsible for records management.
- 25 employees work in the library, but only 1 is a dedicated archives employee.
- 1 full-time records manager and 3 part-time students work in the library.

Part 2: Data, Information, Records Systems

9. To what extent do you function as your own recordkeeper, performing manual functions or using software that expects you to decide how to manage records you create?
Participants were asked to select "always," "sometimes," "rarely," or "never." 202 participants responded, with 138 selecting "always" (68%), 43 selecting "sometimes" (21%), 13 selecting "rarely" (6%), and 8 selecting "never" (4%) (see figure 9).

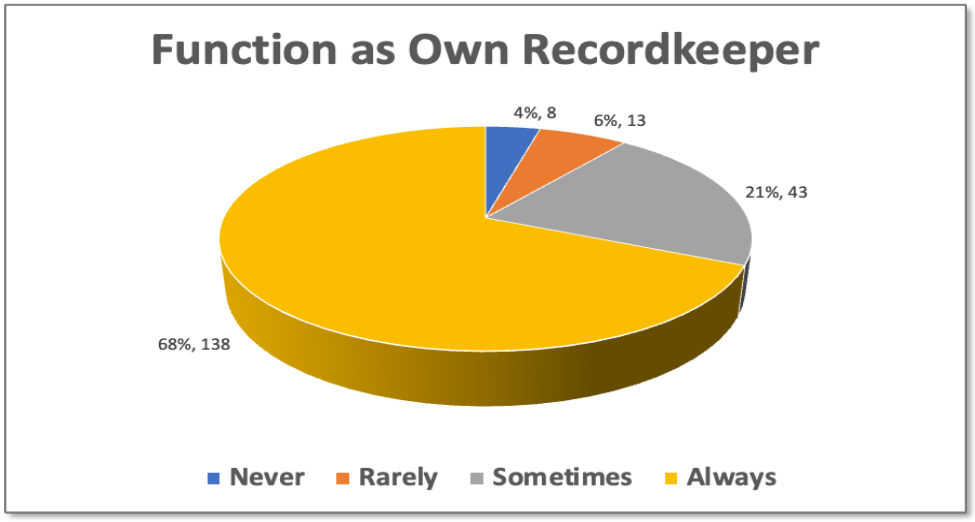


Figure 9. The frequency with which respondents serve as their own recordkeepers.

10. Does your institution employ forms of automation (e.g., workflows, labels, content types) that lessen the burden of performing the recordkeeping function?
Participants were asked to select "yes," "no," or "I don't know." 202 participants responded, with 94 selecting "yes" (47%), 94 selecting "no" (47%), and 14 selecting "I don't know" (7%) (see figure 10).

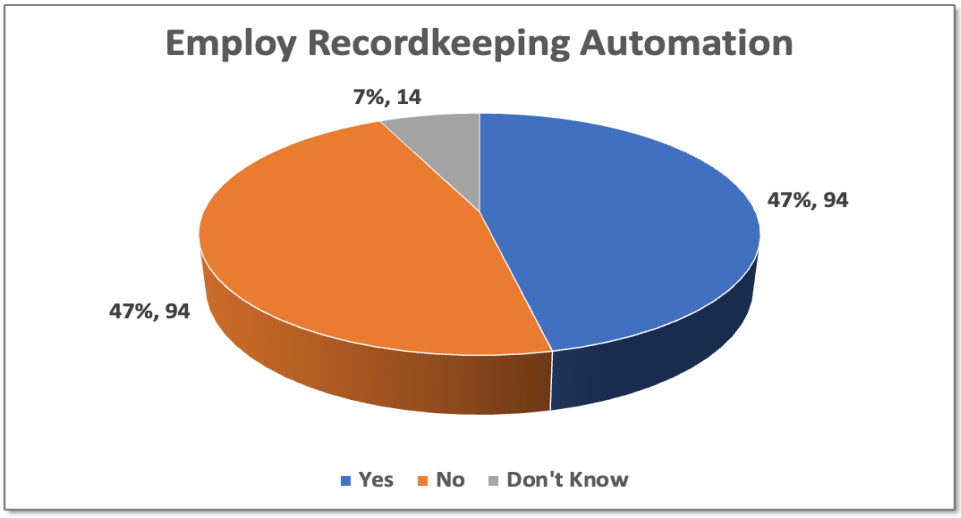


Figure 10. The number of respondents indicating their workplace employs automation to lessen the burden of recordkeeping.

11. If yes, what aspects of recordkeeping are automated (e.g., classification, disposition, tagging, content types)?

This question asked for responses in free text, which were given by 82 respondents. The most frequently cited aspects of recordkeeping were classification (30 responses, 37%), disposition (20 responses, 24%), tagging (20 responses, 24%), content types (11 responses, 13%), and retention (9 responses, 11%) (see figure 11).

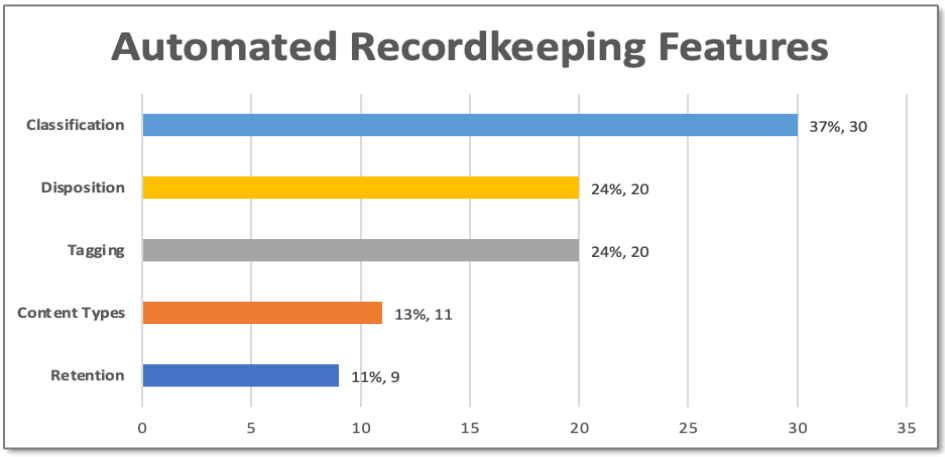


Figure 11. Automated record keeping features currently in use.

12. What applications/technologies are used in your organization to manage digital content (data/information/records)?

We received 194 responses. Respondents were asked to select all applicable applications/technologies from a list provided. The responses are shown in figure 12. An "other" option was available.

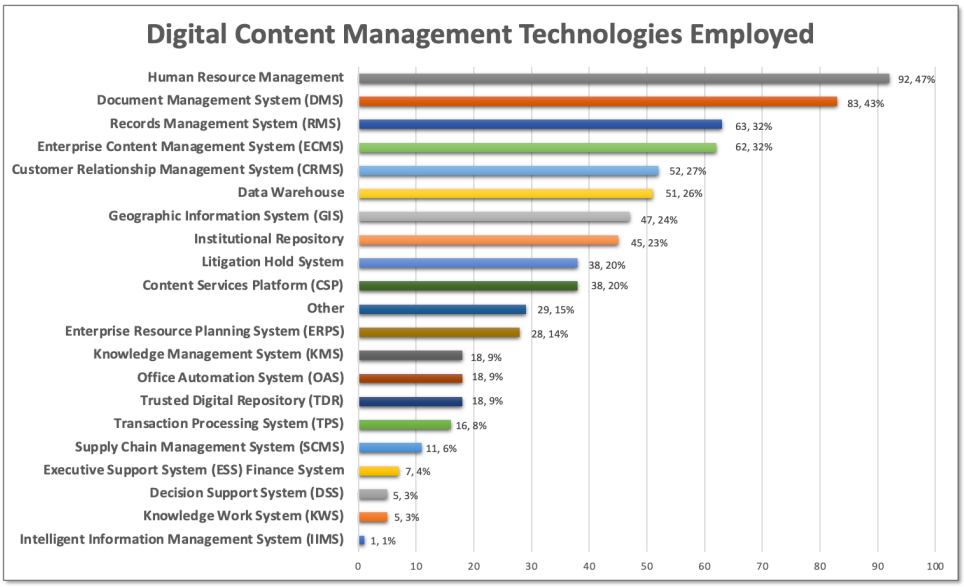


Figure 12. Types of digital content management technologies employed within respondents' organizations.

Of respondents who selected "other," 4 indicated in free text that their institution did not use digital content management technologies; 3 indicated that they were not sure; 2 cited a Collections Management System; 2 cited SAP; and the technologies in table 2 were each mentioned once.

Table 2. Technologies used to manage digital content cited by one respondent each.

Archival Description Software	Digital Archive	Manual registration
CatDV	Digital Asset Management System	Microsoft 365
Case Management System	Digital Preservation System	Microsoft Excel
Cloud Repository	Epic	Netwrix
CollectiveAccess	Firma Electrónica Avanzada (Advanced Electronic Signature (FIEL))	Proprietary database
Content Management System (CMS)	Hyland	SharePoint
Correspondence Management System	Learning Management System (LMS)	

13. Does your organization have a formal written retention and disposition schedule?
Participants were asked to select "yes, on the institutional level," "yes, on the department level," "yes, on both the department and institutional levels," "no," "don't know," or "other." We received 119 responses, as shown in figure 13: Yes, institutional: 96 (81%), 13 selected "other" (11%), and 10 selected departmental level (8%).

