# Computational Archival Science is a Two-Way Street

Bruce Ambacher
*PTAB - Primary Trustworthy Digital Repository Authorisation Body Ltd*
Annandale, VA, USA
bambacher@verizon.net

Mark Conrad
*Advanced Information Collaboratory*
Keyser,WV, USA
conradsireland@gmail.com

***Abstract*— Since its definition in 2018 much of the literature written about CAS has been about archives adopting computational theories, methods and resources. Very little has been written about computational professionals adopting archival theories, methods, or resources. The authors believe that CAS could be substantially enriched if some archival theories, methods, and resources were adopted by computational professionals in developing the systems that create and store vast troves of data. For the purposes of this paper, we will focus on two archival resources – ISO 14721 (OAIS) and ISO 16363 (Trustworthy Digital Repositories). These two resources offer recommendations for long term preservation of digital assets, maintaining understandability of those assets through time, and building trustworthy digital repositories to maintain the provenance and integrity of the repository's collections in such a manner as to provide substantial evidence of the authenticity of the data that it provides to its consumers.**

***Keywords—long-term preservation, OAIS, ISO 16363, information management, trustworthy digital repository, provenance, integrity, understandability, evidence of authenticity, reproducibility, reuse***

## I. Introduction

Computational Archival Science (CAS) is defined 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 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."[1]. Since its definition in 2018 much of the literature written about CAS has been about archives adopting computational theories, methods and resources. Very little has been written about computational professionals adopting archival theories, methods, or resources. See, for example, [6] [7] [8].

The authors believe that CAS could be substantially enriched if some archival theories, methods, and resources were adopted by computational professionals in developing the systems that create and store vast troves of data. For the purposes of this paper, we will focus on two archival resources - ISO 14721 - Open Archival Information System (OAIS) — Reference Model [2] and ISO 16363 - Audit and Certification of Trustworthy Digital Repositories [3][1]. These two resources are referenced throughout archival literature and offer recommendations for long term preservation of digital assets, maintaining understandability of those assets through time, and building trustworthy digital repositories to maintain the provenance and integrity of the repository's collections in such a manner as to provide substantial evidence of the authenticity of the data that it provides to its consumers.

We will begin our paper with a brief overview of ISO 14721. This will be followed by an overview of ISO 16363. We will then discuss some of the benefits of using these resources and how this can enrich CAS.

## II. Overview of OAIS

From the time that organizations began producing digital information there has been an interest in the long term preservation of some of that information. In the mid-1990s the International Organization for Standardization (ISO) asked the Consultative Committee for Space Data Systems (CCSDS) to develop standards for long term storage of digital data. While the CCSDS membership had a great deal of experience in this area, they consulted widely – well beyond the space sciences – and brought in experts from archives, libraries, and other domains. In 1999 a CCSDS Working Group completed a draft of the Reference Model for an Open Archival Information System. It was widely circulated and received many comments and calls for clarifications – especially from the archival community. After incorporating updates, the OAIS Reference Model was published as a CCSDS Recommendation in 2002 and, after further ISO review, as ISO 14721 in 2003. Both CCSDS and ISO require periodic review of their standards. The current version of the ISO Standard is ISO 14721:2012 [4]. A newer version is currently under review with the anticipation that it will be published in the near future.

ISO 14721, the OAIS Reference Model, was the first in a series of related standards focused on long term preservation of digital information – including ISO 16363 discussed below. The CCSDS Working Group chose to begin with the Reference

[1] The free equivalent text of ISO 14721 can be downloaded from https://public.ccsds.org/Pubs/650x0m2.pdf and for ISO 16363 from https://public.ccsds.org/Pubs/652x0m1.pdf .

**978-1-6654-3902-2/21/$31.00 ©2021 IEEE**

Model to establish terms, concepts, and an information model for long term preservation [4]. Because organizations across many domains were developing digital repositories and using domain-specific terms and concepts, the CCSDS Working Group decided that a Reference Model with neutral terminology was needed so that meaningful comparisons could be made between the architectures and designs of existing and future digital repositories. Moreover, the Working Group hoped that as repositories began mapping their terms and concepts to those of the Reference Model it would lead to greater standardization of practices across repositories and this would make it easier for hardware and software vendors to support OAIS Archives.

