Training in Computational Archival Science: Do CAS Educational Frameworks meet Professional Expectations?

Victoria L. Lemieux
School of Information
University of British Columbia
Vancouver, Canada
Orcid ID: 0000-0003-1339-6289

Richard Arias-Hernandez
School of Information
University of British Columbia
Vancouver, Canada
Orcid ID: 0000-0003-2939-1116

***Abstract*— This paper explores the evolving landscape of training for archival professionals in the context of big data and emerging technologies. By comparing two educational frameworks—the CAS framework, developed from computational thinking research and CAS research papers, and the InterPARES framework, based on empirical studies with archivists working with AI/ML, we identify areas of alignment and divergence. While both frameworks share significant concordance, suggesting a growing consensus on integrating computing into archival work, key differences in their approaches (learning outcomes vs. competencies) and focus areas (such as work practices, systems thinking, and cybersecurity) highlight the need for further discourse among archival scholars, educators, and practitioners. These distinctions must be addressed before formalizing CAS educational frameworks. This paper also initiates efforts to integrate emerging technological competencies by bridging the CAS and InterPARES frameworks, emphasizing the value of complementary perspectives from both professional practice and academic research. We argue that such integration is essential for developing robust competency frameworks in archival education, particularly within higher education's professional programs.**

Keywords—computational archival science, education, educational frameworks, archives, artificial intelligence

I. Introduction

Computing technology has transformed archival work over the past few decades with the emergence of born-digital records and the need for archivists to manage and preserve these materials. Digitization has made archival materials more accessible, easier to preserve and possible to manage and explore using novel technologies, such as artificial intelligence (AI). These changes, however, also come with many new challenges.

To address both the opportunities and challenges of technology in archives, a new field called Computational Archival Science (CAS) was established in 2016. CAS integrates computational methods into archival science and focuses on applying computational techniques to improve the efficiency, productivity, and precision of archival work. The development of CAS has been supported through workshops and the creation of new educational frameworks to respond to advances in technology.

While new CAS frameworks have emerged and been presented at IEEE CAS since 2016, there is little empirical data validating these frameworks. This paper seeks to advance the discussion on CAS educational frameworks by presenting empirical data from interviews with archival professionals on the technical skills they perceive as necessary for archivists seeking to apply AI to archival work. Though not a one-to-one match with CAS, in that CAS is broader than using AI alone, the survey provides a first glimpse of how well CAS educational frameworks meet professional expectations.

This paper proceeds as follows: In section one we briefly review the literature relating to computation in archival education, and particularly the demand for technical skills in the performance of archival work. In the following sections, we discuss the methodologies used to develop, and the content of, two proposed educational frameworks for technical training of archival professionals: a Computational Archival Science educational framework and an AI competency framework. We then turn to a comparative analysis of the two frameworks. We conclude our paper with some discussion of limitations and future directions.

II. Background Literature Review

Computing has been transforming archival material and work for several decades now, but it has only been relatively recently that writers on archives have begun to grapple with the question of what kind of training is needed for professionals entering the field in the digital age.

Lemieux and Hofman's survey of archival education literature from North America, Europe, and Africa [1], noted that several papers written between 2001-2017 expressed concern over whether current archival education was adequately preparing archival professionals to handle digital records and archives, especially in the context of increasing digital data [2], [3]. Studies like Lyon et al. [4] highlighted the demand for data-related archival roles, suggesting the need for STEM graduates in archival programs to develop technical and quantitative skills. The importance of practical experience in applying theoretical knowledge to digital records was also noted [4].

It was against a growing realization that archival professionals needed more training in computational knowledge and techniques that Marciano et al. [5] advocated for a new transdisciplinary field, Computational Archival Science (CAS), in response to the rise of big data in archives. Collaborators in the development of the vision for this new field devised a working definition of CAS as:

> *A transdisciplinary field that integrates computational and archival theories, methods, and resources, both to support the creation and preservation of reliable and authentic records/archives and to address large-scale records/archives processing, analysis, storage, and*

access, with the aim of improving efficiency, productivity, and precision, in support of recordkeeping, appraisal, arrangement and description, preservation and access decisions, and engaging and undertaking research with archival material [5].

An inaugural workshop on CAS (see https://ai-collaboratory.net/cas/) was launched in 2016 and has been held annually during the IEEE Big Data Conference. The CAS workshop is now in its 9th year. To date, 102 papers featuring the application of computing to archival work and exploring its implications have been presented.

In 2022, Mark Hedges and Richard Marciano, two of the founders of CAS, and Eirini Goudarouli, guest edited a special issue of the *Journal on Computing and Cultural Heritage* [6]. The special issue was one of the outcomes of a one-year international research networking grant for UK-US collaborations on digital scholarship in cultural heritage institutions, funded by the Arts and Humanities Research Council (AHRC) of the UK in 2019. In the special issue, the editors outline the path of development of CAS and featured several articles aimed at advancing and exploring the ever-expanding application of computing to archival work. The articles explored current debates and future perspectives on the application of AI in archives [7] using machine learning to improve responsiveness to freedom of information requests [8]; and applying computational techniques to oral history archives and in the selection of records for preservation [9], among other topics.

