Relationship between OAIS, ISO 16363, CTS, FAIR and TRUST

|  |  |  |
| --- | --- | --- |
| **David Giaretta** | **John Garrett** | **Mark Conrad** |
| *PTAB LTD*  *UK*  *david@giaretta.org*  *0000-0001-8414-7509* | *Garrett Software*  *USA*  *garrett@his.com*  *0000-0003-4460-9902* | *NARA*  *USA*  *mark.conrad@nara.gov* |
| **J Steven Hughes**  *Jet Propulsion Laboratory*  *USA*  *john.s.hughes@jpl.nasa.gov*  *0000-0003-4851-293X* | **Terry Longstreth**  *Washington Acad. Sciences*  *USA*  *longstreth@acm.org*  *0000-0001-8228-6185* | **Roberta Svanetti**  *Deda Cloud SRL*  *Italy*  *Roberta.Svanetti@dedagroup.it*  *???????????* |
| **Dawei Lin**  *NIAID*  *USA*  *dawei.lin@nih.gov*  *???????* | **Robert Downs**  *CIESIN*  *USA*  *rdowns@ciesin.columbia.edu*  *???????* |  |

**Abstract – This paper compares the integrated set of OAIS related standards (ISO 14721, ISO 16363 and ISO 16919) to CoreTrustSeal, FAIR and TRUST. The aim is to identify areas where the latter three may be clarified and strengthened in future if they are to be used in any kind of repeatable evaluations.**

**Keywords – certification, trustworthiness**

**Conference Topics – 2.4 Sub-theme 4: Building the Capacity & Capability**

# Introduction

This concept paper is an attempt to clarify the relationship between the concepts of ISO 14721(OAIS)[1] ISO 16363[2], and ISO 16919[6] compared to CoreTrustSeal[3], FAIR principles[4] and TRUST principles[5]..

# Overview

In principle, all of the referenced standards documents reflect a desire to clarify the responsibilities of digital libraries, archives or repositories to provide reliable, understandable, shared access to predominantly computer based data. However, only the OAIS series provides a complete beginning-to-end solution with full coverage of a set of stated requirements for usability and data integrity of a repository’s content. Beginning-to-end coverage is reflected in the encapsulation of a framework and set of related processes for ensuring the safety of a repository’s holdings and the assurance that the repository has been vetted by an independent and certified organization to maintain access and understandability of the holdings for an identified and responsible subset of a potentially unlimited user (or reuser) community for a potentially unlimited time period.

# The OAIS related Standards

OAIS and ISO 16363 focus on the ability to preserve, in the sense defined in OAIS, digitally encoded information. Preservation is defined in terms of usability/understandability and therefore the OAIS Information Model is of central importance. Representation Information is a key part of the Information Model concerned with usability/understandability. To be more specific about how much Representation Information is needed, OAIS requires the repository to define the Designated Communities that it serves. OAIS conformance is defined within the standard itself in terms of the Information Model and the Mandatory Responsibilities.

Some important points to remember are:

* OAIS can apply whether or not what is being preserved is of value;
* OAIS applies whether or not the repository is open access;
* OAIS applies to active repositories even those used in ongoing data science activities;
* a specific taxonomy for “metadata” is defined in order to be able to discuss which specific types of metadata, and how much of each, is needed for preservation;
* OAIS recognises that each piece of “metadata” may itself require preservation, using the same principles.
* OAIS defines preservation in a way that allows one to test whether preservation is being carried out properly.
* OAIS specifies how a Designated Community is defined and what its role is.
* OAIS defines Authenticity as being evaluated based on evidence, rather than something withan absolute yes/no quality.
* OAIS includes data migration guidelines for formats or media obsolescence management in the long-term.

The ISO16363 metrics follow the requirements for OAIS conformance, with additional metrics including ones for legal, financial and security considerations. The metrics of ISO 16363 are provided in depth in order to force auditors to look at specific details, hence the sub-sub-sub-metrics.

