TOWARDS A PRESERVATION CONTENT STANDARD FOR EARTH OBSERVATION DATA

Hampapuram Ramapriyan[1, 2], Dawn Lowe[2], Kevin Murphy[3]

[1]Science Systems and Applications, Inc., Lanham, MD, USA

[2]NASA Goddard Space Flight Center, Greenbelt, MD, USA

[3]NASA Headquarters, Washington, DC, USA

ABSTRACT

Earth observation data from a combination of spaceborne, airborne and in situ sensors have been growing rapidly over the last two decades, and have a much longer history. The observational data as well as digital products derived from them constitute a valuable global asset that must be preserved for the future generations. Currently, there are no international standards specifying the artifacts associated with the data and products that need to be preserved so that they can be understood and reused several decades in the future. However, significant efforts have been made by some national and international groups that can contribute towards such a standard. These efforts and several existing standards from the International Standards Organization (ISO) related to data, metadata and preservation are discussed briefly along with a proposal to develop an ISO standard for specifying preservation contents.

Index Terms— Earth Observation, Preservation, Standards, Data, Metadata, Provenance, Context

1. INTRODUCTION

Information from Earth observing missions (remote sensing with airborne and spaceborne instruments, and in situ measurements such as those from field campaigns) is proliferating in the world. Many agencies across the globe are generating important datasets by collecting measurements from instruments on board aircraft and spacecraft, globally and constantly. The data resulting from such measurements are a valuable resource that needs to be preserved for the benefit of future generations. These observations are the primary record of the Earth's environment and therefore are the key to understanding how conditions in the future will compare to conditions today. Earth science observational data, derived products and models are used to answer key questions of global significance. In the near-term, as long as the missions' data are being used actively for scientific research, it continues to be important to provide easy access to the data and services commensurate with current information technology. For the longer term, when the focus of the research community shifts toward new missions and observations, it is essential to preserve the previous mission data and associated information. This will enable a new user in the future to understand how the data were used for deriving information, knowledge and policy recommendations and to "repeat the experiment" to ascertain the validity and possible limitations of conclusions reached in the past and to provide confidence in long term trends that depended on data from multiple missions.

Organizations that collect, process, and utilize Earth observation data today have a responsibility to ensure that the data and associated content continue to be preserved by them or are gathered and handed off to other organizations for preservation for the benefit of future generations. In order to ensure preservation of complete content necessary for understanding and reusing the data and derived digital products from today's missions, it is necessary to develop a specification of such preservation content. While there are existing standards that address archival and preservation in general, there are no existing international standards or specifications today to address what content should be preserved. The purpose of this paper is to outline briefly the existing standards that apply to preservation (section 2) and describe a recent effort in getting an international standard in place for specifying preservation content for Earth observation data and derived digital data products (section 3). The remaining work needed to arrive at a standard will be described in the concluding section (section 4).

2. APPLICABLE EXISTING STANDARDS

There are several standards developed by various technical committees of the International Standards Organization (ISO) that are relevant to Earth Observation (EO) data and derived digital products. A brief discussion of these is given below, summarized from the respective standard documents cited below. The reader is referred to those cited documents for complete details.

2.1. ISO 14271

Space agencies that are members of the international Consultative Committee for Space Data Systems (CCSDS) have developed recommendations titled the Reference Model for Open Archival Information System (RM-OAIS). The most recent update to the OAIS Reference Model is the publication International Standard Organization's (ISO) 14721:2012 "Space data and information transfer systems – Open archival information system (OAIS) – Reference model", which provides a conceptual framework for archiving [1]. This standard is designated as "CCSDS

recommended practice for an OAIS reference model." It is to be noted that the term "open" here refers to the fact that the reference model and other standards based on it are developed in open forums, and does not imply that the archive access is unrestricted. The OAIS framework addresses all functions associated with long-term preservation of information – ingest, archival storage, data management, access, and dissemination. It is applicable to all archives, but specifically those with responsibility to make information available in the long term. It opens the door for the development of more detailed standards in this area. It identifies potential areas (a "road map") for such standards and lists several that have been developed. It does not, however, specify a design or implementation, leaving the option for breaking out the specified functionality at the implementer's discretion.

2.2. ISO 16363

As indicated in the road map in ISO 14271, a standard has been developed by CCSDS and ISO that specifies requirements for certification of trustworthy digital repositories, based on the OAIS Reference Model. This standard, ISO 16363, is also designated by CCSDS as a recommended practice [2]. It is based on the publication by the joint task force of the Research Library Group (RLG) and the National Archives and records Administration (NARA) constituted to address digital repository certification [3]. The ISO 16363 helps auditors of digital repositories with objective criteria and metrics to assess and certify them. It also helps institutions hosting repositories to identify weaknesses and make improvements through self-assessments.