Recognition in the UK of the need for training in computational techniques among archivists has also come from the AEOLIAN network (Artificial Intelligence for Cultural Organisations), funded by a joint program between the US National Endowment for the Humanities (NEH) and the Arts and Humanities Research Council (AHRC) in the UK and led by Dr. Lise Jaillant [10]. Project Partners included the National Library of Scotland; the National Library of Wales; the Wellcome Collection; the History of Parliament Trust; Harvard's Houghton Library; Yale's Digital Preservation team and Music Library; Indiana University Libraries; University of North Carolina at Chapel Hill Libraries; Educopia; and the Frick Collection (NYC). The network produced six workshops, a forthcoming volume of case studies, and two journal special issues [11], [12] on the application of AI/ML in cultural heritage contexts. Focusing on researchers and other users of digital archives, Jaillaint and Aske [13] conducted a survey and semi-structured interviews with experts—including archivists, librarians, digital humanists, literary scholars, historians and computer scientists—to identify the challenges that humanities and social science scholars face when using computational methods and tools, and to offer suggestions to enhance users' computational abilities and improve digital infrastructure for these activities with a view to making archives more useable and accessible. While these findings are applicable to archivists as well, there remains a need for additional research into how best to train archivists in new computational methods and tools.

III. THE CAS COMPETENCY FRAMEWORK.

In 2018, Underwood et al. [14] introduced the first CAS framework mapping computational thinking into archival science education to address the technical knowledge and skills gap in the archival profession. The framework they presented sought to move beyond what most graduate and continuing education programs in archival studies were doing to address the gap in introducing students to information technology as it relates to digital records by proposing an approach based on infusing computational thinking into the graduate archival studies curriculum. Underwood et al.'s framework [14] drew upon Weintrop et al.'s taxonomy of computational thinking [15], formulated from a systematic review of literature, interviews with mathematicians and scientists, and exemplary computational thinking instructional materials as part of an effort to introduce computational thinking into high school science and mathematics curricular materials. The framework consists of 22 computational thinking practices grouped into four main categories: data practices, modeling and simulation practices, computational problem-solving practices, and systems thinking practices.

The following year, Underwood and Marciano [16] elaborated on these initial ideas, presenting several case studies – i.e., review of records for personally identifiable information (PII); development of digital preservation systems research within the InterPARES project; determination of the trustworthiness of recordkeeping systems research within the InterPARES project; and categorization of email for purposes of records retention – that validated the relevance of the taxonomy for archival training by illustrating how archival work already demonstrated the kinds of computational thinking practices used by mathematicians and scientists. Having validated the alignment of this framework with archival use cases, Underwood et al. (see, cases.umd.edu, n.d.) subsequently used it to develop a series of case studies for the training of archivists in graduate programs on archival studies. Each contributed case was designed to consider (1) archival practices, (2) computational thinking practices, and (3) ethics and values. The development of the cases sought to promote: (1) open-source research and educational platforms, (2) cloud-based student learning environments, (3) new pedagogies for educating archivists in computational methods and tools, and (4) establishing a community of practice for sharing computational and archival knowledge. The contributed cases engaged with topics, such as (1) evolutionary prototyping and computational linguistics; (2) graph analytics, digital humanities, and archival representation; (3) computational finding aids; (4) digital curation; (5) public engagement with (archival) content; (6) authenticity; (7) confluences between archival theory and computational methods: cyberinfrastructure and the records continuum; and (8) spatial and temporal analytics.

Based on this foundational CAS work, analysis of the 102 papers presented at IEEE Big Data CAS since its inception, and their experiences with training students in computational archival science, Lemieux and Marciano [17] extrapolated and advanced the following competency framework for CAS:

<u>CAS Graduate Competencies</u>

<u>Technical</u>

1. Graduates possess technical knowledge and understanding enabling them to apply computational thinking, tools and methodologies to effectively analyze and preserve digital records and support archival work. Specifically, graduates have the ability to:

1.1 apply knowledge of technologies to real world record creation and recordkeeping problems and situations

1.2 demonstrate an understanding and skill in "archives-as-data" practices, such a collecting data, creating data, manipulating data, analyzing data and visualizing data

1.3 demonstrate an understanding and skill in how to apply modeling and simulation practices to archival work, such as using modeling to understand a concept, using computational models to find and test solutions, assessing computational models, designing computational models, and constructing computational models.

1.4 within diverse contexts, demonstrate an understanding and skill in computational problem solving, including appreciating when it is appropriate to apply particular computational tools and techniques to specific archival tasks, preparing archival problems for computational solutions, programming, choosing effective computational tools, assessing different computational tools, developing modular computational solutions, creating computational abstractions, and troubleshooting and debugging.

1.5 demonstrate the ability to think in terms of systems and apply systems thinking practices in archival work, including investigating a complex system as a whole, understanding the relationships within a system, thinking in levels, communicating information about a system, and defining systems and managing complexity.