Detailed comparisons between CTS, FAIR and TRUST principles may be viewed as a set of tables at <https://docs.google.com/document/d/1uOL1gHCHyNeG9sOea_KxWb28vMxBjBu-mUiofrNTNPc/edit?usp=sharing>.

For the purposes of this paper, we summarise the differences and illustrate the reasons for these.

ISO 16919 establishes a framework for ensuring that an organization intending to provide trustworthiness certification to any digital repository, is itself registered as competent by an international body reporting to an ISO directorate. As an extension of [ISO 17021:Requirements for Certification Bodies](https://isoupdate.com/standards/iso17021/) (which is incorporated by reference), ISO 16919 is tailored to the requirements and constraints of ISO 16363. The detailed ISO 16363 competencies entailed by an organization purporting to ‘guarantee’ the safety and accuracy of its digital holdings are themselves thereby evaluable and certifiable by an independent third-party with an ISO imprimatur.

# General points about CTS, FAIR and TRUST

CTS, FAIR and TRUST principles do not use OAIS terminology, nor do they provide any cross-mapping of terms, however one can identify some overlaps. CTS refers to the OAIS Functional Model and the Functional Entities, and includes “Appraisal and reappraisal over time ensure data remain relevant and understandable to the Designated Community.” FAIR refers to knowledge representation. There are mentions of formats, dictionaries and standard “metadata”. However, there are no mentions of the OAIS Information Model, nor of Representation Information and Representation Information Networks.

Authenticity is mentioned in CTS in R7 “The repository guarantees the integrity and authenticity of the data”. However, the details indicate that this is limited to what OAIS refers to as Fixity. There is no recognition that Authenticity is not a yes/no issue and cannot be guaranteed. The “R” (for Responsibility) in TRUST says that this is “To be responsible for ensuring the authenticity and integrity of data holdings and for the reliability and persistence of its service.” There is little or no consideration of the possibility of changes (Transformations) of the original digital objects.

None refer to the OAIS Mandatory Responsibilities. This, combined with the lack of support of the OAIS Information Model, shows that none of CTS, FAIR or TRUST provide conformance with OAIS, and therefore with ISO 16363.

The following tables extract some cross mapping with OAIS and ISO16363, to indicate where CTS, FAIR and TRUST may be extended to provide more testable details.

# CoreTrustSeal, OAIS and ISO16363

The following table compares the 16 CTS requirements with the OAIS and ISO 16363.