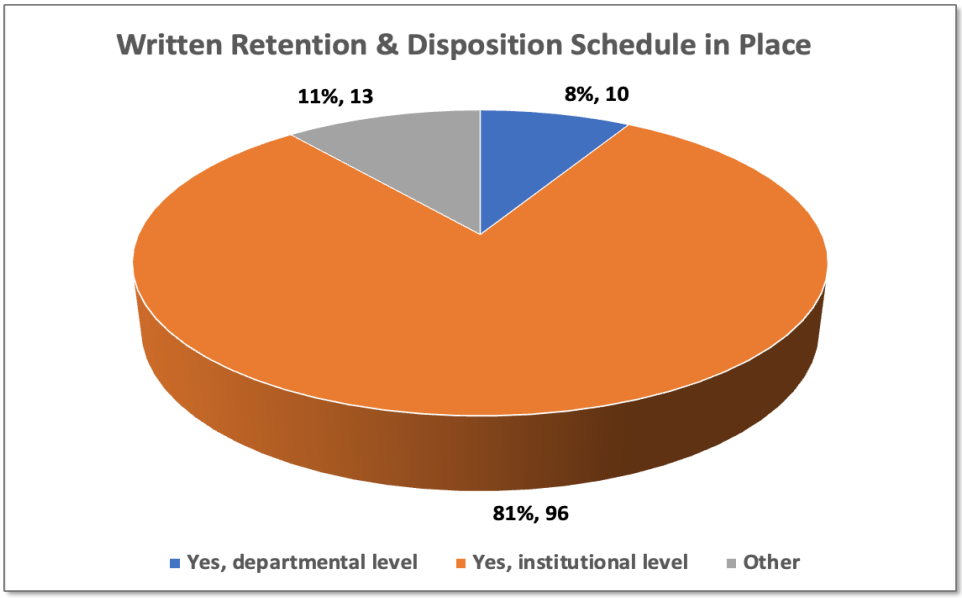


Figure 13. Organizations with existing retention and disposition schedules.

Of the respondents who selected "other," 4 indicated that a schedule was in progress or being drafted; 1 indicated that there was no schedule in place but that their long-term plan is to create one; and 7 answered yes, with additional details:

- Yes for US, No for other countries
- I manage over 300 of them
- We have some schedules but also many "indefinite" classes within the file plan, pending a project to build out the enterprise classification scheme
- Branch level
- Yes, on the state level
- As consultants we draft these for clients
- We follow the SOS schedule

14. Does your organization use electronic records management/content management services or software to manage retention and disposition of records?
Participants were asked to select "yes, paper records," "yes, electronic records," "yes, both paper and electronic records," or "don't know." We received 198 responses, as shown in figure 14: 78 selected "yes, both paper and electronic records" (39%), 53 selected "don't know" (27%), 39 selected "yes, electronic records" (20%), and 28 selected "yes, paper records" (14%).

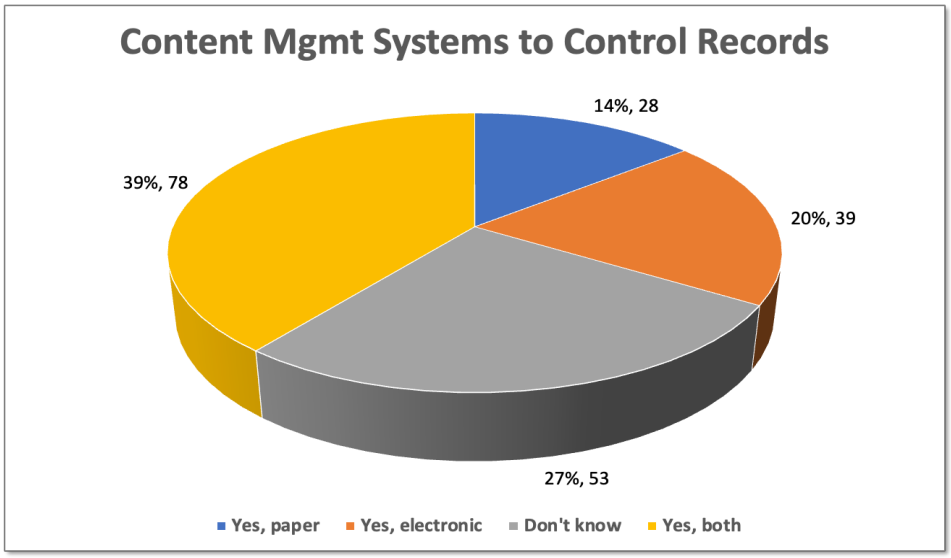


Figure 14. Organizations using content management systems to control paper and/or electronic records.

15. Is automated disposition built into the system?

As shown in figure 15, of 28 responses, 24 selected “no” (86%), and only 4 selected “yes” (14%).

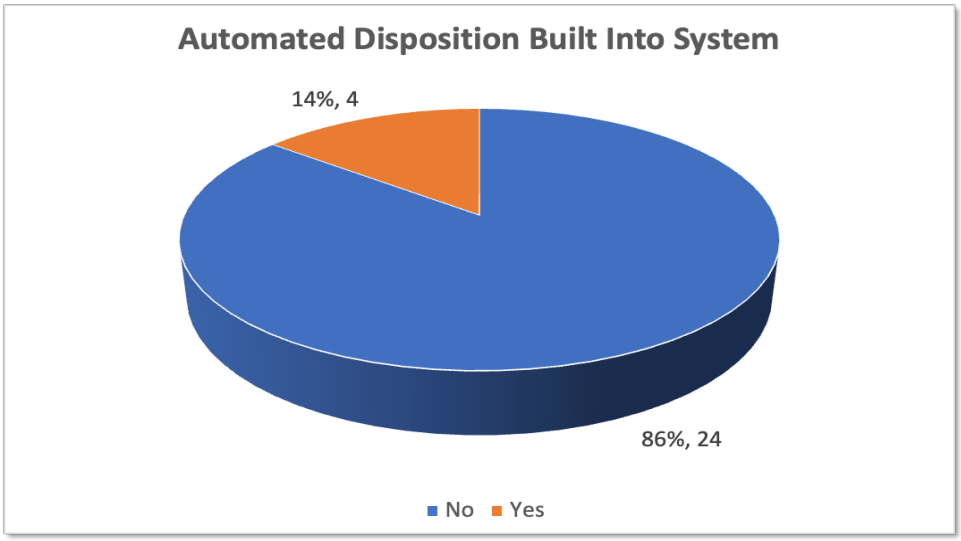


Figure 15. Percentage of automated disposition functionality built into electronic systems.

16. If yes, is there a flag for the records manager (or other party) to review the item or records system for destruction?

Of the 4 responses that selected yes, 100% indicated the presence of a flag for review before destruction.

17. Does your organization maintain an archive (repository for long-term digital records) as part of its content management services?

Of 187 responses, 111 selected “yes” (59%), and 76 selected “no” (41%), as shown in figure 16.

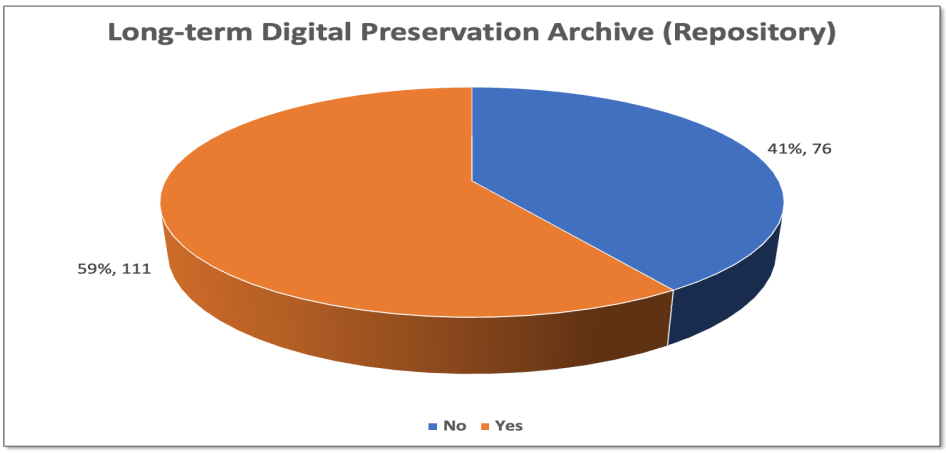


Figure 16. Archive (repository) employed for long term digital records as part of content management services.

18. If yes, what products/services do you employ for long-term digital preservation?

A total of 82 participants responded to this question; some selected more than one product/service from a list provided. The responses are shown in figure 17.

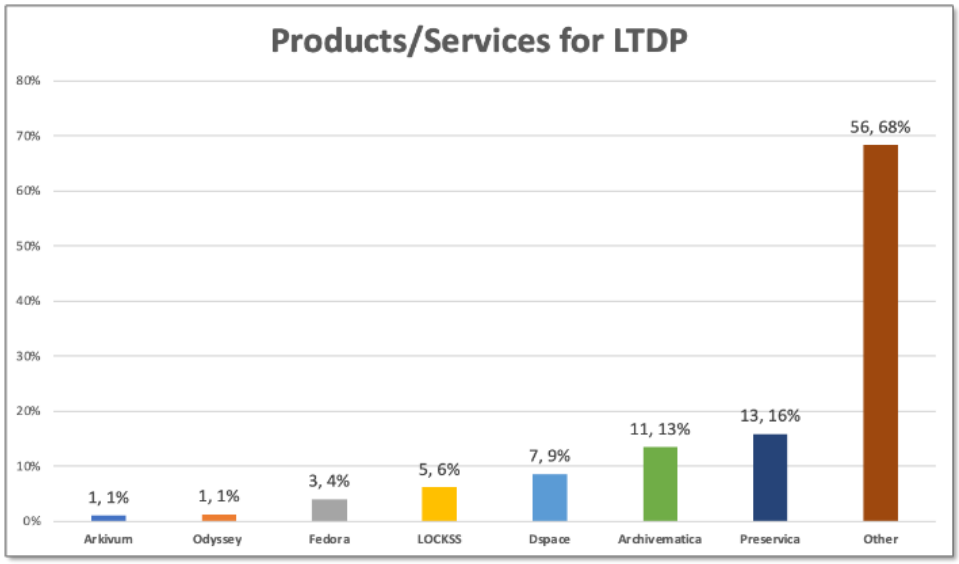


Figure 17. Products/services employed for long-term digital preservation.