<u>Records/Data</u>

2. Graduates are able to apply the foundational theories, methodologies and techniques of computational archival science to understand the effects of computing tools and methods on digital records and recordkeeping. Specifically, graduates have the ability to:

2.1 demonstrate an ability to understand and apply foundational computational archival science theories to gain insights into the nature and evolving forms of digital records and recordkeeping.

2.2 understand and critically evaluate the application of computational techniques and tools, such as those used in processes of "datafication", in relation to their impact upon the trustworthiness of records, such as their accuracy, reliability and integrity.

2.3 understand and critically evaluate the impact of applying computational techniques and tools in relation to the selection and acquisition of records for preservation.

2.4 understand and critically evaluate the impact of applying computational techniques and tools to sensitive information and its consequences for privacy.

2.5 demonstrate an understanding and ability to conduct risk assessments of threats and vulnerabilities to archival cyberinfrastructures and records and develop risk mitigation plans and strategies.

<u>Social</u>

3. Graduates are able to appreciate foundational theories, methodologies and techniques of computational archival science to understand the effects of computing tools and methods on society. Specifically, graduates have the ability to:

3.1 demonstrate an ability to critically examine the discursive aspects of data as well as the inherent power dynamics and cultural positionality manifest within computational infrastructure and the logics of computing.

3.2 assess the social impact and usability of emerging technologies for archival purposes.

3.3 draw upon knowledge of law and ethics in relation to all functions and activities related to applying computing technology to records and archives.

3.4 engage critically and creatively with emerging professional and societal issues, such as bias in training data, arising from the application of computing technology to archival work

3.5 advocate for the responsible and ethical application of computing techniques and tools in archival work.

While the above framework synthesizes much previous work on archival training in CAS, it still lacks empirical validation with the archival profession.

IV. Identifying Artificial Intelligence and Machine Learning Competencies with and for Archival Professionals: An Empirical Study

To address the lack of professional validation of CAS education frameworks, it is possible to draw upon a recently completed study done by InterPARES Trust AI researchers Arias-Hernandez, Fewster and Penniman [18]. In this study they identify professional competencies for archival professionals to make artificial intelligence and machine learning (AI/ML) work for the processing of records. A distinctive feature of Arias-Hernandez et al.'s study [18] was that their research methodology intentionally targeted archival professionals with practical experience applying AI/ML to the processing of records as sources of data. The authors focused on the needs for professional training (rather than academic) and privileged the voices of archivists and records managers for their empirical study.

Two research stages in their study engaged with these participants. First, in-depth semi-structured interviews were conducted in 2023 with 10 archival professionals who had practical experience designing, adapting, or applying AI/ML for the processing of records in actual archives or records management offices [18]. This first stage provided the empirical data for identifying the relevant knowledge, skills, and attitudinal competencies. Second, two workshops, aimed to validate the competencies identified in the previous stage, were organized by Arias-Hernandez and Fewster during the Summer of 2024 in archival conferences: The Association of Canadian Archivists' Annual Meeting 2024 in Edmonton, AB, and Expotecnologia 2024 in Medellin, Colombia. These workshops engaged with around 280 archival professionals and students from Canada, Mexico, Costa Rica, Colombia, Ecuador, Peru, Brasil, Venezuela, Argentina, and Chile. The feedback obtained from attendees at these workshops was used by the researchers to refine the competencies extracted from the interviews, identify priorities for curriculum development, and provide some reflections and discussions on their situatedness in different contexts [19]. The current, validated, and refined competencies from this study are presented below in Tables 1-3. Detailed description of the empirical study and its methodology can be found in Arias-Hernandez et al. [18]. Details on their validation and refinement through workshops, detailed explanation on each competency, and a theoretical discussion on their situatedness can be found in Arias-Hernandez and Rockembach [19].

TABLE I. KNOWLEDGE-RELATED COMPETENCIES

Compe-tency #	*Competency*	*Knowledge Topics*
1.0	Archival principles and practices	Archival theory / functions and AI. e.g., responsibilities of donors when using AI, appraisal, arrangement and description, preservation, access, sensitivity analyses, retention and disposition, archives as data, AI as records, etc.
2.0	Critical data AI/ML, and ethics for archives and records management	Indigenous data sovereignty, data governance, AI/ML data biases, data provenance/paradata, impact of AI on minorities, AI ethics, privacy and copyrights issues of AI, algorithmic transparency, legislation, accountability, etc.
3.0	AI/ML fundamentals and techniques for archives and records management	AI/ML types of models, techniques, and pipelines most-commonly applied to archives and records management, such as: Named Entity Recognition (NER), Natural Language Processing (NLP), topic modeling, sentiment analysis, image and pattern recognition, Large Language Models (LLMs) and GPT.
4.0	Human-computer interaction, user experience and human-information interaction	Basic principles of user experience/interaction with AI/ML apps, tools, and data; interactions with digital records, etc.
5.0	Basic data analysis	Descriptive statistics, data visualization, data operations, data exploration, etc.
6.0	Basic algorithmic thinking	Computational thinking, using algorithms for problem solving and pattern recognition, etc.