|  |  |
| --- | --- |
| **CTS Requirement** | **Relationship to OAIS/ISO 16363 (Reference is to ISO16363 unless otherwise indicated)** |
| R0. Please provide context for your repository.  – Repository Type.  – Brief Description of Repository  – Brief Description of the Designated Community  – Level of Curation Performed.  – Insource/Outsource Partners. If applicable, please list them.  – Summary of Significant  – Other Relevant Information | OAIS 3.2.1 Negotiates For And Accepts Information  OAIS 3.2.3 Determines Designated Community |
| R2. The repository maintains all applicable licenses covering data access and use and monitors compliance. | 3.4.3 The repository shall have an ongoing commitment to analyze and report on risk, benefit, investment, and expenditure (including assets, licenses, and liabilities).  3.1.2 The repository shall have a Preservation Strategic Plan that defines the approach the repository will take in the long-term support of its mission.  3.5.1 The repository shall have and maintain appropriate contracts or deposit agreements for digital materials that it manages, preserves, and/or to which it provides access.  3.5.2 The repository shall track and manage intellectual property rights and restrictions on use of repository content as required by deposit agreement, contract, or license. |
| R3. The repository has a continuity plan to ensure ongoing access to and preservation of its holdings. | 3.1.2.1 The repository shall have an appropriate, formal succession plan, contingency plans, and/or escrow arrangements in place in case the repository ceases to operate or the governing or funding institution substantially changes its scope.  3.1.2.2 The repository shall monitor its organizational environment to determine when to execute its succession plan, contingency plans, and/or escrow arrangements.  3.4.1 The repository shall have short- and long-term business planning processes in place to sustain the repository over time. |
| R7. The repository guarantees the integrity and authenticity of the data. | 3.3.5 The repository shall define, collect, track, and appropriately provide its information integrity measurements.  4.2.9 The repository shall provide an independent mechanism for verifying the integrity of the repository collection/content.  3 4.1.2 The repository shall clearly specify the information that needs to be associated with specific Content Information at the time of its deposit.  4.1.4 The repository shall have mechanisms to appropriately verify the identity of the Producer of all materials.  4.4.1.2 The repository shall actively monitor the integrity of AIPs. |
| R8. The repository accepts data and metadata based on defined criteria to ensure relevance and understandability for data users. | 4.2.5 The repository shall have access to necessary tools and resources to provide authoritative representation information for all of the digital objects it contains. |
| R10. The repository assumes responsibility for long-term preservation and manages this function in a planned and documented way. | 3.5.1.3 The repository shall have written policies that indicate when it accepts preservation responsibility for contents of each set of submitted data objects.  3.1.1. The repository shall have a statement that reflects a commitment to the preservation of, long term retention of, management of, and access to digital information  3.3.2 The repository shall have Preservation Policies in place to ensure its Preservation Strategic Plan will be met.  3.3.3 The repository shall have a documented history of the changes to its operations, procedures, software, and hardware.  4.2.1.2 The repository shall have a definition of each AIP that is adequate for longterm preservation, enabling the identification and parsing of all the required components within that AIP.  4.2.2 The repository shall have a description of how AIPs are constructed from SIPs.  4.2.6 The repository shall have documented processes for acquiring Preservation Description Information (PDI) for its associated Content Information and acquire PDI in accordance with the documented processes.  4.3.1 The repository shall have documented preservation strategies relevant to its holdings.  5.1.1.1.6 The repository shall have procedures in place to monitor and receive notifications when software changes are needed.  5.2.1 The repository shall maintain a systematic analysis of security risk factors associated with data, systems, personnel, and physical plant. |
| R14. The repository enables reuse of the data over time, ensuring that appropriate metadata are available to support the understanding and use of the data. | OAIS and ISO 16363 with specific reference to the Designated Community  4.2.5 The repository shall have access to necessary tools and resources to provide authoritative representation information for all of the digital objects it contains.  4.2.5.1 The repository shall have tools or methods to identify the file type of all submitted Data Objects.  4.2.5.2 The repository shall have tools or methods to determine what Representation Information is necessary to make each Data Object understandable to the Designated Community.  4.2.5.3 The repository shall have access to the requisite Representation Information.  4.2.5.4 The repository shall have tools or methods to ensure that the requisite Representation Information is persistently associated with the relevant Data Objects. |

# FAIR, OAIS and ISO16363

|  |  |
| --- | --- |
| FAIR Principles | OAIS and ISO16363 |
| To be Findable: |  |
| OAIS and ISO 16363 | OAIS and ISO 16363 |
| F2. data are described with rich metadata (defined by R1 below). | OAIS: Information Model |
| To be Accessible: | OAIS has only a small amount to say about accessibility, except where it is consistent with the Access Rights Information, and with the requirement that a member of the Designated Community can get the information needed to be able to make a judgement about Authenticity. FAIR does not mention either of these concerns. |
| To be Interoperable: |  |
| I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation. | OAIS provides more detailed requirements for Representation Information.  But there are no restrictions in OAIS/ISO 16363. |
| I2. (meta)data use vocabularies that follow FAIR principles. | Interoperability requires more than vocabularies. |
| I3. (meta)data include qualified references to other (meta)data. | This may be the same as Representation Information Network. |
| To be Reusable: |  |
| R1. meta(data) are richly described with a plurality of accurate and relevant attributes. | OAIS provides more detailed requirements for Representation Information and, to a lesser extent for Provenance Information. Fixity Information, Reference Information and Access Rights Information. |
| R1.1. (meta)data are released with a clear and accessible data usage license. | 3.5.2 The repository shall track and manage intellectual property rights and restrictions on use of repository content as required by deposit agreement, contract, or license.  For “metadata” also is implicit in OAIS and ISO 16363 where the “metadata” is itself being preserved. |
| R1.2. (meta)data are associated with detailed provenance. | Data needs Provenance Information.  For “metadata” this is implicit in OAIS and ISO 16363 where the “metadata” is itself being preserved. |
| R1.3. (meta)data meet domain-relevant community standards. | No restriction in OAIS/ISO 16363 on which standards to use. In any case standards will change over time. |