Of the “other” responses, 7 cite internal/homegrown systems, 3 cite servers, and 3 cite cloud storage. SharePoint, FileNet, LaserFiche, and Omeka are each cited by 2 respondents. The software/services shown in Table 3 were selected by one respondent each.

Table 3. Products/ services indicated by a single respondent.

Amazon Web Services	Fiserv Nautilus	MetaArchive
Archivematica	Google Drive	Micro Focus Content Manager
ArchivEra	Hyland Onbase	Microsoft Excel
ArchivesSpace	iARXIU	MINISIS TiTAN
Cloud Storage	InfoQuest	National Archives
Collabspace	Internet Archive	OnBase
Constellio	Islandora	Online Archive of California
CONTENTdm	Libsafe	PastPerfect
DropBox	LIBSAFE Go	SaaS solutions
ENKI	LTO	SADAI
FIRIS	Mandarin M5	Synergy by Jack Henry & Associates

19. What information and documentation does your organization keep about its digital preservation process (e.g., technical metadata, laws & standards, documentation of decision points and actions, policies, process mapping)?

We received 159 free text responses. 45 of these responses referred to higher-level policies; 26 indicated that there was no formal documentation of the digital preservation process; 24 indicated that a documentation scheme was in development; 23 indicated that technical metadata was kept; 22 cited procedures/workflows; 16 cited some form of standards/best practices being documented; 16 indicated that there was some process/decision documentation; 14 indicated that applicable laws/regulations were documented; 12 respondents indicated that they did not know what kind of documentation was kept; 12 indicated that documentation included retention schedules; 9 cited descriptive metadata; 8 indicated guidelines/guidance documents; 3 indicated that there was very little, or ad hoc documentation; 3 cited SOPs; 2 cited historical documentation; 2 cited data maps; and 2 cited checksums (see figure 18).

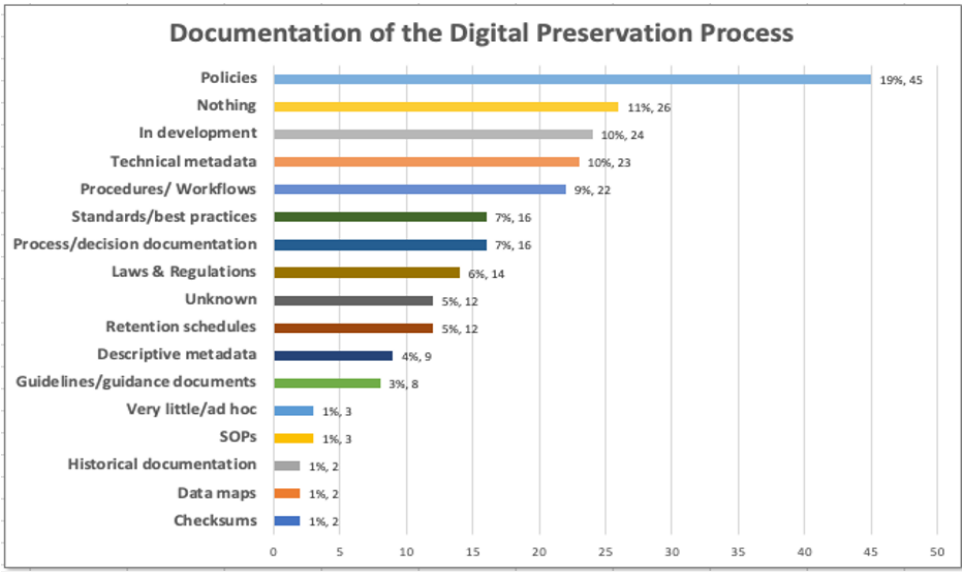


Figure 18. Types of documentation about the digital preservation process retained by the organization.

Part 3: Retention and Disposition

20. What file formats do you create, receive, manage, and/or exchange?

Individuals were asked to select as many file formats as applied. The file formats that are most handled by survey respondents are PDF, DOC/DOCX, and JPEG, as shown in figure 19.

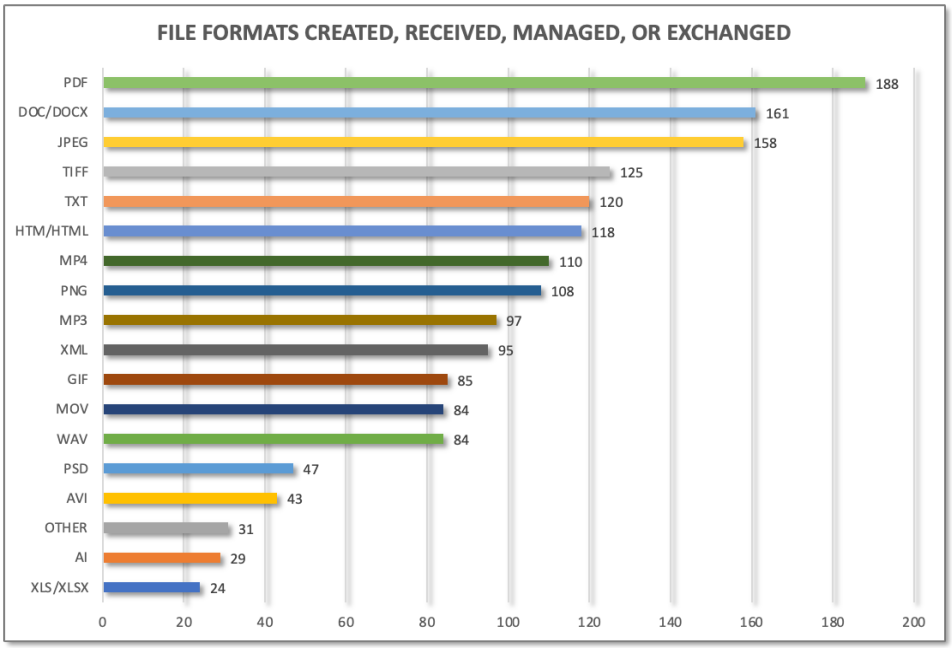


Figure 19. Types of file formats in use within organizations.

Among the "other" file formats (selected by only 1 or 2 respondents each) are PST, CSV, ZIP, PPT, GBD, CDR, PSP, JPG2, CAD, MSG, EML, DB2, Office Formats, and Markdown.

A few comments related to the volume of file formats in use are:

- Private.
- Kept in native format.
- Older file types.
- There are too many different types to mention; some proprietary.
- I'm sure a whole host of other record types once you consider GIS data, data from our Police Dept, etc. not to mention old file formats that are lurking in our network drives and floppy disks from decades ago.
- We have identified 145 digital formats in our holdings so far.
- 850+ file extensions were found during a pilot network drive scan project.

21. What types of content do you create/receive, manage, and/or exchange?

Respondents were asked to select all that applied and were provided with an option to list other content types. As shown in figure 20, documents were selected by all 191 responses to this question. The next natural grouping included metadata (129, 68%), still images (121, 63%), and videos (117, 61%).

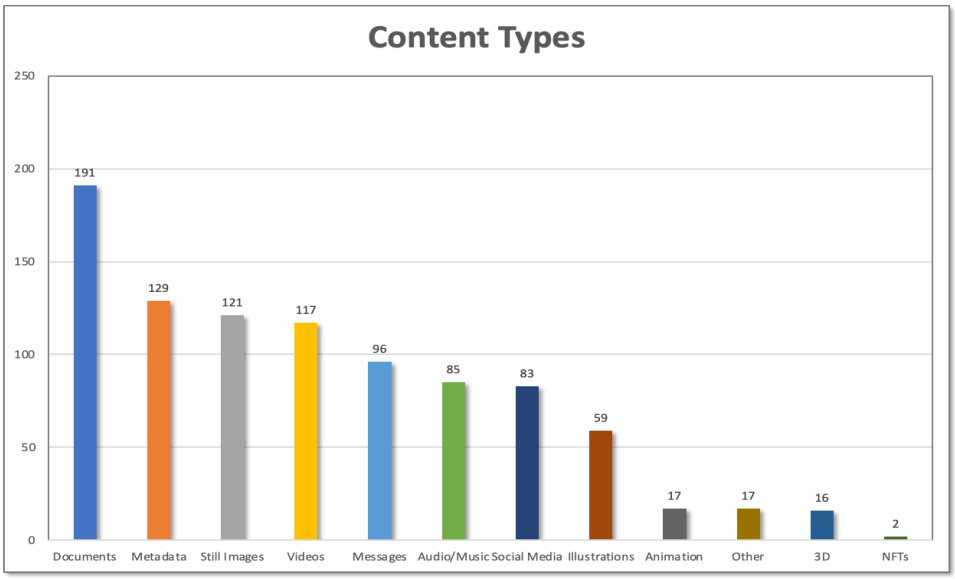


Figure 20. Content types in use within organizations.