The Reference Model does not prescribe a particular design or implementation. It includes a set of minimal requirements that an organization must carry out.

The OAIS shall:

– Negotiate for and accept appropriate information from information Producers.

– Obtain sufficient control of the information provided to the level needed to ensure Long Term Preservation.

– Determine, either by itself or in conjunction with other parties, which communities should become the Designated Community and, therefore, should be able to understand the information provided, thereby defining its Knowledge Base.

– Ensure that the information to be preserved is Independently Understandable to the Designated Community. In particular, the Designated Community should be able to understand the information without needing special resources such as the assistance of the experts who produced the information.

– Follow documented policies and procedures which ensure that the information is preserved against all reasonable contingencies, including the demise of the Archive, ensuring that it is never deleted unless allowed as part of an approved strategy. There should be no ad-hoc deletions.

– Make the preserved information available to the Designated Community and enable the information to be disseminated as copies of, or as traceable to, the original submitted Data Objects with evidence supporting its Authenticity.

(ISO 14721:2012 Section 3.1)

Section 3.2 of the Reference Model provides examples of ways to meet each requirement in order to operate an OAIS Archive.

The Reference Model provides a framework for understanding archival concepts needed for long-term digital information preservation and access. It is not possible to discuss all of these concepts in this paper, but it is worth highlighting a few of these concepts. Readers of this paper are encouraged to read the entire ISO 14721 to get a fuller understanding of what is necessary to develop and maintain an OAIS Archive.

"The term 'Archive' has come to be used to refer to a wide variety of storage and preservation functions and systems. Traditional Archives are understood as facilities or organizations which preserve records, originally generated by or for a government organization, institution, or corporation, for access by public or private communities. The Archive accomplishes this task by taking ownership of the records, ensuring that they are understandable to the accessing community, and managing them so as to preserve their information content and Authenticity." (ISO 14721:2012 Section 2) The OAIS Reference Model breaks down these traditional archival activities and concepts into their lowest level components and models them using UML to ensure the long-term preservation and continued access to digital information for a particular community of users of that information.

It is important to begin with definitions of some of the key terms (ISO 14721:2012 Section 1.7.4.):

**Archive:** An organization that intends to preserve information for access and use by a Designated Community.

**Long Term Preservation:** The act of maintaining information, Independently Understandable by a Designated Community, and with evidence supporting its Authenticity, over the Long Term.

**Long Term:** A period of time long enough for there to be concern about the impacts of changing technologies, including support for new media and data formats, and of a changing Designated Community, on the information being held in an OAIS. This period extends into the indefinite future.

**Information**: Any type of knowledge that can be exchanged. In an exchange, it is represented by data.

**Data**: A reinterpretable representation of information in a formalized manner suitable for communication, interpretation, or processing. Examples of data include a sequence of bits, a table of numbers, the characters on a page, the recording of sounds made by a person speaking, or a moon rock specimen.

**Representation Information**: The information that maps a Data Object into more meaningful concepts. An example of Representation Information for a bit sequence which is a FITS file might consist of the FITS standard which defines the format plus a dictionary which defines the meaning in the file of keywords which are not part of the standard.

Another example is JPEG software which is used to render a JPEG file; rendering the JPEG file as bits is not very meaningful to humans but the software, which embodies an understanding of the JPEG standard, maps the bits into pixels which can then be rendered as an image for human viewing.

**Designated Community**: An identified group of potential Consumers who should be able to understand a particular set of information. The Designated Community may be composed of multiple user communities. A Designated Community is defined by the Archive and this definition may change over time.

**Knowledge Base**: A set of information, incorporated by a person or system, that allows that person or system to understand received information.

**Independently Understandable**: A characteristic of information that is sufficiently complete to allow it to be interpreted, understood and used by the Designated Community

without having to resort to special resources not widely available, including named individuals.

**Content Information:** A set of information that is the original target of preservation or that includes part or all of that information. It is an Information Object composed of its Content Data Object and its Representation Information.