TABLE II. SKILL-RELATED COMPETENCIES

Compe-tency #	*Competency*	*Skills involved*
7.0	Management and collaboration skills	Engaging, teaming up and collaborating effectively with subject-matter experts, technical experts and data stakeholders, project management, marketing of AI tools/services in the archives, etc.
8.0	Data-related skills	Collecting and selecting data, digitizing, datafication, critical analyses (sensitivity analyses, data bias analyses), utilitarian analyses (statistical and mathematical analyses), using data as evidence, using data transformation tools, etc.
9.0	Tool-testing and adaption to the workflow skills	Experimenting and tweaking apps and tools, adapting tools to archival workflows, reading manuals, problem solving with online communities and information, etc. Identification of work tasks or processes to be supported by AI/ML, testing, deployment, and evaluation of AI/ML in work tasks and processes, creation and defense of business cases for incorporation of AI/ML into the workflow based on collected evidence, assessment and evaluation of AI/ML to be deployed at the workplace.
10.0	AI/ML skills	Labeling data, feature engineering, training/refining models, selecting and evaluating ML models, etc.
11.0	Basic programming	Scripting, basic coding, tweaking code, using ML libraries, use of coding environments, working with files, troubleshooting, debugging, etc.

TABLE III. ATTITUDE-RELATED COMPETENCIES

Compe-tency #	*Competency*	*Attitudes displayed*
12.0	Self-regulated learning of digital tools	Autonomous, self-regulated and independent learner of digital technologies, detailed-oriented.
13.0	Intellectual curiosity towards AI/ML for archives and records management	Intellectually curious towards AI/ML development for archives and records management. Willing or showing initiative to experiment with AI/ML on archives and records management. Open-minded, early adopter of technology.
14.0	Critical thinking to evaluate new technologies	Reflective practitioner, social theory and value-informed practitioner, praxis-informed social critic of technology, detailed-oriented.
15.0	Disposition to work effectively in teams	Collaborator, coordinator, clear communicator, leader/follower, empathetic, respectful, assertive, active listener, willing to take on workload and deliver, committed to contribute to the success of teamwork, accountable.

V. Comparative Analysis of CAS and InterPARES Frameworks

In this section, we map specific elements (see Appendix A) in the CAS educational framework [17] to the InterPARES educational framework for professional competencies for archival professionals in AI/ML (Arias Hernandez et al [18] in order to validate the results of prior work on CAS educational frameworks and to identify gaps and differences in coverage.

We begin by noting that the two frameworks are differently structured. While the InterPARES framework divides competencies into 'Knowledge', 'Skills', and 'Attitudes', the CAS framework divides them into 'Technical', 'Informational', and 'Social'. Thus, in the CAS framework, specific learning outcomes may combine knowledge, skill and attitudinal competencies, and cover them at a different level of depth or specificity than found in the InterPARES framework. Neither approach is inherently better or worse than the other. Learning outcome-oriented frameworks focus on specific knowledge and skills that students should acquire, which are often assessed through exams and coursework. Thus, they may be more suited to a traditional academic setting. Competency-Based Learning is broader, focusing on what students can do with the knowledge and skills in real-world contexts, often emphasizing practical, real-life applications. A competency-oriented framework may be better for vocational training or professional programs where the ability to apply knowledge and skills in complex situations is key. There remains a discussion to be had, therefore, as to which approach is best suited to archival training in the context of the application of technology to archival work.

Despite the different approaches between the two frameworks, we found a surprising degree of agreement between them. Of the 29 unique elements we extrapolated from the InterPARES framework, there was agreement with CAS competencies for 24 elements (82%), albeit few matches at the same level of specificity or focus, as noted above. Of the 15 CAS competencies, 13 of them were mapped to the InterPARES framework (86%). This considerable and mutual overlap is interesting, given that the frameworks have arisen from two distinct research teams working independently of one another. The high degree of concordance between the two frameworks suggests that archival researchers, if not archival professionals, are beginning to converge in their thinking about what is needed for the training of new archivists in the digital age.

Nevertheless, there were some key differences in coverage between the two frameworks. Looking first at what was included in the InterPARES framework but not in the CAS framework, we note that absent from the CAS framework are specific skills and attitudes associated with computational work practices such as (1) engaging, teaming up and collaborating effectively with subject-matter experts, technical experts and data stakeholders, etc. and (2) autonomous, self-regulated and independent learner of digital technologies, detailed-oriented. We believe that these are important skills and attitudinal competencies and that they highlight a key difference between traditional archival work practices (i.e., 'the lone arranger') and those needed in the digital age (i.e., ability to work as a part of a large multidisciplinary team and engage in continuous, self-directed learning of new technologies). Also missing from the CAS framework is what can be characterized as 'management-oriented competencies' such as project management, marketing, and business case development. Again, we consider these to be important to include in any CAS educational framework given their centrality to working with technology where the need for strong project management skills and the ability to justify the acquisition of new software and computing services is often heightened.