2.3. ISO 16919

ISO 16919 describes how to audit archives for compliance with the requirements [4]. Also a CCSDS recommended practice, this standard provides requirements to be met by bodies that audit and certify candidate trustworthy digital repositories. It is also helpful for digital repository staff to understand the certification process. In the context of this standard, trustworthiness of repositories means "that they can be trusted to maintain, over the long-term, the understandability and usability of digitally encoded information placed into their safekeeping." The standard describes how a third party can inspire confidence that it has been performed the certification with: impartiality, competence, responsibility, openness, confidentiality, and responsiveness to complaints.

2.4. ISO 19115-1 and 19115-2

ISO 19115 is standard with a general title "Geographic Information – Metadata". It has three parts. ISO 19115-1 covers fundamentals, 19115-2 provides extensions for imagery and gridded data and 19115-3 consists of technical specifications for an XML schema implementation of metadata fundamentals. The primary purpose of ISO 19115-1 is to describe digital information that has a geographic extent [5]. However, it can be used to describe information resources that do not have a geographic extent – e.g., documents, software and repositories. Implementation of this metadata standard by resource providers will help effective and complete characterization of resources; facilitate management; enable understanding and appropriate use; help discovery, access and reuse; and assist users in deciding whether a given resource is useful to them.

ISO 19115-2 augments ISO 19115-1 with additional structure to describe more extensively the derivation of geographic imagery and gridded data [6]. It provides the structure needed to represent properties of instrument acquiring data, instrument geometry, production processes, etc. It covers metadata needed for describing derivation of "geographic information from raw data, including the properties of the measuring system, and the numerical methods and computational procedures used in the derivation."

2.5. ISO 19157

ISO 19157:2013 standardizes components and structures of data quality measures, which are important to compare different datasets to determine which of them best fit a user's needs [7]. It provides principles for describing data quality and concepts for handling information on data quality, and a consistent manner in which such information is determined and reported. It also provides guidelines for evaluation of quantitative quality information for geographic data.

2.6. ISO 19165

ISO 19165, a standard for "Geographic Information - Preservation of digital data and metadata" was developed during 2014-2016, and considers geographic information preservation in general [8]. It enumerates several distinguishing characteristics of geospatial data (e.g., relation to a well-defined section of the Earth, exchange by using theme-specific and sophisticated exchange formats, large data volumes, and existence of several levels-of-detail of the same dataset), and postulates that such data should be "preserved together with relevant metadata content that fully addresses these structural characteristics." It also indicates that while its focus is on geospatial data, the principles brought forth by this standard can be applied to other types of data as well.

It defines the requirements for the long-term preservation of digital geospatial data, including "metadata, representation information, provenance, context and any other content items that capture the knowledge that are necessary to fully understand and reuse the archived data."

3. TOWARDS A CONTENT STANDARD

While all of the standards discussed above provide a good basis for developing a content standard for preserving Earth observation data and derived digital products, currently there is no standard that meets this need.

The standard ISO 19165 states: "In preserving data, future users need to understand what they are working with (context information) and how the data were created (provenance information). Because most Earth science data involve complex physics and mathematics, the metadata shall include sufficient documentation (or pointers thereto) that provide the derivation of the algorithms used to generate the dataset. Likewise, the metadata shall include pointers to calibration data and ancillary data that were needed to produce the dataset. The specific content items needed to preserve the full provenance and context of the data and associated metadata depend on the needs of the designated community and types of datasets (e.g., maps, remotely sensed data from satellites and airborne instruments, and physical samples). Follow-up parts to this standard may be developed detailing content items appropriate to individual disciplines."

A New Work Item Proposal (NWIP) to develop a standard titled "Geographic information -- Preservation of digital data and metadata -- Part 2: Content specifications for Earth observation data and derived digital products" has recently been approved by the ISO Technical Committee for Geographic Information/Geomatics (TC 211). This new standard, ISO 19165-2, will provide detailed specifications of the preservation content for Earth observation data and derived digital products as an extension to ISO 19165. There has been significant amount of work outside the ISO environment that can be used as a basis for the proposed standard. This work is described briefly below.

3.1. ESA

The European Space Agency (ESA) formed a Long Term Data Preservation (LTDP) Working Group in 2007 for defining and promoting a coordinated approach to reserve and curate European Earth observation space data assets. One of the outputs of this working group was the "Earth Observation Preserved Data Set Content" (EO PDSC), a document providing guidance to data holders on preservation. There have been several versions of this document, the latest having been published in 2012 [9].

3.2. Earth Science Information Partners (ESIP)

Earth Science Information Partners (ESIP) is a U.S. based group of organizations with international membership. It has been active since 1998 and currently has over 180 organizational members including government agencies, universities and commercial as well as nonprofit entities [10]. In 2011, The Data Stewardship Committee within the ESIP developed an "emerging" Provenance and Context Content Standard (PCCS), which is essentially an enumeration of data and related items that need to be preserved from missions, projects or investigations that support long-term global change research [11]. This list benefitted from the categories of content called for preservation by a report of the US Global Change Research Program's workshop held in 1998 [12] as well as NASA and NOAA experiences with many Earth observing satellite missions. The purpose of the list is to provide a starting point for developing a standard for content. Thus, the focus is on "what" needs to be preserved, rather than "how".