**Information Package:** A logical container composed of optional Content Information and optional associated Preservation Description Information. Associated with this Information Package is Packaging Information used to delimit and identify the Content Information and Package Description information used to facilitate searches for the Content Information.

**Submission Information Package (SIP):** An Information Package that is delivered by the Producer to the OAIS for use in the construction or update of one or more AIPs and/or the associated Descriptive Information.

**Archival Information Package (AIP**): An Information Package, consisting of the Content Information and the associated Preservation Description Information (PDI), which is preserved within an OAIS.

**Dissemination Information Package (DIP):** An Information Package, derived from one or more AIPs, and sent by Archives to the Consumer in response to a request to the OAIS.

From these definitions we can observe a few things about an OAIS Archive. The Archive is an organization that intends to maintain information – not just data – for access and use by a Designated Community. To do this the Archive must provide access to enough Representation Information so that the contents of the Archive remain Independently Understandable to the Designated Community based on the Designated Community's Knowledge Base over the Long Term.

There are a number of implications to these statements. They include, the Archive must: have a thorough knowledge of the content it holds – both from a technical and semantic standpoint; identify its Designated Community(ies) carefully; monitor the Knowledge Base of its Designated Community over time to ensure it does not need to increase the Representation Information for its holdings so that they remain Independently Understandable by the Designated Community; monitor the hardware, software, and data it uses for obsolescence and replace it when necessary; maintain meticulous documentation of the chain of custody of all its holdings to provide evidence of the Authenticity of the information it provides to consumers; and ensure that it has the appropriate resources – finances, personnel, etc. – to carry out all of its responsibilities.

The Reference Model provides guidance on how to address all of these issues and much more. It includes an Information Model covering the lifecycle of an Archive's holdings from receipt through dissemination. In the following figures we present some of the central concepts of the Information Model. Fig. 1 shows a high-level model of an Information Package.

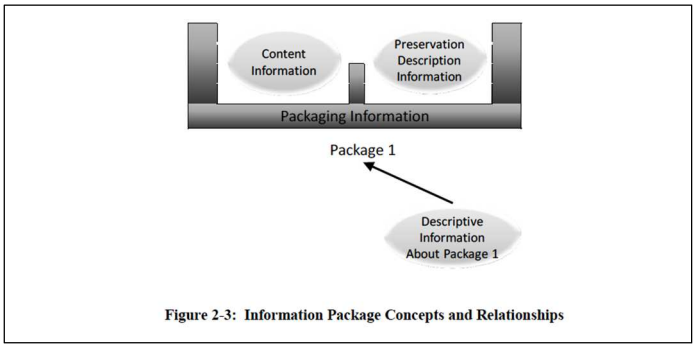

Figure 2-3: Information Package Concepts and Relationships

Fig. 1. OAIS Information Package (ISO 14721:2012 Section 2.2.2)

Fig. 2 illustrates the relationships between the three main types of Information Packages, the OAIS, and the Producer and Consumer.

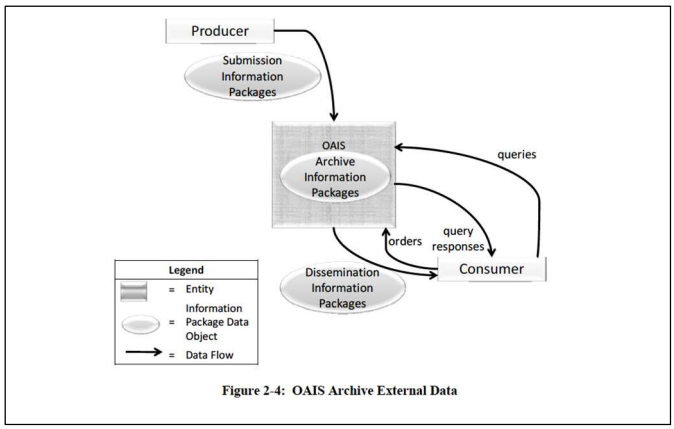

Figure 2-4: OAIS Archive External Data

Fig. 2. Information Package Relationships (ISO 14721:2012 Section 2.3)

Fig. 3 presents a more detailed view of the Archival Information Package (AIP).