Conversely, included in the CAS framework but missing from the InterPARES framework is reference to the ability for higher level systems thinking and systems-oriented analysis. Also not included in the InterPARES framework is any competency related to cybersecurity (i.e., "Demonstrate an understanding and ability to conduct risk assessments of threats and vulnerabilities to archival cyberinfrastructures and records and develop risk mitigation plans and strategies"). In both cases of missing elements this is possibly due to the InterPARES study's research method and sampling. Included participants worked on applications of AI/ML that were contained and experimental. In other words, their applications were not fully integrated to other systems, nor deployed operationally. Thus, cybersecurity considerations were less likely to show up. In relation to cybersecurity, interviewees provided knowledge competencies on safety, security, custodianship, and control of digital records used in AI/ML applications under the label of "data governance." However, since their AI/ML applications were not fully operational within their organizations, their take on the topic of cybersecurity was limited rather than holistic in coverage.

Additionally, an area given less emphasis in the InterPARES framework that was included in the CAS framework is 'Understand and critically evaluate the application of computational techniques and tools, such as those used in processes of "datafication", in relation to their impact upon the trustworthiness of records, such as their accuracy, reliability and integrity'. This element was included as a knowledge competency subsumed in the InterPARES framework competency 1.0 (Archival principles and practices) under the label "archives as data." Interviewed participants in the original study did note the need to understand digital records as different than data (unless the record itself is a dataset) especially since data constructed during datafication goes through multiple stages of alterations and noted that authenticity and integrity of digital records in AI environments continue to be a central concern for them. However, the current version of the competency framework does not include it explicitly as a skill competency, unlike in the CAS framework. This is a notable absence given that the need for the traditional archival skill of assessing and preserving records' trustworthiness and authenticity is only rising amidst concerns about deep fakes and the potential of AI to 'forge' records of all types.

Though the absence or relative lack of emphasis of certain elements from the InterPARES framework means that their inclusion in the CAS framework was not validated empirically by the InterPARES study's interviews with archival professionals, we would not go so far as to say that this implies that these elements are unnecessary or unimportant in any CAS educational framework, nor that they would not be expected of and by archival professionals. Indeed, given the rise in cyberattacks on all types of GLAM institutions and concerns about deep fakes and the impact of AI on authenticity of records (see, e.g., [20] and [21]), the exclusion of cybersecurity and the lack of specific emphasis on trustworthiness and authenticity of records in a CAS educational framework hardly makes sense. Nevertheless, lack of professional validation argues for further discussion of these ‘missing’ or less explicitly mentioned elements.

VI. Conclusion, Limitations, and Next Steps

In this paper, we have sought to advance the discussion on training for archival professionals in the age of big data and the emergence of new technologies that can be used in archival work. To do so, we have compared two distinct educational frameworks: one (the CAS framework) developed by CAS researchers, which draws upon a prior computational thinking framework and CAS research papers [[17] and another (the InterPARES framework) that arises from empirical research grounded in the views of professional archivists working with AI/ML [18]. Despite differences between the origins, structure, and coverage of the two frameworks, we found a high degree of concordance, suggesting an emerging consensus around the training of new archival professionals in the application of computing to archival work. On the other hand, there were key differences in approach (i.e., learning outcome-oriented vs. competency-oriented) and coverage (i.e., work practices, management competencies, higher level systems thinking, emphasis on the trustworthiness of records, and cybersecurity) that warrant further discussion among archival researchers, educators and professionals prior to formalizing CAS educational frameworks.

Further work is also necessary to continue integrating compatible technology competencies frameworks that are currently emerging for archivists and records managers. We began this work in this paper by comparing and bridging between two compatible educational frameworks proposed by CAS and by an InterPARES study. Different approaches to development of these frameworks are valuable to minimize blind spots that may come from adopting a singular perspective. For example, a solely professional perspective cannot cover for what academic research and scholarship can provide to a competency and educational framework. These perspectives should be complementary, especially for professional programs in higher education.

Acknowledgment

This study is an outcome of InterPARES Trust AI, an international research partnership led by Drs. Luciana Duranti and Muhammad Abdul-Mageed, University of British Columbia, and funded by the Social Sciences and Humanities Research Council of Canada (SSHRC), https:interparestrustai.org.

References

[1] V. Lemieux (2019). Education [Chapter 12]. In L. Duranti and C. Rogers, eds. Trusting Records in the Cloud, pp. 245-265. Facet Publishing.

[2] S. Katuu, (2009). Archives and Records Management Education and Training: what can Africa learn from Europe and North America?, Information Development, 25 (2), 133–45. https://doi.org/10.1177/0266666909104714

[3] S. Katuu (2015). Managing Records in a South African Health Care Institution – A Critical Analysis. PhD Thesis, University of South Africa.

[4] L. Lyon, L., E. Mattern,, A. Acker, and A. Langmead (2015) Applying Translational Principles to Data Science Curriculum Development, iPRES. In *iPRES 2015 Proceedings*.