3.3. NASA

Based on the PCCS mentioned above, NASA developed its Earth Science Preservation Content Specification (PCS) [13] in late 2011 and has been using it as a requirement for its new missions. For missions that had been in progress or completed before the PCS was developed, it is used as a check list in order to capture and preserve as much of the relevant content items as possible on a best efforts basis. The content items are grouped into the following eight categories: Preflight/Pre-Operations Calibration, Science Data Products, Science Data Product Documentation, Mission Data Calibration, Science Data Product Software, Science Data Product Algorithm Input, Science Data Product Validation, and Science Data Software Tools.

3.4. CEOS WGISS

The Data Stewardship Interest Group within the Working Group on Information Systems and Services (WGISS) of the Committee on Earth Observation Satellites (CEOS) Earth Observation Preserved Data Set Content has adopted the EO PDSC [9] developed by ESA and has evolved it into a more global reference for data preservation [14]. As is the case with references [9], [12] and [13], this document also focuses on "what" needs to be preserved, but also provides guidance on "when" the various content items should be preserved in the course of a mission's lifecycle.

4. CONCLUSION

Earth observation data and derived digital products are valuable global assets that need to be preserved for the benefit of future generations. To ensure that they remain understandable and reusable decades into the future, it is essential to identify all the associated artifacts that should be preserved along with the data. Frameworks for archival systems, metadata and data quality standards, as well as recent work in national and international agencies provide a good basis for developing a standard that specifies the content to be preserved. Such a standard will provide uniform guidelines for all organizations around the world involved in Earth observations. Planning for preservation at the beginning of a project or mission will help ensure that the preservation content items are captured at the appropriate

times and are not lost as the individuals or groups familiar with those items move on to other activities. This is especially important where several large organizations need to interact for successful implementation of missions, and different teams are responsible for different content items that need to be eventually preserved. Clearly, it is not sufficient that the content items be preserved, but they should also be easily accessible to future users. Establishing persistent identifiers to preserved artifacts and using linked data concepts to connect related items would benefit such access greatly.

5. ACKNOWLEDGMENTS

Ramapriyan's contribution to this paper was supported by NASA's contract with Science Systems and Applications, Inc. Lowe and Murphy contributed to the paper as a part of their duties as employees of NASA.

REFERENCES

[1] ISO 14721:2012 *Space data and information transfer systems – Open archival information system (OAIS) – Reference model*, https://www.iso.org/standard/57284.html.

[2] ISO 16363:2012 *Space data and information transfer systems -- Audit and certification of trustworthy digital repositories*, https://www.iso.org/standard/56510.html.

[3] RLG-NARA Task Force 2007 *Trustworthy Repositories Audit & Certification: Criteria and Checklist*, Version 1.0. Chicago: CRL, February 2007, http://www.crl.edu/sites/default/files/d6/attachments/pages/trac_0.pdf.

[4] ISO 16919:2014 *Space data and information transfer systems -- Requirements for bodies providing audit and certification of candidate trustworthy digital repositories*, https://www.iso.org/standard/57950.html.

[5] ISO 19115-1:2014 *Geographic information -- Metadata -- Part 1: Fundamentals*, https://www.iso.org/standard/53798.html.

[6] ISO 19115-2:2009 *Geographic information -- Metadata -- Part 2: Extensions for imagery and gridded data*, https://www.iso.org/standard/39229.html.

[7] ISO 19157:2013 *Geographic information -- Data quality*, https://www.iso.org/standard/32575.html.

[8] ISO 19165 *Geographic information -- Preservation of digital data and metadata*, https://www.iso.org/standard/67325.html.

[9] ESA:2012 *EO Preserved Data Set Content v4.0*, LTDP-GSEG EOPG-RD-11-0003, July 2012.

[10] ESIP 2017 http://esipfed.org/ [Last accessed Sept. 7, 2017].

[11] ESIP:2011 *ESIP Federation Provenance and Context Content Standard* http://wiki.esipfed.org/index.php/Provenance_and_Context_Content_Standard.

[12] USGCRP, 1999, Global Change Science Requirements for Long-Term Archiving, Report of the Workshop, October 28-30, 1998, National Center for Atmospheric Research, Boulder, CO. Sponsored by NASA and NOAA, through the USGCRP Program Office, DOI: http://dx.doi.org/10.7930/J0CZ353N.

[13] NASA:2011 *Earth Science Preservation Content Specification* https://earthdata.nasa.gov/files/423-SPEC-001_NASA%20ESD_Preservation_Spec_OriginalCh01_0.pdf.

[14] CEOS:2015 *Earth Observation Preserved Data Set Content (PDSC)*, http://ceos.org/document_management/Working_Groups/WGISS/Interest_Groups/Data_Stewardship/Recommendations/EO%20Preserved%20Data%20Set%20Content_v1.0.pdf.