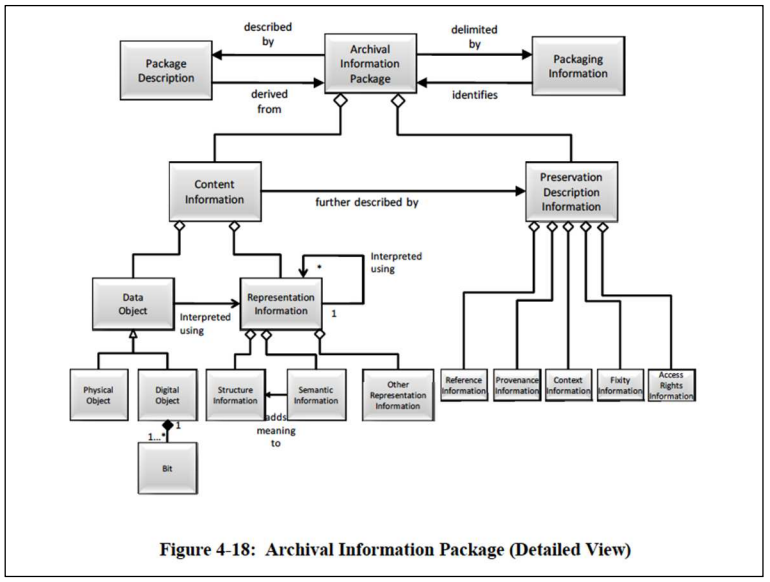

Figure 4-18: Archival Information Package (Detailed View)

Fig. 3. AIP Detailed View (ISO 14721:2012 Section 4.2.2.3)

ISO 14721 also includes a Functional Model that provides a detailed overview of what an Archive may need to do to carry out its mission. Fig. 4 shows a high-level view of the Functional Model.

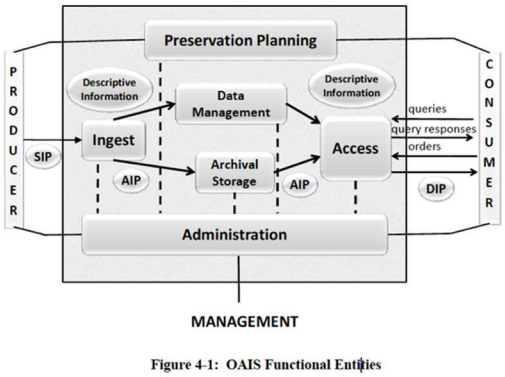


Fig. 4. OAIS Functional Model. (ISO 14721:2012 Section 4.1)

## III. Overview of ISO 16363

One aspect of the OAIS Reference Model that often is overlooked is Section 1.5, (Road Map for Development of Related Standards). This section included an item "standard(s) for accreditation of archives," reflecting the long-standing desire for a standard against which digital repositories may be audited and on which an international accreditation and certification process may be based.

A second notable feature of OAIS was the determination to not provide a system or application specific "solution." Indeed, the developers specifically warned that OAIS was not a "plug and play" solution to their digital preservation concerns.

A decade of international effort utilizing multiple international working group studies and recommendations took place before the certification issues were moved into the CCSDS process. The most significant effort was the National Archives and Records Administration (NARA) Research Libraries Group (RLG) led international study. Participants included national libraries, universities, digital preservation groups, computing centers, government agencies and the Internet Archive. This effort began in 2002 and ended two years later with the issuance of *Trusted Repositories: Audit and Certification* which included a Checklist for the Certification of Trusted Digital Repositories. Following additional refinements based on comments from concerned parties, the task force conducted pilot audits of digital preservation repositories to refine the criteria, gain field experience and identify areas for additional study.

The important next phase began when the CCSDS established a "Birds of a Feather" group within its Mission Operations and Information Management Systems (MOIMS) sector. The new group contained many people who developed OAIS. The task was to further refine, clarify and expand the checklist metrics and guide the resulting standard through CCSDS and ISO approval.