[5] R. Marciano, V. Lemieux, M. Hedges, M., Esteva, W. Underwood, M. Kurtz, and M. Conrad (2018). Archival Records and Training in the Age of Big Data. In Jaeger, P. T. and Sarin, L. C. (eds), Re-Envisioning the MLS, Bingley, Emerald Group Publishing, http://dcicblog.umd.edu/cas/wp-content/uploads/sites/13/2016/05/submission_final_draft.pdf

[6] M. Hedges, R. Marciano, and E. Goudarouli (2022). Introduction to the Special Issue on Computational Archival Science. J. Comput. Cult. Herit. 15, 1, Article 1, 2 pages. https://doi.org/10.1145/3495004

[7] G. Colavizza, T. Blanke, C. Jeurgens, and J. Noordegraaf (2021). Archives and AI: An Overview of Current Debates and Future Perspectives. J. Comput. Cult. Herit. 15, 1, Article 4 (February 2022), 15 pages. https://doi.org/10.1145/3479010

[8] J. Baron, M.. Sayed, and D.. Oard (2022). Providing More Efficient Access to Government Records: A Use Case Involving Application of Machine Learning to Improve FOIA Review for the Deliberative Process Privilege. J. Comput. Cult. Herit. 15, 1, Article 5 (February 2022), 19 pages. https://doi.org/10.1145/3481045

[9] D. Venkata, P. Young, M. Bell, and A. Green (2021). Alexa, Is This a Historical Record? J. Comput. Cult. Herit. 15, 1, Article 7 (February 2022), 20 pages. https://doi.org/10.1145/3479008

[10] AEOLIAN Network (2024). AEOLIAN Network [Website]. https://www.aeolian-network.net/

[11] L. Jaillant (2024). Introduction to the Special Issue “Applying Innovative Technologies to Digitised and Born-Digital Archives”. J. Comput. Cult. Herit. 16, 4, Article 74e (December 2023), 3 pages. https://doi.org/10.1145/3639829

[12] G. Layne-Worthey and J.S. Downie (2024). Guest editorial: Artificial intelligence for cultural heritage materials. *Journal of Documentation*, *80*(5), 1025-1030. https:// doi.org/10.1108/JD-09-2024-275

[13] L. Jaillant and K. Aske (2024). Are Users of Digital Archives Ready for the AI Era? Obstacles to the Application of Computational Research Methods and New Opportunities. *ACM Journal on Computing and Cultural Heritage*, *16*(4), 1-16.

[14] W. Underwood, D. Weintrop, M. Kurtz, and R. Marciano (2018, December). Introducing computational thinking into archival science education. In *2018 IEEE International Conference on Big Data (Big Data)* (pp. 2761-2765). IEEE.

[15] D. Weintrop, E. Beheshti, M. Horn, K. Orton, K. Jona, L. Trouille, and U. Wilensky (2016). Defining computational thinking for mathematics and science classrooms. *Journal of science education and technology*, *25*, 127-147.

[16] W. Underwood and R. Marciano (2019, December). Computational thinking in archival science research and education. In *2019 IEEE International Conference on Big Data (Big Data)* (pp. 3146-3152). IEEE.

[17] Lemieux, V. L., & Marciano, R. Teaching Computational Archival Science: Context, Pedagogy, and Future Directions. https://ai-collaboratory.net/wp-content/uploads/2024/09/Teaching-Computational-Archival-Science-Context-Pedagogy-and-Future-Directions-2024-submitted.pdf

[18] R. Arias-Hernandez, K. Fewster and S. Penniman (2024). Artificial Intelligence and Machine Learning Competencies for the Archival Professions. Proceedings of the Association for Information Science and Technology, 61(1).

[19] R. Arias-Hernandez and M. Rockembach (2025 forthcoming). Building Trustworthy AI Solutions: Integrating Artificial Intelligence Literacy into Records Management and Archival Systems. AI & Society. Springer Nature.

[20] The British Library (2024, March 8). Learning Lessons from the Cyber-Attack [Report]. https://www.bl.uk/home/british-library-cyber-incident-review-8-march-2024.pdf/

[21] C. Zhu (2024, October 23). Internet Archive hack a signal that cultural institutions are cyber criminals' newest target. CBC Radio. https://www.cbc.ca/radio/asithappens/internet-archive-hack-1.7359959

