Towards Unifying NASA Earth Science Enterprise-Wide Metadata Around International Standards: Study Results and Recommendations

S.J.S. Khalsa[a,*], S.F. Browdy[b], B.H. Wiess[c]

[a]CIRES, University of Colorado, Boulder, CO 80309 USA – sjsk@nsidc.org
[b]OMS Tech, Orlando, Florida, USA – steveb@omstech.com
[c]Jet Propulsion Laboratory, Pasadena, CA – Barry.h.weiss@jpl.nasa.gov

Abstract – NASA's Earth Science Data and Information System (ESDIS) Project assembled a team of data systems and metadata experts to analyze requirements and recommend best approaches for adopting the ISO 19100 series of standards for describing the data and services offered by NASA Earth Science data systems. NASA has recognized the advantages of adopting the ISO 19115 metadata standard for future Earth-observing missions. The Earth Science data centers have also recognized these advantages. The team considered the applicability, limitations, and possible profiles of this standard for the diverse data sets maintained by NASA data centers and missions. The team found compelling reasons for adopting the ISO 19100 standard to help NASA reach its interoperability goals. Achieving full integration will take time. Planning the integration will require resolution of tension between the near-term needs of new missions and heritage systems with large amounts of records using other data models.[**]**

Keywords: metadata, data center, Earth science missions, ISO 19115, interoperability

1. INTRODUCTION

NASA Earth Science Data systems include data production facilities for satellite, airborne and in situ observing systems, funded research projects that generate data products, data archive and distribution systems (data centers), and metadata clearinghouses. Metadata is a common enabling element to all these systems, employed in many different ways and represented by a variety of conceptual models and encoding standards. Use of a single common metadata standard would enhance interoperability and reduce systems engineering and data management costs. The international metadata standard ISO 19115 is gaining increasing acceptance worldwide (ANZLIC 2007; JRC 2010) and is being considered for NASA's upcoming Decadal Survey missions. At the same time, an immediate and complete adoption of a single metadata standard across all NASA data systems would be extremely costly. The likely benefit of a immediate adoption would not justify the investment.

NASA's Earth Science Data and Information System (ESDIS) Project established a Tiger Team to assess the metadata needs and current practices of current and planned systems to determine whether ISO 19115 and related 19100 standards would enhance stakeholders' current data models and provide benefits to NASA Earth science as a whole. This team, christened MENDS – Metadata Evolution for NASA Data Systems – sought to determine the optimal path for integration of current stakeholders' data systems with a common metadata standard.

2. METHODS

2.1 Stakeholder Characterization

The three main stakeholder groups in NASA's Earth Science Enterprise are:

- Science missions,
- Data centers,
- Metadata clearinghouses.

Science missions are the primary data producers. Data products originate from science data systems which process data streams coming from satellite-, airborne- or in situ instruments. Complex algorithms and numerical models employ these data as well ancillary data from other instruments and experiments to retrieve various geophysical measurements. Data centers archive and distribute science data products and also produce value-added products. NASA's two metadata clearinghouses, the Global Change Master Directory (GCMD) and ECHO provide discovery services.

Each stakeholder group had representation on the MENDS team and contributed to the understanding of the requirements, aspirations, constraints, and issues surrounding the usage of ISO 19115. All helped to arrive at a set of recommendations for ESDIS regarding the adoption of ISO 19115. The methods used to collect input included review of submitted documentation, surveys, development of use cases, dialog during team conference calls, and mapping of existing data models to ISO 19115.

2.2 Metadata Standards

The International Organization for Standardization (ISO) Technical Committee for Geographic Information (TC211) first published 19115:2003 *Geographic information – Metadata*, and subsequently published 19115-2:2009 – *Extensions for imagery and gridded data*. ISO 19119:2005 *Services* provides additional classes describing service metadata. The implementation of these abstract standards is specified in ISO 19139:2007 *Metadata – XML schema implementation*. A recently published standard from TC211 that is relevant to many NASA data products is ISO 19130:2010 – *Imagery sensor models for geopositioning*. In fact, none of the ISO/TC211 standards or technical specifications are intended to be used alone. Rather, as a family they aim to facilitate the availability, access, integration, understanding and sharing of geographic information.

The MENDS team also considered other metadata standards currently in use. The team concluded that none of them had the capabilities for the full range of intended applications within NASA data systems.

* Corresponding author
** This work was funded by the NASA ESDIS Project

2.3 Metadata Usage

A metadata flow diagram was created (Fig. 1) to better visualize how metadata is currently exchanged and used among stakeholders' systems, and to help identify the points where issues related to the adoption of ISO 19115 might arise. Data products can originate in a NASA Science Data system, or in a Science Investigator-Led Processing System (SIPS). SIPS can generate products from NASA-funded research efforts and can also be part of the processing stream that generates higher-level geophysical products from lower-level products obtained from a Mission Science Data system.

The three most common metadata standards currently in use by NASA Earth science data systems are the Earth Science Data Model (ESDM), developed for NASA's EOSDIS Core System, the ECHO data model, developed for the EOS Clearing HOuse and the Directory Interchange Format (DIF) used by the GCMD for catalog-level metadata. Many of NASA's data centers maintain metadata catalogs using custom data models. The encoding formats used for exchange of metadata between systems include ASCII, XML, ODL and DIF, which is both a format and content standard.