Among the content types suggested in response to "other" are web archives, cartographic and architectural plans, engineering drawings, GIS, maps, databases, reports, executable files, private, graphics, data sets, excel, drawings and schematics, and Microsoft Teams files.

Three respondents shared the following comments:

- Any content that is part of research data. Usually tabular, but can be any format.
- All types of structured, semi-structured, and unstructured data.
- Fax, disc, USB storage, print. No email.

A recent content type for which information professionals may find themselves responsible is Non-fungible Tokens (NFTs), selected by two respondents. NFTs are blockchain-based tokens that each represent a unique asset such as a piece of art or media.

22. In your opinion, are any records in your organization retained past their disposition due date?

Of the 174 responses to this question, the vast majority (157, 90%) indicated that records were being held beyond the retention period set. Only 17 (10%) stated that they had no records held beyond the retention period (see figure 21).

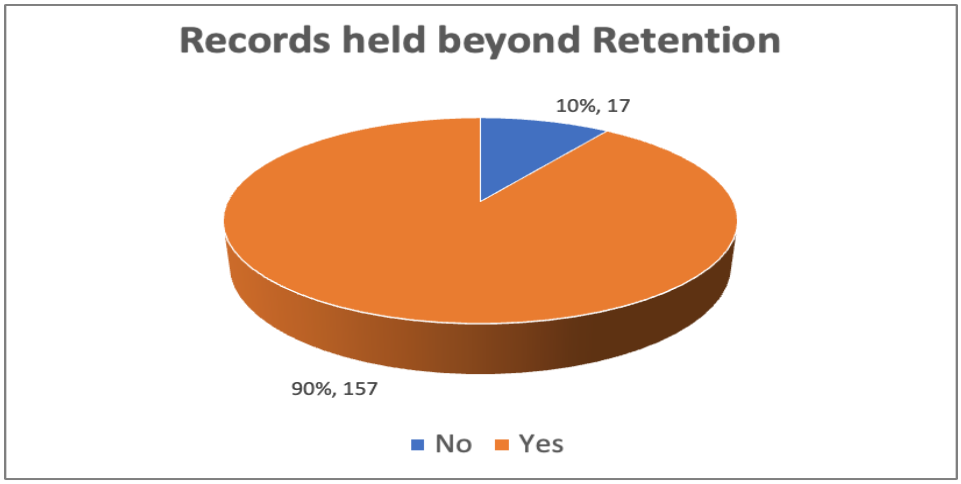


Figure 21. Number of respondents indicating their organization holds records beyond their retention date.

23. If yes, what steps might be taken to resolve this issue?

The responses were varied, but when the responses were categorized, they fell within four themes: Processes, Policies, People, and Systems.

Processes (Procedures). The greatest number of responses implied that new or modified processes (procedures) were needed; for example, several individuals indicated cumbersome manual disposition processes were in place. While some thought the answer was additional funding to employ more staff, others saw this as a need for automated disposition. One respondent cited the need to identify some classifications that can be disposed of automatically and destroy those records. Two individuals cited the need to improve the litigation hold process to enhance efficiency and reduce backlog. One respondent stated they are looking into creating an initiative to delete ROT and then will begin a path towards digital transformation.

***Policies*.** Several respondents indicated they had no records retention and disposition policy but were in the process of developing one. One respondent indicated the need to identify all copies and one single source of truth. One stated that there is currently no institution-wide records retention schedule, although some departments have their own. One stated they are working to identify records past their retention period and begin disposition of paper records. They then plan to begin working to identify and create a retention schedule for electronic records. Another indicated a need to complete the records retention schedule (RRS).

People. Both staff and upper management presented challenges to the retention and disposition of records. Several cited a need for additional funding to employ additional

staff. One respondent reported repeatedly requesting funding for additional staff to aid state agencies to meet retention timeframes, but this (according to the respondent) will be an ongoing request. Others cited the "need to keep everything attitude" of employees refusing to dispose of records, indicating a need for change management. A similar concern is management's refusal to promote the program. One respondent suggested this may change if something occurs to shift the thinking of senior management, such as a lawsuit or something else that costs money. Another frequent concern is the need to train and educate records creators and users. One respondent stated the need to educate employees that all information are records, not just what is in their EDRMS. Several respondents stressed the need for communications and collaboration with individuals in their department and other departments, such as Information Technologies and Archives.

***Systems*.** While some of the respondents believed they needed additional staff to help with deletion of records, others looked for technology solutions. One respondent stated they needed software that allows for deletion. One stated the solution for them would be to establish a connection between the ECMS and business systems where records are received or created—or to export those records to an ECM—and then carry out the disposition process. One individual cited a need to acquire tools to help manage legacy systems. Several others expressed a desire to utilize features of Microsoft 365, SharePoint, and Purview (a unified data governance solution to manage records, including retention and disposition).

Solutions

Rather than state what could be done, some of the respondents shared what they were doing to delete records that were retained beyond their disposition dates. Some cite specific products; others cite policy development, process improvement, and employee training/education. A select few comments that illustrate the various stages of maturity toward an effective records management program are shown below:

- The first set of retention schedules has officially been approved only at the beginning of 2022: it will take some time before their use may gain traction in the corporate environment.
- We will soon have our first records management policy, which will initiate training development around RM, and resources to support compliance and implementation.
- We are educating our employees and working with coordinators and IT to identify and dispose of old materials.

- We are just in the process of implementing disposition to our electronic records. When we migrate an application, we typically will apply retention at that point.
- We're in the process of a big cleanup process across shared drives, as well as planning to add retention policies and labels into M365 content (where the shared drive content is moving to).
- Comprehensive data mapping is a newer initiative that requires considerable cross-divisional cooperation. We are working towards that. My hope is to have a complete inventory and identified systems of record within the next two years.
- We've enabled retention policies in M365 and are building retention into systems and applications.
- Moving forward, … we will use technology to assist us in review and capture or disposal of current content assets.
- We are in the process of digitizing our physical records and then taking all our digital records and putting as many of them into SharePoint as possible. The rest will remain on our network drives. We plan to connect SharePoint (and possibly later the file shares) to Collabspace. The file shares will also be connected to Netwrix. Between these 2 tools, we can better classify, retain, and dispose of records based upon the retention requirements we have agreed upon as an agency that considers legal requirements, risk analysis, public interest, and staff needs.

24. Does your organization use any other systems for back-end tasks (e.g., Enterprise Resource Planning [ERP] solutions) such as supply chain management, financial management, project management, etc. that are sources of data/information/records that must be managed?

As shown in Figure 22, 40 (29%) of the 139 respondents to this question stated their organization did not use other systems that are sources of data/information/records that must be managed. However, 99 (71%) indicated that ERP systems were in place that contained data/information/records to be managed.

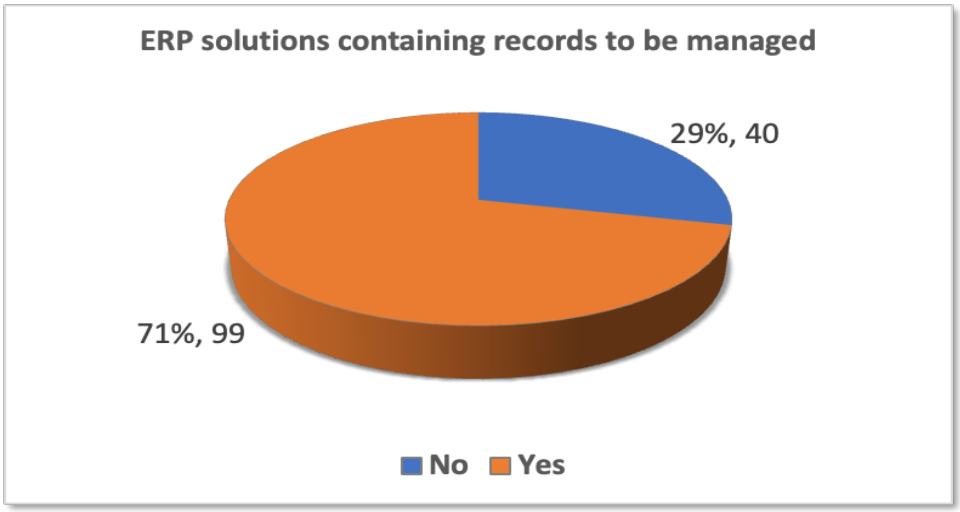


Figure 22. ERP solutions that contain digital content that must be managed.

25. If yes, which types of ERP systems? Do they support retention and disposition of content? A total of 94 individuals responded to this question, many indicating the use of multiple systems in their organization. Figure 23 illustrates the types of systems most often recognized as containing digital content that must be managed. The numbers shown are actual numbers and not percentages.