APPENDIX A: MAPPING OF CAS COMPETENCY FRAMEWORK TO INTERPARES COMPETENCY FRAMEWORK

CT= Competency; K= Knowledge; S=Skill; A=Attidude

	CAS Competency Framework	CT	**InterPARES Competency Framework**	CT
1	Understand and critically evaluate the impact of applying computational techniques and tools in relation to the selection and acquisition of records for preservation (2.3)	K, A	Responsibilities of donors when using AI (1.0)	K
2	Understand and critically evaluate the impact of applying computational techniques and tools in relation to the selection and acquisition of records for preservation (2.3). Apply computational thinking, tools and methodologies to effectively analyze and preserve digital records and support archival work (1.0)	K,S, A	Appraisal, arrangement and description, preservation, access using AI (1.0)	K
3	Apply computational thinking, tools and methodologies to effectively analyze and preserve digital records and support archival work (1.0) Understand and critically evaluate the impact of applying computational techniques and tools to sensitive information and its consequences for privacy (2.4)	K,S, A	Sensitivity analysis using AI (1.0)	K
4	Apply knowledge of technologies to real world record creation and recordkeeping problems and situations (1.1)	S	Retention and disposition using AI (1.0)	K
5	Demonstrate an understanding and skill in "archives-as-data" practices, such a collecting data, creating data, manipulating data, analyzing data and visualizing data (1.2) Understand and critically evaluate the application of computational techniques and tools, such as those used in processes of "datafication", in relation to their impact upon the trustworthiness of records, such as their accuracy, reliability and integrity (2.2)	K	Archives as data (1.0)	K
6	Demonstrate an ability to understand and apply foundational computational archival science theories to gain insights into the nature and evolving forms of digital records and recordkeeping (2.1)	K,S	AI as records (1.0)	K
7	Demonstrate an ability to critically examine the discursive aspects of data as well as the inherent power dynamics and cultural positionality manifest within computational infrastructure and the logics of computing (3.1)	K,A	Indigenous data sovereignty (2.0)	K
8	and skill in "archives-as-data" practices, such a collecting data, creating data, manipulating data, analyzing data and visualizing data (1.2)	K	Data governance (2.0)	K
9	Engage critically and creatively with emerging professional and societal issues, such as bias in training data, arising from the application of computing technology to archival work (3.4)	K,A	AI/ML data biases (2.0)	K
10	Demonstrate an understanding and skill in "archives-as-data" practices, such a collecting data, creating data, manipulating data, analyzing data and visualizing data (1.2)	K, S	Data provenance/paradata (2.0)	K
11	Demonstrate an ability to critically examine the discursive aspects of data as well as the inherent power dynamics and cultural positionality manifest within computational infrastructure and the logics of computing (3.1)	K,A	Impact of AI on minorities (2.0)	K
12	Draw upon knowledge of law and ethics in relation to all functions and activities related to applying computing technology to records and archives (3.3)	K	Privacy and copyrights issues of AI (2.0)	K
13	Engage critically and creatively with emerging professional and societal issues, such as bias in training data, arising from the application of computing technology to archival work (3.4)	K	Algorithmic transparency, legislation, accountability, etc. (2.0)	K
14	Demonstrate an understanding and skill in "archives-as-data" practices, such a collecting data, creating data, manipulating data, analyzing data and visualizing data (1.2) Demonstrate an understanding and skill in how to apply modeling and simulation practices to archival work, such as using modeling to understand a concept, using computational models to find and test solutions, assessing computational models, designing computational models, and constructing computational models. (1.3)	K, S	AI/ML types of models, techniques, and pipelines most-commonly applied to archives and records management, such as: Named Entity Recognition (NER), Natural Language Processing (NLP), topic modeling, sentiment analysis, image and pattern recognition, Large Language Models (LLMs) and GPT (3.0)	K
15	Demonstrate an understanding and skill in how to apply modeling and simulation practices to archival work, such as using modeling to understand a concept, using computational models to find and test solutions, assessing computational models, designing computational models, and constructing computational models (1.3) Within diverse contexts, demonstrate an understanding and skill in computational problem solving, including appreciating when it is	K,S	Basic principles of user experience/interaction with AI/ML apps, tools, and data; interactions with digital records, etc. (4.0)	K