# TRUST Principles, OAIS and ISO16363

|  |  |
| --- | --- |
| TRUST principles | OAIS/ISO 16363 |
| **Transparency** |  |
| To be compliant with this principle, repositories should ensure that, at a minimum, the mission statement and scope of the repository are clearly stated. | 3.1.1 The repository shall have a statement that reflects a commitment to the preservation of, long term retention of, management of, and access to digital information..  3.1.2 The repository shall have a preservation strategic plan that defines the approach the repository will take in the long. -term support of its mission |
| **Responsibility**. TRUSTworthy repositories take responsibility for the stewardship of their data holdings and for serving their user community.  Responsibility is demonstrated by: |  |
| Adhering to the designated community’s metadata and curation standards, along with providing stewardship of the data holdings e.g. technical validation, documentation, quality control, authenticity protection, and long-term persistence. | OAIS does not demand use of Designated Community’s standards. |
| Providing data services e.g. portal and machine interfaces, data download or server-side processing. |  |
| Managing the intellectual property rights of data producers, the protection of sensitive information resources, and the security of the system and its content. | 3.5.2 The repository shall track and manage intellectual property rights and restrictions on use of repository content as required by deposit agreement, contract, or license. |
| **User Focus** |  |
| A TRUSTworthy repository needs to focus on serving its target user community. Each user community likely has differing expectations from their community repositories, depending in part on the community’s maturity regarding data management and sharing. A TRUSTworthy repository is embedded in its target user community’s data practices, and so can respond to evolving community requirements. We take a broad view of ‘user community’ as these could include users depositing or accessing data; those accessing data holdings computationally; and indirect stakeholders such as funders, journal editors, other institutional partners or citizens. | OAIS and ISO16363 define the role of the Designated Community. |
| **Sustainability** |  |
| Ensuring sustainability of a TRUSTworthy repository is necessary to ensure uninterrupted access to its valuable data holdings for current and future user communities. Continued access to data is dependent upon the ability of the repository to provide services over time, and to respond with new or improved services to meet evolving user community requirements. | OAIS and ISO16363 require risk management and succession planning. |
| **Technology** |  |
| A repository depends on the interaction of people, processes, and technologies to support secure, persistent, and reliable services. Its activities and functions are supported by software, hardware, and technical services. Together, these provide the tools to enable the delivery of the TRUST Principles. |  |
| A TRUSTworthy repository may demonstrate the fitness of its technological capabilities by: | ISO16363 section 5 has a large number of metrics covering specific details which must be considered. |
| Implementing relevant and appropriate standards, tools, and technologies for data management and curation. |  |
| Having plans and mechanisms in place to prevent, detect, and respond to cyber or physical security threats. |  |

# REFERENCES

1. See http://public.ccsds.org/publications/archive/650x0m2.pdf
2. See https://public.ccsds.org/Pubs/652x0m1.pdf
3. See http://www.coretrustseal.org
4. See https://www.go-fair.org/fair-principles/
5. See <https://www.nature.com/articles/s41597-020-0486-7>
6. See //https://public.ccsds.org/Pubs/652x1m2.pdf