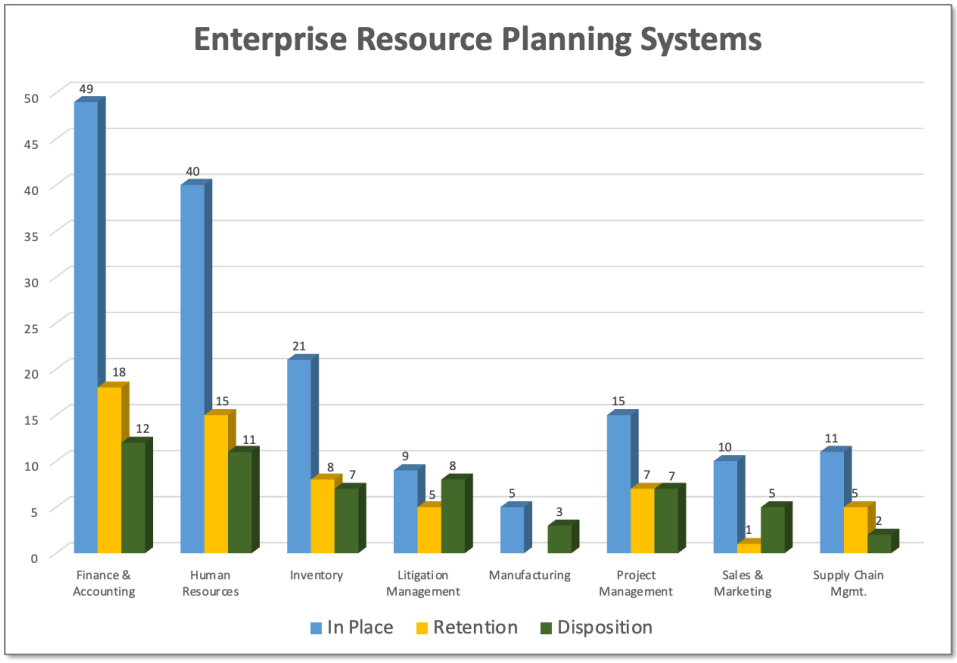


Figure 23. Enterprise Resource Planning Systems containing digital content that must be managed.

Finance and accounting systems were selected by the largest number of respondents, 79 of 94 (84%). Human Resource Systems is next, selected by 66 of the 94 respondents (70%). Inventory management (36, 38%), project management (29, 31%), and litigation management (22, 23%)

follow. The final three systems are supply chain management (18, 19%), sales and marketing (16, 17%), and manufacturing (8, 9%).

Retention and disposition features were not present in all systems, according to respondents. These features were indicated for less than half of the systems listed for finance and accounting as well as human resources. This would need further investigation, perhaps by interviewing the business units to better understand the features of the systems. Litigation Management stands out as having disposition available in all but one of the 9 systems cited. Why retention was indicated for only 5 of those 8 having disposition features also merits further exploration. Manufacturing is also worth noting, since 3 of the 5 systems contained disposition features.

Several respondents selected "other" and indicated the following: asset management, maintenance tracking, land management, procurement management, customer relationship management, student systems, M365, and one stating "lots of others."

26. Are you familiar with the International Council on Archives' requirements for records retention and disposition?

Of the 200 respondents to this question, 80 (40%) were familiar with the International Council on Archives' requirements for retention and disposition; 120 (60%) were not (see figure 24).

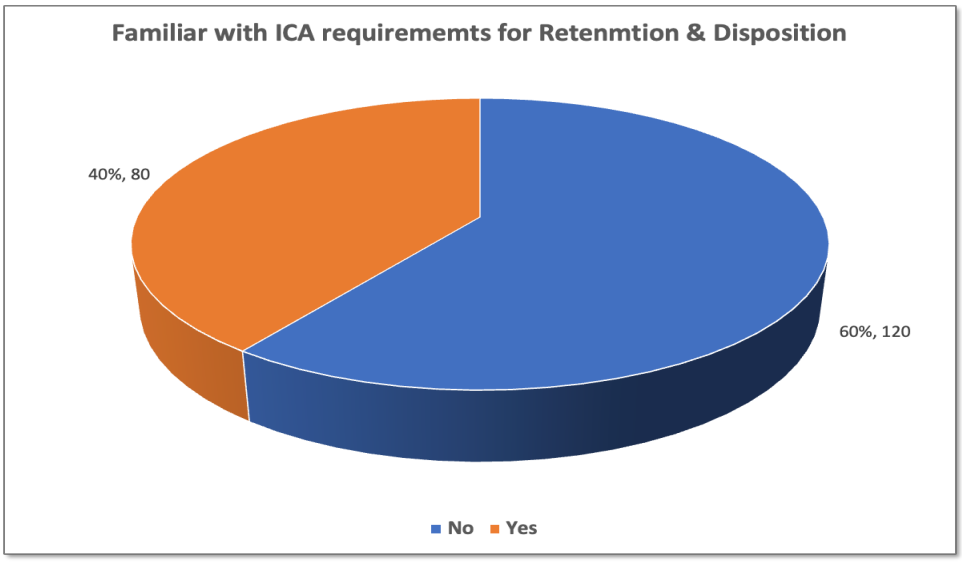


Figure 24. Number of respondents familiar with ICA requirements for retention and disposition.

27. If you answered yes to Question 26, do you believe your systems meet the ICA requirements for retention and disposition?

As shown in Figure 25, most respondents (76%) who stated they are familiar with the ICA requirements do not believe their systems meet those requirements for retention and disposition.

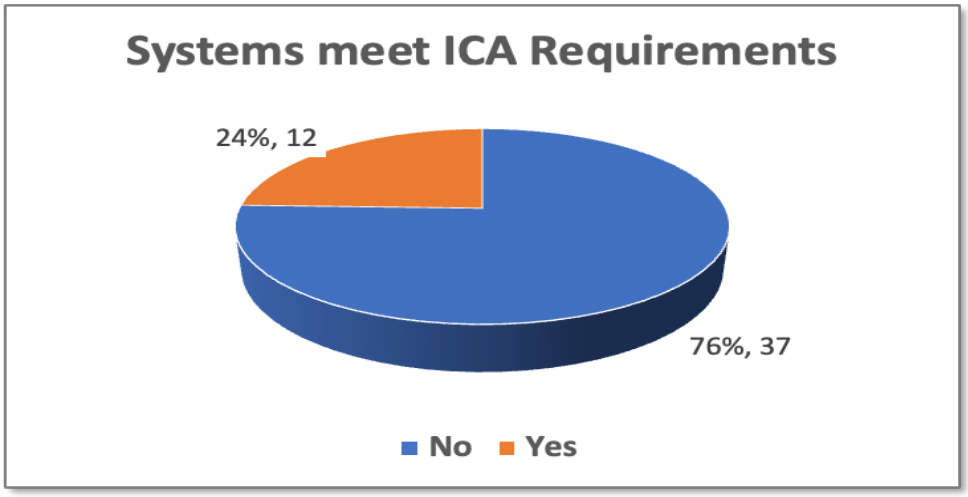


Figure 25. Responses indicating belief systems in place meet ICA requirements for retention and disposition.

Part 4: Artificial Intelligence

28. How familiar are you personally with the use of AI-based technology applied to recordkeeping (including archives and records management)?

As shown in figure 26, 200 respondents replied to this question. Of those, 75 respondents selected “not familiar at all” (38%), 67 selected “slightly familiar” (34%), 41 selected “moderately familiar” (21%), 12 selected “very familiar" (6%), and 5 selected “extremely familiar” (3%).[1]

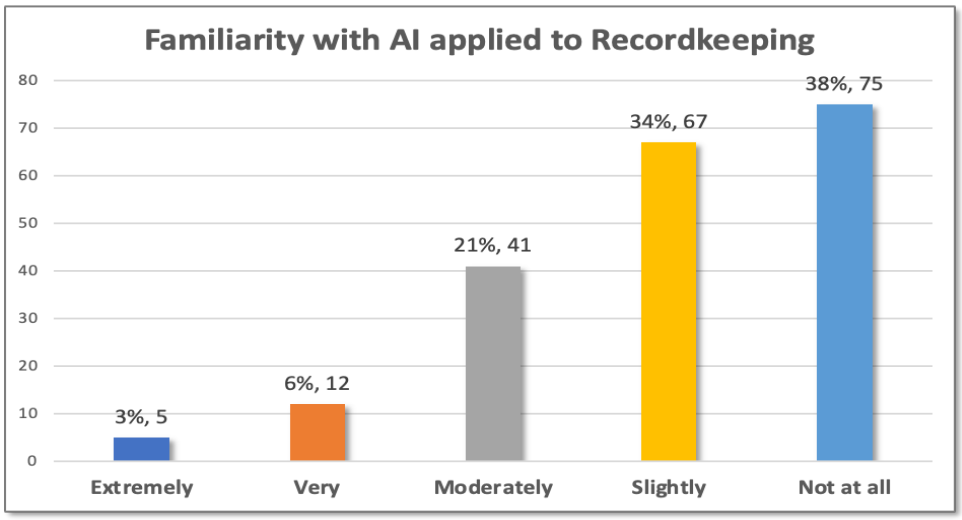


Figure 26. Respondents’ level of familiarity with artificial intelligence applied to recordkeeping.

More than half of the respondents (72%) reported being either slightly or not at all familiar with artificial intelligence applied to recordkeeping tasks. Only 9% reported being very or extremely familiar with artificial intelligence applications for recordkeeping.

29. What is your definition of Artificial Intelligence?

171 respondents answered this question, providing their definitions of AI. While responses varied, several commonalities emerged across definitions:

- Applications that simulate human intelligence or otherwise perform tasks associated with human intelligence, such as analysis and prediction
- “Machine learning” (many respondents either define AI as “machine learning,” or include machine learning in their definition)
- Applications that can generate predictions and decisions based on inputs
- Applications that can analyze large data sets
- Applications that can perform tasks more efficiently than humans

[1] Percentages total 102% due to rounding percentages to whole numbers.

- Applications that can improve based on new feedback
- Applications that can make decisions “on their own,” or without human intervention or feedback

A selection of the 171 responses is presented in Appendix 1.

30. Does your institution/organization use AI-supported activities in the management of data/information/records?

Of the 202 responses to this question, 129 selected “no” (64%), 41 selected “don’t know” (20%), and 32 selected “yes” (16%), as shown in figure 27.