	appropriate to apply particular computational tools and techniques to specific archival tasks, preparing archival problems for computational solutions, programming, choosing effective computational tools, assessing different computational tools, developing modular computational solutions, creating computational abstractions, and troubleshooting and debugging (1.4)			
16	Demonstrate an understanding and skill in "archives-as-data" practices, such a collecting data, creating data, manipulating data, analyzing data and visualizing data (1.2)	K,S	Descriptive statistics, data visualization, data operations, data exploration, etc. (5.0)	K
17	Demonstrate an understanding and skill in how to apply modeling and simulation practices to archival work, such as using modeling to understand a concept, using computational models to find and test solutions, assessing computational models, designing computational models, and constructing computational models (1.3) Within diverse contexts, demonstrate an understanding and skill in computational problem solving, including appreciating when it is appropriate to apply particular computational tools and techniques to specific archival tasks, preparing archival problems for computational solutions, programming, choosing effective computational tools, assessing different computational tools, developing modular computational solutions, creating computational abstractions, and troubleshooting and debugging (1.4)	K,S	Computational thinking, using algorithms for problem solving and pattern recognition, etc. (6.0)	K
18	Not explicitly addressed	NA	Engaging, teaming up and collaborating effectively with subject-matter experts, technical experts and data stakeholders, project management, marketing of AI tools/services in the archives, etc. (7.0)	S
	Demonstrate an understanding and skill in "archives-as-data" practices, such a collecting data, creating data, manipulating data, analyzing data and visualizing data (1.2)	K,S	Collecting and selecting data, digitizing, datafication, critical analyses (sensitivity analyses, data bias analyses), utilitarian analyses (statistical and mathematical analyses), using data as evidence, using data transformation tools, etc. (8.0)	S
19	Within diverse contexts, demonstrate an understanding and skill in computational problem solving, including appreciating when it is appropriate to apply particular computational tools and techniques to specific archival tasks, preparing archival problems for computational solutions, programming, choosing effective computational tools, assessing different computational tools, developing modular computational solutions, creating computational abstractions, and troubleshooting and debugging (1.4)	K,S	Experimenting and tweaking apps and tools, adapting tools to archival workflows (9.0)	S
20	Within diverse contexts, demonstrate an understanding and skill in computational problem solving, including appreciating when it is appropriate to apply particular computational tools and techniques to specific archival tasks, preparing archival problems for computational solutions, programming, choosing effective computational tools, assessing different computational tools, developing modular computational solutions, creating computational abstractions, and troubleshooting and debugging (1.4)	K,S	Reading manuals, problem solving with online communities and information, etc. (9.0)	S
21	Within diverse contexts, demonstrate an understanding and skill in computational problem solving, including appreciating when it is appropriate to apply particular computational tools and techniques to specific archival tasks, preparing archival problems for computational solutions, programming, choosing effective computational tools, assessing different computational tools, developing modular computational solutions, creating computational abstractions, and troubleshooting and debugging (1.4)	K,S	Identification of work tasks or processes to be supported by AI/ML, testing, deployment, and evaluation of AI/ML in work tasks and processes, creation (9.0)	S
22	Not explicitly covered	NA	Defense of business cases for incorporation of AI/ML into the workflow based on collected evidence (9.0)	S
23	Within diverse contexts, demonstrate an understanding and skill in computational problem solving, including appreciating when it is appropriate to apply particular computational tools and techniques to specific archival tasks, preparing archival problems for computational solutions, programming, choosing effective computational tools, assessing different computational tools, developing modular computational solutions, creating computational abstractions, and troubleshooting and debugging (1.4) Assess the social impact and usability of emerging technologies for archival purposes.(3.2)	K,S	Assessment and evaluation of AI/ML to be deployed at the workplace (9.0)	S

24	Demonstrate an understanding and skill in "archives-as-data" practices, such a collecting data, creating data, manipulating data, analyzing data and visualizing data (1.2) Demonstrate an understanding and skill in how to apply modeling and simulation practices to archival work, such as using modeling to understand a concept, using computational models to find and test solutions, assessing computational models, designing computational models, and constructing computational models (1.3)	K,S	Labeling data, feature engineering, training/refining models, selecting and evaluating ML models, etc. (10.0)	S
25	Demonstrate an understanding and skill in computational problem solving, including appreciating when it is appropriate to apply particular computational tools and techniques to specific archival tasks, preparing archival problems for computational solutions, programming, choosing effective computational tools, assessing different computational tools, developing modular computational solutions, creating computational abstractions, and troubleshooting and debugging (1.4)	K,S	Scripting, basic coding, tweaking code, using ML libraries, use of coding environments, working with files, troubleshooting, debugging, etc.(11.0)	S
26	Not explicitly covered	NA	Autonomous, self-regulated and independent learner of digital technologies, detailed-oriented (12.0)	A
27	Not explicitly covered	NA	Willing or showing initiative to experiment with AI/ML on archives and records management. Open-minded, early adopter of technology (13.0)	A
28	Demonstrate an ability to critically examine the discursive aspects of data as well as the inherent power dynamics and cultural positionality manifest within computational infrastructure and the logics of computing (3.1) Engage critically and creatively with emerging professional and societal issues, such as bias in training data, arising from the application of computing technology to archival work (3.4)	K,A	Reflective practitioner, social theory and value-informed practitioner, praxis-informed social critic of technology, detailed-oriented (14.0)	A
29	Not explicitly covered	NA	Collaborator, coordinator, clear communicator, leader/follower, empathetic, respectful, assertive, active listener, willing to take on workload and deliver, committed to contribute to the success of teamwork, accountable (15.0)	A
30	Demonstrate the ability to think in terms of systems and apply systems thinking practices in archival work, including investigating a complex system as a whole, understanding the relationships within a system, thinking in levels, communicating information about a system, and defining systems and managing complexity (1.5)	K,S	Not explicitly covered	N A
31	Understand and critically evaluate the application of computational techniques and tools, such as those used in processes of "datafication", in relation to their impact upon the trustworthiness of records, such as their accuracy, reliability and integrity (2.2)	K	This is covered by knowledge of theory on "archives as data" in 1.0: Archival theory / functions and AI. e.g., responsibilities of donors when using AI, appraisal, arrangement and description, preservation, access, sensitivity analyses, retention and disposition, archives as data, AI as records, etc. (1.0)	K
32	Demonstrate an understanding and ability to conduct risk assessments of threats and vulnerabilities to archival cyberinfrastructures and records and develop risk mitigation plans and strategies (2.5)	K,S	Not explicitly covered	N A