Using their skills and professional experience, this diverse international group of cultural heritage digital preservation people, data experts and CCSDS interested scientists utilized weekly teleconferences to examine the TRAC metrics, consolidate where necessary, and provide clearer meaning to each metric. The group also provided a standard structure for each metric consisting of the statement of the metric presented as a responsibility statement, supporting text that amplifies the metric and indicates why it is necessary as part of a Trustworthy Digital Repository, examples of the ways the repository can demonstrate it is meeting the metric, and an informative discussion of issues regarding the metric and its intent.

ISO 16363, *Audit and Certification of Trustworthy Digital Repositories* underwent review in the CCSDS community and the ISO system of international deliberation and voting. The standard achieved ISO standard status in late 2012 and is undergoing a mandatory five year review in 2021.

In the summer of 2011 the working group developing ISO 16363 conducted test audits at six repositories – three in Europe and three in the United States – to test the metrics in a real environment and thus refine the draft standard. The test sites were promised anonymity, but most of them have since self-identified

An ISO 16363 audit involves a two stage process. The first stage is a review of the candidate repository's written documentation and supporting evidence, including a narrative on the repository, its collections and its users. The candidate repository benefits from its review of its policies, procedures and operations and may see and make improvements prior to the audit. Much of Stage 1 is conducted remotely. The review may elicit additional concerns from the audit team that can be addressed in the Audit phase.

The bulk of the Stage 2 Audit focuses on the metrics. The candidate repository has the opportunity to amplify on its written responses to each metric, to provide a tour of its facilities, to demonstrate its operational procedures, and to allow for staff discussion and demonstration. Much of the Audit is structured to ensure all aspects of the metrics are addressed. Other parts of the onsite audit are unstructured to allow for addressing unforeseen issues and to allow auditors to get a "feel" for the candidate repository and its operations. The Audit Team prepares a detailed report to the candidate repository and to a separate certification team which reviews the recommendation and determines whether the candidate digital repository can be certified as a Trustworthy Digital Repository. If the candidate is certified, it receives a certificate that is valid for three years if annual surveillance audits are conducted and no significant changes occur.

The ISO 16363 metrics are arranged in three major categories. Section Three, Organizational Infrastructure, focuses on administrative issues such as:

- Financial sustainability
- Succession planning
- Technical operations, contracts, licenses
- Preservation procedures
- Designated Community
- Internal review and external audit

These issues include the need for documentation of the organization's mission statement, collection policy, organizational plan, adequacy of staffing, preservation policy, financial sustainability, staff training and succession planning.

Examples of metrics in each of the three major sections are included after the description of each section. One example of an Organizational Infrastructure metric is:

3 ORGANIZATIONAL INFRASTRUCTURE

3.1 GOVERNANCE AND ORGANIZATIONAL VIABILITY

3.1.1 The repository shall have a mission statement that reflects a commitment to the preservation of, long term retention of, management of, and access to digital information.

**Supporting Text:** This is necessary in order to ensure commitment to preservation, retention, management and access at the repository's highest administrative level.

**Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement:** Mission statement or charter of the repository or its parent organization that specifically addresses or implicitly calls for the preservation of information and/or other resources under its purview; a legal, statutory, or government regulatory mandate applicable to the repository that specifically addresses or implicitly requires the preservation, retention, management and access to information and/or other resources under its purview.

**Discussion:** The repository's or its parent organization's mission statement should explicitly address preservation. If preservation is not among the primary purposes of an organization that houses a digital repository then preservation may not be essential to the organization's mission. In some instances, a repository pursues its preservation mission as an outgrowth of the larger goals of an organization in which it is housed, such as a university or a government agency, and its narrower mission may be formalized through policies explicitly adopted and approved by the larger organization. Government agencies and other organizations may have legal mandates that require they preserve materials, in which case these mandates can be substituted for mission statements, as they define the purpose of the organization. Mission statements should be kept up to date and continue to reflect the common goals and practices for preservation.

Section Four, Digital Object Management, focuses on aspects of a digital repository's preservation planning and operations:.