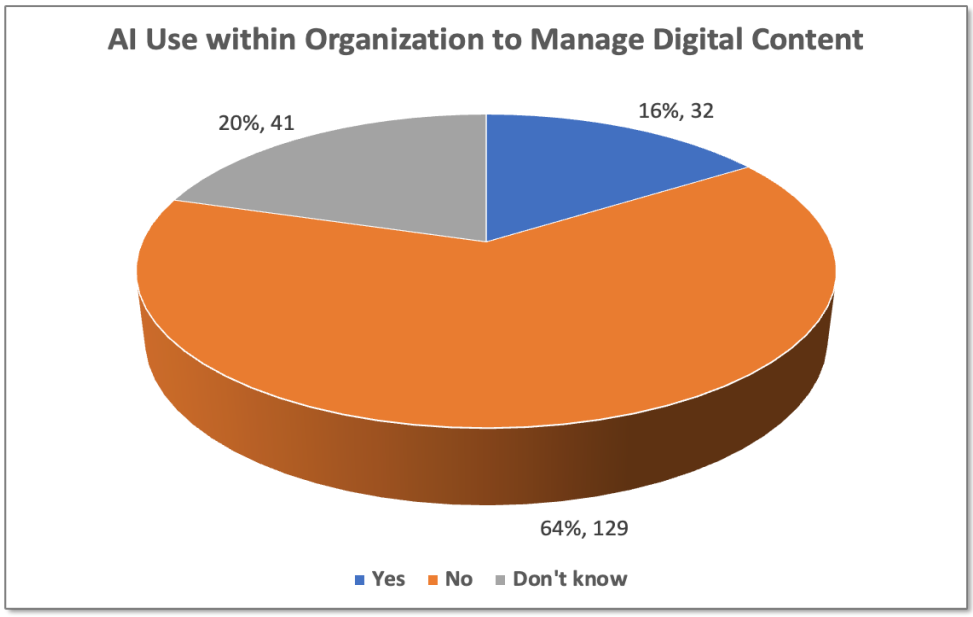


Figure 27. Use of AI to manage digital content.

31. If yes, for which of the following tasks does your organization use AI-supported activities?

As shown in figure 28, 31 respondents replied to this question by selecting applicable tasks from a list provided.

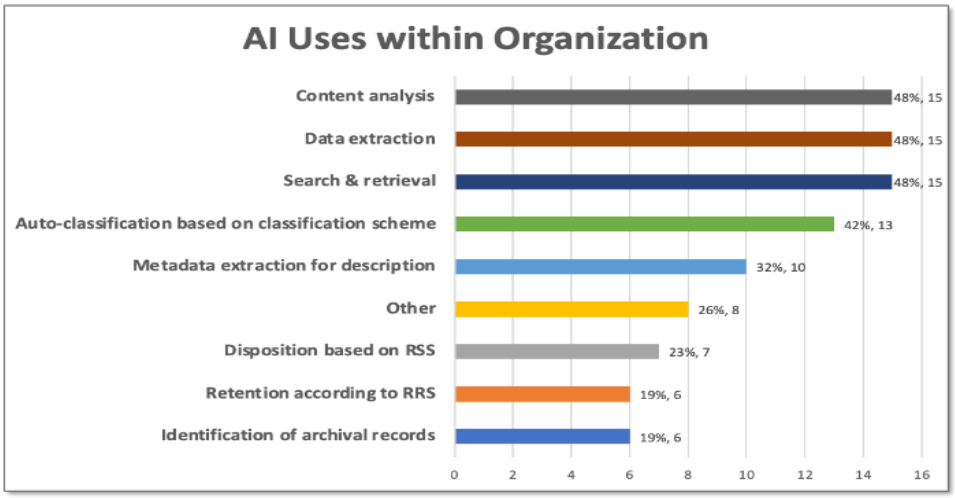


Figure 28. Uses of AI application within organizations.

As shown in figure 28, content analysis, data extraction, and search and retrieval were the most common uses, followed closely by auto-classification based on classification schemes. Disposition and retention based on records retention schedules and identification of archival records were the least common uses of AI within the organizations represented.

Of the respondents who selected "other," 7 specified additional uses:

- All of the above
- Chat bots
- Creation of archives manuals
- Grid / Market Ops likely use
- Presentation of related party information to clear conflicts
- Research and model development tool
- Search engine

32. From the perspective of information governance, who in your organization does/would you expect to take the lead in employing AI-based technology to improve recordkeeping in the workplace?

Respondents were asked to select from a list of 8 options and "other." As shown in figure 29, 198 responses were received.[2]

[2] Percentages total 101% due to rounding percentages to whole numbers.

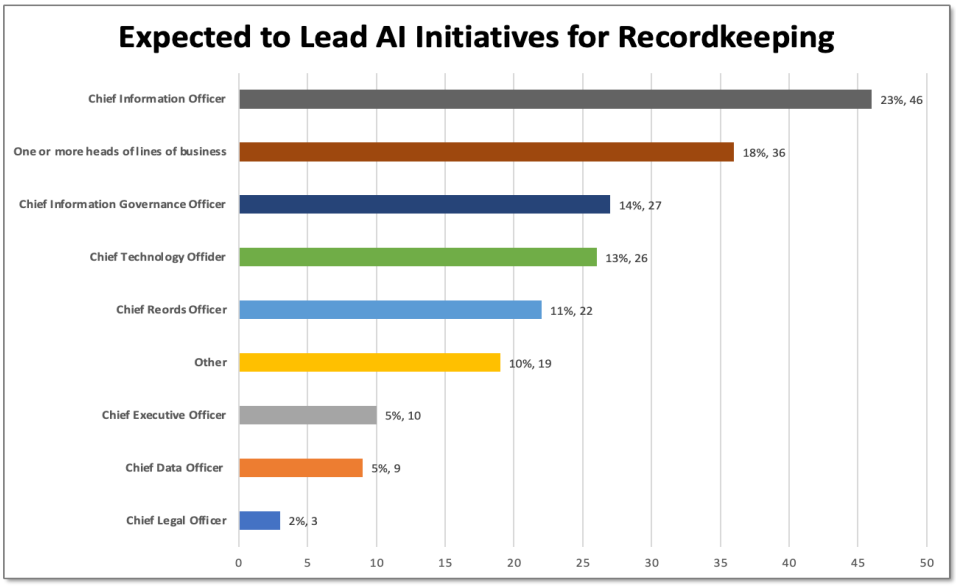


Figure 29. Titles of individuals respondents believed would most likely lead AI initiatives for recordkeeping within their organizations.

The "other" responses are reproduced below:

- The Leader of the Digital Transformation Team
- IT Director and myself (records manager)
- Me, in partnership with our vendor
- Director of Information Technology
- Provincial archivist
- Archivist of Manitoba
- Director of Records Management
- Unknown… It may even be me.
- Up for debate. IT will say it's their domain.
- Me
- Data Governance Officer
- Archivist & Privacy Officer
- Hopefully nobody
- Director of IT & Info Services
- Director
- CFO
- Information Management Officer & Director of IT

33. To the best of your knowledge, does your institution/organization use AI-supported activities in digital preservation processes?

Of 200 respondents, 154 selected "no" (77%), 36 selected "don't know" (18%) and 10 selected "yes" (5%), as shown in figure 30.

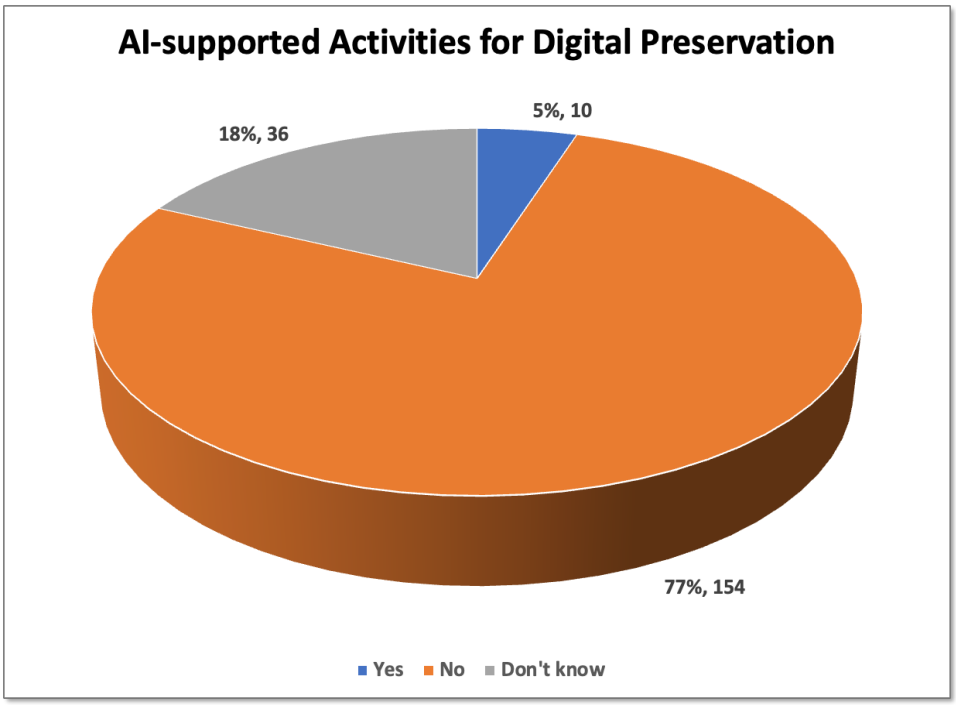


Figure 30. AI-supported activities in use within organizations for digital preservation.

