

Interoperability and standardization aspects in the data preservation domain

Esther Conway¹, Spiros Ventouras¹, Mirko Albani² and Rosemarie Leone²

¹Science and Technology Facilities Council, Rutherford Appleton Laboratory, Didcot, Oxon, UK

²ESA ESA-ESRIN, Via G. Galilei, CP 64, 00044 Frascati, Italy

1. INTRODUCTION

A cooperative and harmonized collective approach on Long Term Data Preservation (LTDP) is ongoing in Europe with the goal to coordinate and optimize European efforts in the LTDP area. The aim is to ultimately preserve the entire European EO space data set for the benefit of all European countries and users, while at the same time reducing the overall costs of data preservation. ESA is coordinating the LTDP initiative[1] in Europe and has an ongoing LTDP programme with the objective to guarantee the preservation of the data from all EO ESA missions and third party missions managed by ESA for the long term, ensuring their accessibility and usability, as part of a joint and cooperative approach in Europe.

In 2010 the LTDP project conducted a survey of over 25 standards accrediting or developing organisations. The Survey resulted in a long candidate list which contained in excess of 600 different standards and procedures that were inspected for relevance to LTDP for the EO community. Many standards and procedures were excluded on grounds that they were not highly relevant or broadly applicable for long term preservation of Earth Observation data sets. The result of this process was a shortlist of approximately 150 standards and procedures presented in the LTDP survey document [2].

The survey constituted a fundamental starting point which allows relevant results from the described standards and procedures to be identified and further analyzed to support the LTDP guidelines. This paper will discuss the complexity of the standards landscape and how it can be used to meaningfully support data preservation and interoperability for the EO community.

2. STANDARDS, PRESERVATION AND INTEROPRABILITY

The International standards organization ISO [3] defines a standard as being a document, established by consensus and approved by a recognized body, that provides, for common and repeated use, rules, guidelines or characteristics for activities or their results, aimed at the achievement of the optimum degree of order in a given context. Where standards should be based on the consolidated results of science, technology and experience, and aimed at the promotion of optimum community benefits.

Standards have the ability to bring such community benefits in the form of digital preservation, access and interoperability. We categorized standards into the following technical areas:

- Authentication
- Digital Document Preservation Standards
- File Formats and Description Languages
- Identifier Standards
- Imaging Standards
- Interoperability Protocols
- Metadata Content Standards and Cataloguing
- Metadata Description Standards
- Metadata and Packaging Standards
- Metadata Structure Standards
- Query languages
- Reference Models and Frameworks
- Earth Observation and Geographic Data Standards
- Thesauri and Word Lists
- XML DTD Schema and Mark Up Languages
- Web Services
- Security
- Ingestion

Standards from these diverse areas are capable of supporting preservation access and interoperability in a number of ways, provision of reference models, audit and certification, reductions of cost, development of requirements, stabilization and mitigation of risk through use of common formats, ensuring common semantic understandings, are amongst some of the modes of support described below.

3. STANDARDS SUPPORTING THE LTDP THEMES

Standards such as OAIS [4] and TRAC [5] audit and certification are broadly applicable to all LTDP themes at a high level. However each of the LTDP themes can be supported through standards, which address the more specific requirements of the Earth Observation community, we will now discuss some typical aspects under the LTDP themes below.

3.1. Preserved Data Set Composition

The OAIS standard requires that the archival information package is composed of representation, packaging and descriptive information. Standards also have a role for data content, utilizing standard formats can increase longevity through breadth of community/technology support. They also reduce preservation costs through shared high quality structural documentation. Employment of standards such as SAFE ensure well structured and described archival information package for an appropriately composed preserved data set.

3.2 Archived Operation and Organization

The OAIS standard mandates many of key operations required of an Archive with complimentary audit/certification standards such as TRAC requiring evidence to support such practices are in place and well documented. While the LTDP working group believes in the standardization of archive practices but they also appreciate the bespoke nature of many archives operational requirements. For example guidelines 2.3 states that equipment maintenance should be done in conformance with manufacturers recommendations. Here a standardized procedure should be put in place, but needs to be done at the individual archive level.

3.3. Archive Security

As above many archives also have individual security issues with bespoke requirements, with the very nature of security risk being notoriously difficult to determine in any absolute sense. However there are standards which support archives in the development of security standards in the case the ISO Information technology security techniques. In addition to standards which allow the development of security requirements there are also standards which facilitate the capture of information to control access such as the OGC GEO rights management standard [6]. This illustrates a case where a standard does not provide a solution but facilitates the creation of one.