- Ingest process
- Archival Information Package creation
- Preservation planning
- Metadata creation/modification
- AIP integrity monitoring
- Authentic versions in the future
- Technology watch
- Access management

Digital Object Management is the heart of acquisition procedures and validation of the source of the data and the accuracy of the received data. This section includes guidance on internal procedures such as interactions with the data creator(s), assigning unique identifiers to each acquisition, documenting ingest and validation procedures, and ensuring the completeness and accuracy of each data package. Preservation planning and description should be associated with each acquisition to permit discovery and use by the Designated Community. Managing that access is vital.

An example metric from Digital Object Management is:

4 DIGITAL OBJECT MANAGEMENT

4.2 INGEST: CREATION OF THE AIP.

4.2.2 The repository shall have a description of how AIPs are constructed from SIPs.

**Supporting Text:** This is necessary in order to ensure that the AIP(s) adequately represents the information in the SIP(s).

**Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement:** Process description documents; documentation of the SIP-AIP relationship; clear documentation of how AIPs are derived from SIPs.

**Discussion:** In some cases, the AIP and SIP will be almost identical apart from packaging and location, and the repository need only state this. In other cases, complex transformations (e.g., data normalization) may be applied to objects during the ingest process, and a precise description of these actions may be necessary to reflect how the AIP(s) has been adequately transformed from the information in the SIP(s). The AIP construction description should include documentation that gives a detailed description of the ingest process for each SIP to AIP transformation, typically consisting of an overview of general processing being applied to all such transformations, augmented with description of different classes of such processing and, when applicable, with special transformations that were needed.

Some repositories may need to produce these complex descriptions case by case. Under such circumstances case diaries or logs of actions taken to produce each AIP should be created and maintained. In these cases, documentation should be mapped to individual AIPs, and the mapping should be available for examination. Other repositories that can run a more production-line approach may have a description for how each class of incoming objects is transformed to produce the AIP. It must be clear which definition applies to which AIP. If, to take a simple example, two separate processes each produce a TIFF file, it must be clear which process was applied to produce a particular TIFF file.

Section Five, Infrastructure and Security Management brings together metrics on both internal and external security. The major areas of focus are technical infrastructure risk management and security risk management.

- Technologies, Infrastructure & Security
- General system infrastructure requirements
    - Multiple copies

- Offsite storage
- Data integrity measures

- Appropriate technologies
  - Migration, refreshment
- Security – IT, fire, flood, people

The focus is preventing unwarranted access to the repository systems, ensuring only those who are authorized to do so have access to the systems and the data objects. Activities in this section include developing and employing technology watch features to ensure the currency of all systems, technologies and data objects are protected. Preventing or mitigating and recovering from security risks is a vital aspect of overall risk management.

An example metric from Infrastructure and Security Management is:

5 INFRASTRUCTURE AND SECURITY RISK MANAGEMENT

5.1 TECHNICAL INFRASTRUCTURE RISK MANAGEMENT

5.1.1 The repository shall identify and manage the risks to its preservation operations and goals associated with system infrastructure.

5.1.1.3 The repository shall have effective mechanisms to detect bit corruption or loss.

**Supporting Text:** This is necessary in order to ensure that AIPs and metadata are uncorrupted, or any data losses are detected and fall within the tolerances established by repository policy (see 3.3.5).

**Examples of Ways the Repository Can Demonstrate It Is Meeting This Requirement:** Documents that specify bit error detection and correction mechanisms used; risk analysis; error reports; threat analysis; periodic analysis of the integrity of repository holdings.

**Discussion:** The objective is a comprehensive treatment of the sources of data loss and their real-world complexity. Any data or metadata that is (temporarily) lost should be recoverable from backups. Routine systematic failures must not be allowed to accumulate and cause data loss beyond the tolerances established by the repository policies. Mechanisms such as checksums (MD5 signatures) or digital signatures should be recognized for their effectiveness in detecting bit loss and incorporated into the overall approach of the repository for validating integrity

## IV. ENRICHING CAS

Weintrop et. al. [5] identified 4 categories containing 22 computational thinking practices in science and math education with detailed descriptions of each:

**Data Practices**

• Collecting Data

• Creating Data

• Manipulating Data

• Analyzing Data

• Visualizing Data

**Modeling and Simulation Practices**

• Using Computational Models to Understand a Concept

• Using Computational Models to Find and Test Solutions

• Assessing Computational Models

• Designing Computational Models

• Constructing Computational Models

**Computational Problem-Solving Practices**

• Preparing Problems for Computational Solutions

• Computer Programming

• Choosing Effective Computational Tools

• Assessing Different Approaches/Solutions to a Problem

• Developing Modular Computational Solutions

• Creating Computational Abstractions

• Troubleshooting and Debugging

**Systems Thinking Practices**

• Investigating a Complex System as a Whole

• Understanding the Relationships within a System

• Thinking in Levels

• Communicating Information about a System

• Defining Systems and Managing Complexity

Several of the authors of the foundational paper on Computational Archival Science have identified Weintrop's taxonomy as a good source for enriching archival practices [6] [7] [8]. It should be noted that some of the Computational Thinking practices do not align well with traditional archival practices. For example, "Creating Data" and "Manipulating Data" are not something that archivists typically do. Archives create supporting/explanatory information to go with the data they receive from a producer, but they do not manipulate or change the received data in such a way that it would no longer be traceable to the original data. Such a change might call into question the authenticity of the information that the archive disseminated.

Because of the resource intensive nature of gathering or producing big data, researchers and funders of big data projects place a high value on data reuse, provenance, authenticity, documentation, and sustainability. The OAIS Reference Model and the Audit and Certification of Trustworthy Digital Repositories offer guidance on all of these issues. Here we list a few examples.

The concept of the Designated Community(ies) requires that a repository think about who it is collecting data for and what is the Knowledge Base of that community. Big data projects may start with a single purpose in mind with little consideration of

how the data they collect could be reused by others. Archivists have centuries of experience helping researchers find and reuse data for their own purposes. Archivists could provide computational professionals with advice on potential secondary uses of their information and other Designated Communities that might have an interest in a project's data. Armed with this information the repository can ensure that it gathers sufficient Representation Information (i.e., documentation) at the same time it gathers or produces the big data to facilitate its reuse. Gathering substantial Representation Information at the time of big data generation is much less resource intensive than trying to (re-)create it at a later time.

Both ISO standards offer guidance on acquiring and maintaining provenance information for collections. This information combined with other recommended practices for digital preservation and data management – including not changing the digital content as received from the producer - serve as substantial evidence of the authenticity of the information disseminated by the repository. OAIS posits that a repository cannot declare its information authentic. Rather the repository must provide evidence of the authenticity of its holdings so that the users of the holdings can judge the authenticity of the disseminated information. If the authenticity of the disseminated information is called into question, the reputation of the big data project may be damaged.

Preserving and providing access to data – even over the medium term, let alone the long term – is not a "one-and-done" operation. The repository must monitor and respond to such things as bit preservation, guarding against the obsolescence of the hardware, software, and the data itself, physical and cybersecurity, data replication, etc. The repository must also be concerned with such things as changes to the Knowledge Bases of the Designated Communities and the resulting need for additional Representation Information to ensure the data continues to be understandable, the availability of resources to sustain the data, data loss due to transformations of the data, an unbroken and well-documented chain of custody to serve as provenance information, and a successor repository should the existing repository no longer be able preserve and disseminate the data.

Introducing some of these archival practices into computational practices could enhance Computational Archival Science. Incorporating these practices from the beginning of a big data project would produce data that is more sustainable, reusable, and has substantial evidence of its authenticity. This would lead to recognition by the users and the funders of the big data project of the high quality of the data produced and the value received for money spent on the project. Billions of dollars have been spent on projects where the data is no longer usable after a few years. In response, the Big Data Interagency Working Group held a workshop entitled, "Measuring the Impact of Digital Repositories." Subsequently the National Science & Technology Council issued a report from the workshop including the recommendation to, "Expand the use of trustworthy digital repository certification initiatives to ensure that a repository's digital material is authentic, reliable, accessible, and usable on a continuing basis [5]."