34. If yes, briefly explain (name and describe the process)

Six responses were received and are reproduced in full below:

- We have an open-text auto-classification system that we are teaching how to determine which emails need to be saved and which don't. We "feed it" exemplar emails to "show it" the kind that don't need to be saved. It focuses on the emails of high-level officials, and it routinely disposes of emails that can be destroyed.
- SCADA used to monitor water flow, reclamation & related activities.
- Procesamiento digital y uso de herramientas tecnológicas. [Digital processing and use of technological tools.]
- Todo el sistema de Gestión documental, es amigable, y se esta implementando en todas las áreas laborales. [The entire document management system is friendly and is being implemented in all work areas.]
- Automated digital preservation with intelligent application of policy and updates.
- Elaboración de registros donde se concentra la información generada. [Preparation of records where the information generated is concentrated.]

35. Does your institution/organization plan to employ AI-enabled technologies in the future?

As shown in figure 31, of 106 responses, 37 selected “yes, for the first time” (35%), 31 selected “yes, additional AI tools” (29%), 21 selected “no” (20%), and 17 selected “other” (16%).

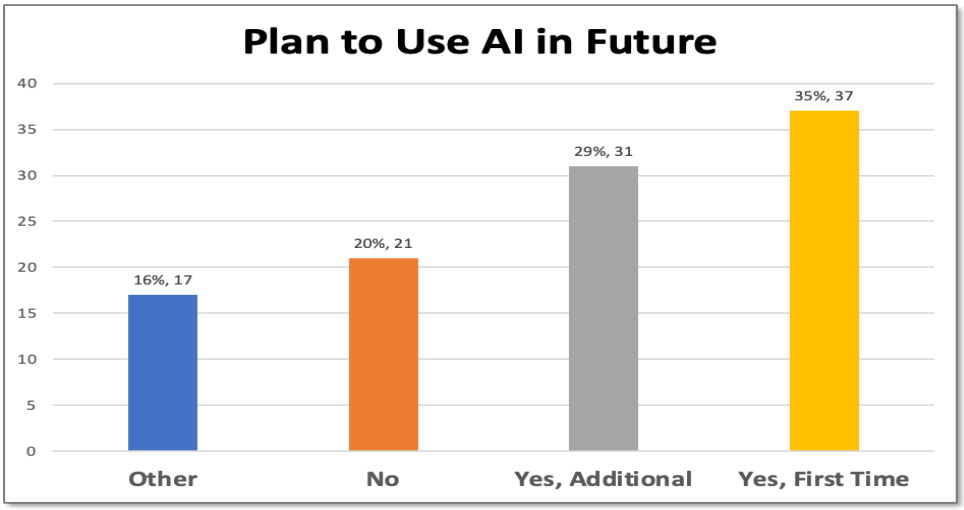


Figure 31. Responses indicating plans to use AI-enabled technologies in the future.

More than half (68%) reported plans to either use AI for the first time or to add artificial intelligence initiatives to those already in place. Although 20% did not have plans to use AI in the future, the 16% that selected the other option reported a mix of sentiments rather than knowledge—ranging from the hope their institution would employ AI to the hope that it would not. Those 15 responses are shown in figure 32.

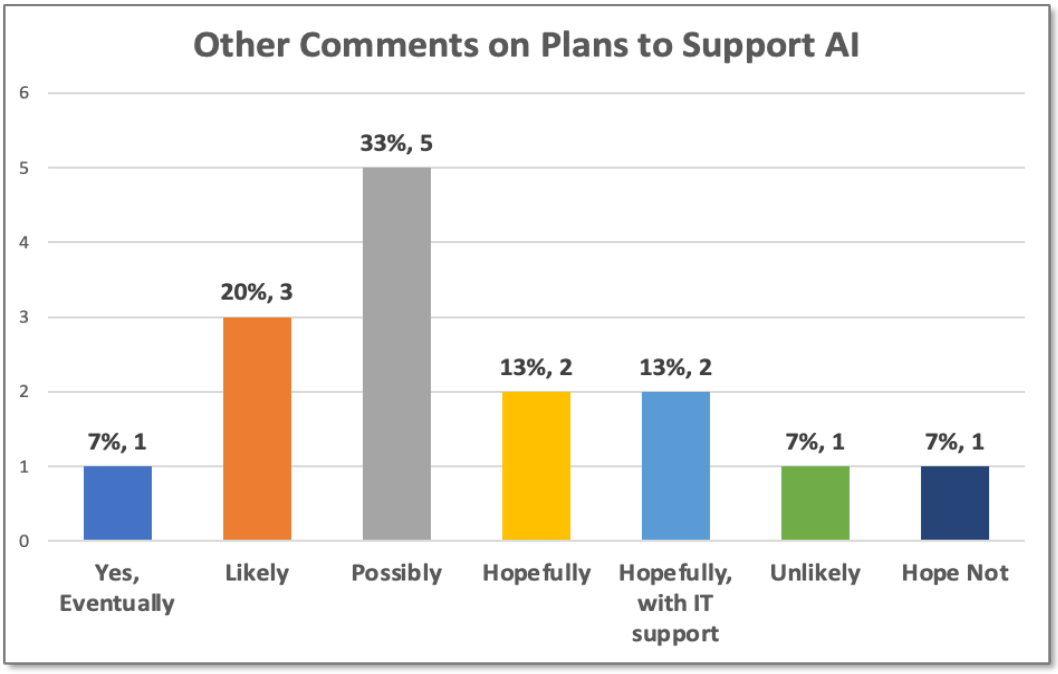


Figure 32. Other responses to the question asking about plans to support artificial intelligence within the organization.

36. For which of the following reasons would your institution/organization employ AI tools or technology?

Respondents were asked to select five items from a list of pre-selected options. Responses are shown in figure 33.

Of the 193 respondents, the greatest number selected “Increased efficiency” (167, 87%), with 140 selecting “Improved productivity” (73%), 122 selecting “Better quality/reduced human error” (63%), 117 selecting “Reduced workloads” (61%), 94 selecting “Improved monitoring” (49%), 93 selecting “Improved decision making” (48%), 82 selecting “Better customer service” (42%), 59 selecting “New capabilities and business model expansion” (31%), 44 selecting “Enhanced communication” (23%), 25 selecting “Better talent management (e.g., hiring, productivity)” (13%), and 22 selecting “other” (11%).

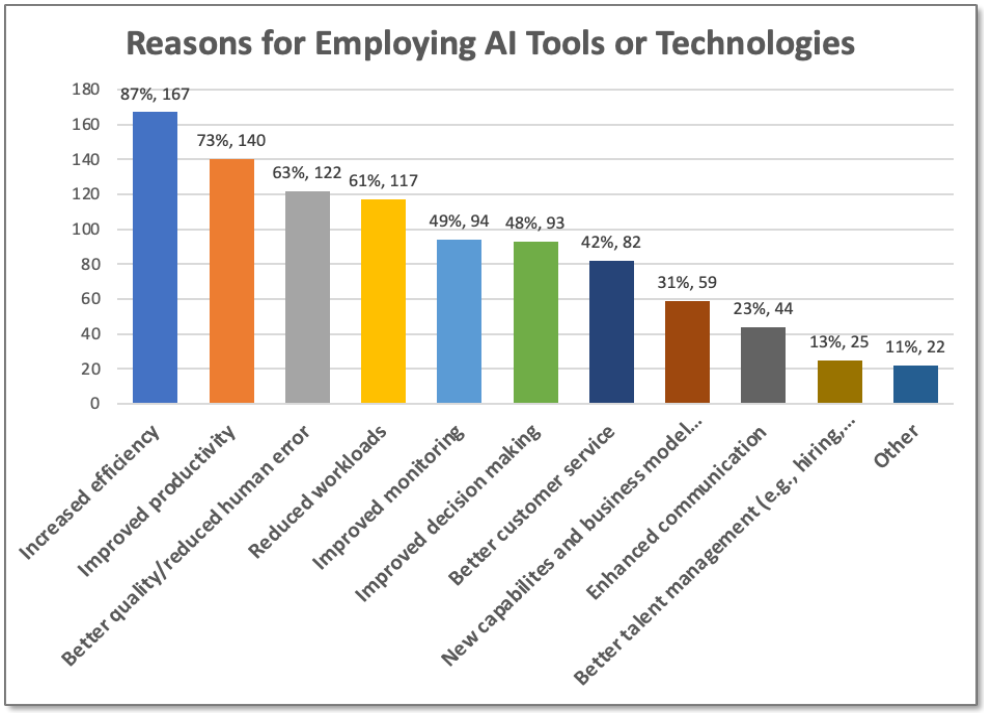


Figure 33. Benefits of the use of AI within the organization.

Respondents who selected “other” had the option to provide information in free text. As shown in figure 34, of these respondents, 5 cited improved compliance; 4 cited data management/records management; 3 cited innovation or improved quality of services; and 2 cited data analysis or research.