3.4 Data Ingestion

Data ingestion standards such as PAIMAS identify and provide a structure for the interactions which take place between an information producer and a deposit archive, and PAIS which formally define the digital information objects to be transferred by an information Producer to an Archive. These standards can be used for all types EO data under this theme however LTDP Guideline 4.2 Metadata generation is the most difficult guideline to support given the sheer number of OGC and other competing standards that are available. Examples of universally applicable standards in the EO domain from ISO 19100 series are

- Geographic Information – Spatial Schema
- Geographic Information – Temporal Schema

The survey document however contains many more valuable standards in this area, which are useful in specific cases. Section 2.13 Earth Observation and Geographic Data contain many more of these more specialized standards which possess a high degree of relevancy in specific domains. This highlights the need to maintain an external more detailed resource which is complimentary to the LTDP guidelines, preventing overload by information which is not broadly applicable.

3.5. Archive Maintenance

The guidelines stipulate that formalized descriptions of archive formats are required. Data description languages such as DRB [7], EAST [8] and DFDL [9] are technologies capable of producing such standardized descriptions.

3.6 Data Access and Interoperability

Data description standards described above are can also support data access and interoperability by providing structural representation information. These descriptions are capable of supporting future transformations into facilitating movement to new standards and interoperability with other existing ones. We should also note that referencing technology based standards which will be superseded, places a burden of frequent review upon any resource which recommends them.

3.7. Data Exploitation and Re-Processing

Guideline 6.6 EO products harmonization again here the LTDP project has identified an area where there is need for a standardization activity. By exploring the individual LTDP themes we can identify area where the community needs to collaborate in order to find common solutions which will naturally bring stability through a large user base.

3.8 Data Appraisal

Another area where a standardization gap has been identified is in area of data appraisal. The guidelines draw on established community procedures but a standardization of this process would be beneficial. Again the prospect of enhancing the long term viability of data through formalizing is something that has been identified through the formulation of the LTDP guidelines.

4. IN CONCLUSION THE CHALLENGE OF MAKING STANDARDS SERVE THE EARTH OBSERVATION COMMUNITY

Given the sheer number of standards the challenge becomes how to deliver information on standards to the Earth Observation community and identify gaps in a meaningful way. The LTDP guidelines need to be related an evolving document which is reactive to the needs of community possessing both generic and more specific data preservation requirements. There also exists the need to distinguishing between standards which are

- One of many ways of supporting a guideline
- A recommended way of universally supporting a guideline
- A recommended way supporting a guideline for some special cases
- The only satisfactory way of enforcing a guideline

Preservation and interoperability cannot be achieved through the use of standard alone. They are however an integral part of the digital preservation landscape for a variety of reasons. It is a key aim of the LTDP initiative to support the EO community in navigating this complex landscape through the judicious selection, appraisal and recommendation of standards. The LTDP initiative is also considering appropriate responses to the identification of gap areas.

6. REFERENCES

- [1] LTDP Common Guidelines issue 1.1 Sept 2010 http://earth.esa.int/gscb/ltdp/EuropeanLTDPCommonGuidelines_Issue1.1.pdf
- [2] LTDP Survey of Standards and Procedures http://ltdpts.eo.esa.int/documentation/-/document_library_display/fB2f/view/12615
- [3] ISO - International Organization for Standardization <http://www.iso.org/iso/home.html>
- [4] Consultative Committee for Space Data Systems (2002). Reference Model for an Open Archival Information System (OAIS). Recommendation for Space Data Systems Standard. CCSDS Blue Book. <http://public.ccsds.org/publications/archive/650x0b1.pdf>
- [5] Consultative Committee for Space Data Systems (2011). Audit And Certification of Trustworthy Repositories, CCSDS Magenta Book <http://public.ccsds.org/publications/archive/652x0m1.pdf>
- [6] Open Geospatial Consortium <http://www.opengeospatial.org/>
- [7] Data Request Broker (DRB) <http://www.gael.fr/drbl/>
- [8] EAST Specification <http://public.ccsds.org/publications/archive/644x0b2s.pdf>
- [9] Data Format Description Language (DFDL) <http://www.ogf.org/dfdl/>