## V. Conclusion

In this paper we have given a high-level overview of two ISO standards that are widely acknowledged as seminal works in the archival literature. ISO 14721 (OAIS) and ISO 16363 (Trustworthy Digital Repositories) offer recommendations for long term preservation of digital assets, maintaining understandability of those assets through time, and building trustworthy digital repositories to maintain the provenance and integrity of the repository's collections in such a manner as to provide substantial evidence of the authenticity of the data that it provides to its consumers. These are two examples of recommended archival practices the authors believe could improve computational practices, and in so doing enhance Computational Archival Science.

This paper has discussed some of the impact and benefits that accrue to a certified Trustworthy Digital Repository. Not all repositories that undergo an audit will be certified. Others will become certified with a Plan of Action to address shortcomings over the ensuing time period, usually up to three years. Repository staff should view this negotiated Plan of Action as an opportunity to improve their operations and ensure preservation of useable data over time. The Plan of Action also can be presented to higher management as a blueprint for additional resources to address specific issues. The audit and certification process also enhances the validity and meaning of the more current measures used to justify a program and legitimize operations.

Whether or not a repository chooses to seek certification of their digital repository, these two ISO Standards can lead to improvements in their repository if they are incorporated into the planning and execution of the repository. Moreover, use of these standards should lead to closer collaboration between computational and archival professionals. This in turn should lead to enhanced Computational Archival Science.

## References


[1] Marciano, R., et. al. "Archival records and training in the age of big data," In J. Percell, L. C. Sarin, P. T. Jaeger, J. C. Bertot (Eds.), "Re-Envisioning the MLS: Perspectives on the Future of Library and Information Science Education," (*Advances in Librarianship*, Volume 44B, pp.179-199). Emerald Publishing Limited. (https://ai-collaboratory.net/wp-content/uploads/2020/10/Marciano-et-al-Archival-Records-and-Training-in-the-Age-of-Big-Data-final.pdf)

[2] ISO 14721:2012 Open archival information system (OAIS) — Reference model (https://www.iso.org/standard/57284.html) Also available as Consultative Committee for Space Data Systems, Recommended Practice CCSDS 650.0-M-2 (https://public.ccsds.org/Pubs/650x0m2.pdf).

[3] ISO 16363:2012 Audit and certification of trustworthy digital repositories (https://www.iso.org/standard/56510.html) Also available as Consultative Committee for Space Data Systems, Recommended Practice CCSDS 652.0-M-1 (https://public.ccsds.org/Pubs/652x0m1.pdf)

[4] Garrett, J., et. al. "Certification of digital archives - A brief history and status report," in *Proceedings for the 2015 PV Conference, 3-5 November 2015, Darmstadt, Germany*.

[5] "Measuring the impact of digital repositories: Recommendations," Big Data IWG, Networking and Information Technology Research and Development Subcommittee, Committee on Science & Technology Enterprise, National Science & Technology Council, July 20, 2018. (https://www.nitrd.gov/pubs/BD-IWG-Digital-Repository-Recommendations-2018.pdf)

[6] W. Underwood, D. Weintrop, M. Kurtz, and R. Marciano. *"*Introducing Computational Thinking into archival science education*"*, in *IEEE Conference on Big Data Conference (BigData) 2018, CAS Workshop*, *Seattle, Washington*, pp.2761-2765 (https://ai-collaboratory.net/wp-content/uploads/2020/03/1.Underwood.pdf)

[7] R. Marciano et. al. *"*Reframing digital curation practices through a Computational Thinking Framework*"*, in *IEEE Big Data Conference (BigData), 2019, CAS Workshop, Los Angeles CA*. (**https://ai-collaboratory.net/wp-content/uploads/2020/04/ReframingDC-UsingCT_final-2019.pdf)**

[8] R. Marciano, W. Underwood, M. Hannae, C. Mullane, A. Singh and Z. Tethong. *"*Automating the detection of personally identifiable information (PII) in Japanese-American WWII incarceration camp records**."** in *IEEE Conference on Big Data Conference (BigData) 2018, CAS Workshop*, *Seattle, Washington*, pp.2725-2732 .(https://ai-collaboratory.net/wp-content/uploads/2020/03/2.Marciano.pdf)