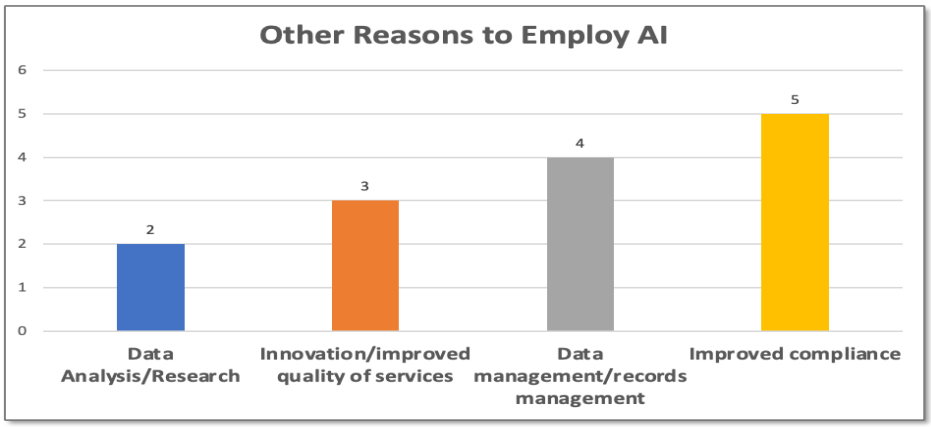


Figure 34. Additional reasons cited for the use of AI within the organization.

Insights and Recommendations

Survey Demographics

Insight: Based on respondents' declared roles and areas of specialization and expertise, many respondents occupying the role of "archivist" and "records manager" are actually employing a wider range of skills and expertise beyond what is traditionally associated with those roles; namely, competencies in the field of information governance that extend beyond traditional archival and records management duties.

Recommendation: There is a need for institutional change to catch up to and recognize that archivists and records managers are capable of doing more than their traditional roles entailed. In particular, archivists and records managers are moving increasingly into information governance roles, and this change should be recognized and addressed at institutional levels.

Data, Information, Records Systems

Insight: For the maintenance of long-term digital records, high-level policies and retention schedules are common in institutions, but more granular systematic documentation of processes is rare.

Insight: Many systems in use for managing data, information, and records are more traditional systems (e.g., Document Management Systems, Records Management Systems) compared to the most current technology according to scholarly literature and best practices, such as content services platforms. A large number and variety of applications and technologies are being used to manage information and records, with very little standardization across institutions.

Insight: There is also very little standardization of preservation strategies and technologies.

Recommendation: Institutions should build out more systematic documentation to connect high-level policies to granular activities and processes; documentation of the more granular processes and activities may be an important element of automation and effective AI. Information professionals must continue to advocate for the development of preservation standards and technology that leverages the power of new technologies, including artificial intelligence.

Retention and Disposition

Insight: Archivists and records managers are dealing with a large number of file formats and content types. This fact, along with the large number of systems in use within their organizations, poses challenges to both records managers and archivists.

Insight: 90% of respondents say they have records retained beyond the disposition date. The challenge of records being kept beyond disposition due date is primarily due to the unmanageable number and variety of types of records.

Recommendation: Respondents identified the need for 4 types of solutions—processes, policies, people, and systems. Of these solutions, a big takeaway is that systems need to be automated—systems are retaining records, but not disposing of them. Successful automation of disposition actions for a variety of record types will rely on the effective documentation of granular activities and processes.

Artificial Intelligence

Insight: It is common for institutions to have plans or at least aspirations to employ AI solutions in recordkeeping, but there is very little familiarity with AI for recordkeeping among records managers and archivists.

Recommendation: Employee training programs must be designed to support the organization's AI strategy and may address AI basics; risks and benefits of the use of AI; available AI tools and technologies to facilitate recordkeeping, and effective use of AI tools and technologies made available to them.

Recommendation: In addition to increasing awareness of the risks, benefits, and potential uses of AI in recordkeeping, institutions and archivists/records managers should endeavor to understand how their current and historical recordkeeping processes may inform AI models and how to effectively document existing workflows and decision points for future incorporation into AI solutions.

Acknowledgements

InterPARES Trust[AI] Researchers, Assistants, and Partners

Primary Investigator: Dr. Patricia C. Franks, San Jose State University

Investigators: Jason R. Baron, University of Maryland; Dr. Lois Evans, Land Title and Survey Authority of British Columbia; Patricia L. Moore, Carleton University; Samuel Tweneboah-Koduah, Gannon University; Sarah Weldon, Iron Mountain

Research Assistants: Melissa Kemp, San Jose State University; Michael Carelse and Linnet Chappelka, University of British Columbia

Industry Partners:

Robert M Gerbrandt Managing Director, IG Consulting Global Head of Information Governance Advisory	**Graham Sibley** Chief Executive Officer

Organizations that confirmed their assistance disseminating information about the survey.

Academy of Certified Archivists (ACA)
American Medical Informatics Association (AMIA)
ARMA International
ARMA Calgary
ARMA Triangle Chapter
ARMA Vancouver Island
Association of Canadian Archivists (ACA)
Association for Intelligent Information Management (AIIM)

Canadian Health Information Management Association (CHIMA)
Certified Information Governance Officers Association (CIGOA)
Council of State Archivists (CoSA)
International Association of Privacy Professionals (IAPP)
IG Guru (Information Governance News)
Institute of Certified Records Managers (ICRM)
Information and Records Management Society Ltd. (IRMS)
National Association of Government Archives and Records Administrators (NAGARA)
Society of American Archivists (SAA)
Society of Georgia Archivists
Society of Information Risk Analysts (SiRA)
Society of North Carolina Archivists

About ItrustAI

InterPARES Trust AI (2021-2026) is a multi-national interdisciplinary project co-directed by Dr. Luciana Duranti (Principal Investigator) and Dr. Muhammad Abdul-Mageed of the School of Information, University of British Columbia, Canada. The aim of the research project is to design, develop, and leverage Artificial Intelligence to support the ongoing availability and accessibility of trustworthy public records by forming a sustainable, ongoing partnership producing original research, training students and other highly qualified personnel (HQP), and generating a virtuous circle between academia, archival institutions, government records professionals, and industry, a feedback loop reinforcing the knowledge and capabilities of each party.

The ITrust^{AI} goals are to:

1. Identify specific AI technologies that can address critical records and archives challenges;
2. Determine the benefits and risks of using AI technologies on records and archives;
3. Ensure that archival concepts and principles inform the development of responsible AI; and
4. Validate outcomes from Objective 3 through case studies and demonstrations.

Learn more about ITrust AI at https://interparestrustai.org/trust

Appendix 1: Selective Comments: What does AI mean to me?

A total of 171 of the 214 respondents who completed this survey provided their own definition of AI. Below are some of those definitions.

#1.
AI is the study of, engineering of computer systems, or the systems themselves, that simulate or instantiated intelligence, human or otherwise, where intelligence is a catch all for the full list of cognitive terms such as: reasoning, inferring, predicting, planning, understanding, explaining, perceiving, speaking, deciding, acting, and learning. (respondents' synthesis of multiple definitions)

#2.
A system or program that can learn and extrapolate from its initial store of information and instructions, to complete tasks or to communicate with users in an increasingly disconcerting imitation of human speech and decision-making patterns.

#3.
AI and ML (machine learning) are when an algorithm can be developed to scan and appraise a set of data and then "learn" from it. A human can tell the AI whether it is correct or incorrect and it can then improve its algorithm over time based on input.

#4.
It is the combination of algorithms set up with the purpose of creating machines that have the same capabilities as humans.

#5.
Artificial Intelligence is any technological learning capability. AI does not create intelligence, but it instead processes extant information to create accurately predictive algorithms as applied to new information.

#6
Artificial Intelligence (AI) is a branch of computer science that focuses on creating machines and software capable of performing tasks that would normally require human intelligence, such as visual perception, speech recognition, decision-making, and language translation.

#7
It is a field of science that works with a set of concepts, technologies and algorithms that allow computers to reproduce some characteristics of human reasoning and behavior.

#8
Hold on, let me ask ChatGPT :)

Appendix 2: List of Professional Associations

The survey was sent to the following professional associations, with requests to circulate the survey among their membership. *Thanks to all who shared the information without letting us know* 😉

Academy of Certified Archivists (ACA)
AIIM (Association for Intelligent Information Management)
American Library Association (ALA)
ARMA Arizona
ARMA Calgary
ARMA Edmonton
ARMA International
ARMA Montreal
ARMA New Brunswick
ARMA Ottawa
ARMA Saskatchewan
ARMA Southwestern Ontario Chapter
ARMA Triangle Chapter
ARMA Vancouver
ARMA Vancouver Island
ARMA Winnipeg
Association of Canadian Archivists (ACA)
Association of Moving Image Archivists (AMIA)
Building Industry Consulting Services International (BICSI)
Canadian Health Information Governance Association (CHIMA)
Canadian Information Processing Society (CIPS)
Canadian Library Association (CLA)
Certified Information Governance Officers Association (CIGOA)
Computing Technology Industry Association (CompTIA)
Council of State Archivists (CoSA)
First Nations Information Governance Centre (FNIGC)
IEEE Computer Society
Information and Records Management Society (IRMS)
Information Systems Audit and Control Association (ISACA)
Information Systems Security Association (ISSA)
Institute of Certified Records Managers (ICRM)
International Association of IT Asset Managers (IAITAM)

International Association of Privacy Professionals (IAPP)
Michigan Archival Association
Midwest Archives Conference
National Association of Government Archives and Records Administrators (NAGARA)
New England Archivists
Nuclear Information and Records Management Association (NIRMA)
RIMPA Global (was Records and Information Management Professionals Australasia)
Risk Management Association (RMA)
Risk Management Society (RIMS)
Society for Information Management (SIM)
Society of American Archivists (SAA)
Society of California Archivists
Society of Florida Archivists
Society of Georgia Archivists
Society of Information Risk Analysts (SiRA)
Society of North Carolina Archivists
Society of Ohio Archivists
Society of Rocky Mountain Archivists
Society of Tennessee Archivists