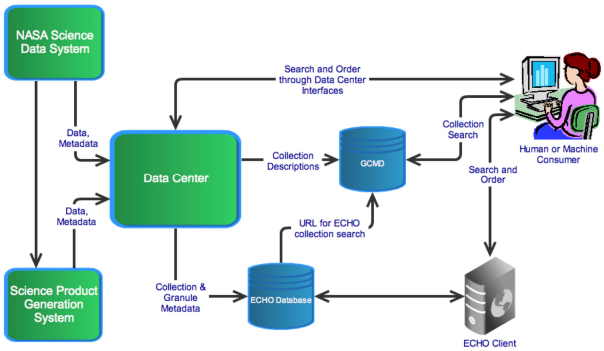

Figure 1. High-level schematic depicting the principal categories of metadata repositories and channels of metadata sharing in NSAA Earth science data systems.

3. ISSUES

Through its investigations, the MENDS team discovered a number of issues that had bearing on the adoption of ISO 19115 and influenced the recommendations to ESDIS. The most prominent of these are summarized in this subsection.

3.1 Resources

The two greatest concerns and the biggest impediments to the adoption of ISO 19115 by NASA data systems are cost and the availability of relevant expertise among the stakeholders. To maximize the proper and effective use of ISO metadata, those who implement the change must gain a thorough understanding of the standard. The complexity of the specifications and the volume and density of documentation require a substantial time investment.

3.2 Heritage Systems

NASA Earth science data systems currently house a vast amount of data. Replacing the metadata that describes these data sets, individual data files and the elements within data files would be a huge undertaking. Conversion to the ISO format on demand is the solution being considered by ECHO, and some data centers already offer this service for catalog-level metadata.

3.3 New Missions

As new NASA missions begin designing their data production systems, they must first select a metadata standard The Soil Moisture Active-Passive (SMAP) mission planned is scheduled for a launch in late 2014, which means that the SMAP mission must choose a metadata standard very soon. This deadline imposes some urgency upon NASA to develop processes and create guidelines for the adoption of a metadata standard.

3.4 Profiles and Extensions

The tremendous flexibility of the ISO metadata standard means that there may be multiple ways to represent the same information. ISO 19115 defines a process for creating profiles and extensions of the standard to tailor its application to specific domains. The MENDS team examined the North American Profile of ISO 19115 (INCITS 2009) and found that many elements critical to complete documentation of NASA data were missing. If NASA were to develop its own specific guidelines for the application of ISO 19115 to NASA datasets it could help prevent inconsistencies and incompatibilities that might arise as stakeholder groups develop their implementations independently.

3.5 Metadata Hierarchy

Metadata serves a multitude of purposes. It is used to describe the characteristics of the instrument that produced the data, the structure, encoding and spatio-temporal extent of the data product and the services that a data center might offer for product ordering, subsetting, merging and visualization. Metadata are also important for conveying to users the lineage and the quality of a data product. The applicability of these metadata elements vary based on their scope, which can be characterized as follows:

- *Collection metadata* describe an entire set of data products or files. Values of collection metadata apply to all of the products in the specified collection.
- *Granule metadata* describe a single instance (granule) within a data collection.
- *Parameter metadata* describe a specific data element or layer within a granule. Examples are metadata describing units of measure, and the names and dimensions of data arrays.

Collection and granule metadata are typically used in search and discovery of data products. Granule and parameter metadata are usually necessary for data interpretation. ISO 19115 can be used for all three levels of metadata. For consistency, it would make sense to adopt ISO for all metadata that appear in NASA Earth science data. However, popular Earth science data encoding formats like netCDF and HDF already have conventions for Parameter metadata.

4. RECOMMENDATIONS

The MENDS team found compelling reasons for NASA data systems to evolve towards a consistent application of ISO 19115 and related standards. Indeed, the adoption of ISO 19115 will enable a more cost-effective and thorough means to achieve integration across various data systems that generate, deliver and archive NASA's Earth Science data products. But the team also recognized that this will involve tiers of adoption and varying timelines for implementation by different stakeholder groups. The key recommendations for achieving successful adoption of ISO 19115 are summarized in the following subsections.

4.1 Early Adopters

Some NASA stakeholders have already begun working on ISO 19115 implementations so there is an urgent near-term need to define NASA conventions. The experiences of these early adopters will help in the formulation of conventions that will become the basis of any future extension or profile of this standard. The process of defining an extension or profile must be measured and systematic and should involve the broadest possible spectrum of producers and consumers of NASA data products. As a result, the process will evolve. The designers and early adopters of NASA ISO 19115 conventions will not benefit from uniformity across all stakeholder groups. These designers and early adopters will no doubt make some decisions that will later have to be reversed or amended.

4.2 Managing the Evolution of NASA Conventions

As missions and projects come on board, NASA designers and developers will encounter issues, and will recommend additions or modifications to the NASA ISO 19115 conventions. ESDIS will need to retain a group of individuals who review these recommendations and make appropriate decisions. ESDIS should provide these individuals with specific criteria for making decisions that build upon the NASA ISO 19115 conventions. To retain maximum conformance with the ISO 19115 standard, any additions or modifications that the NASA teams recommend and ESDIS adopts, should be forwarded through representatives to the ISO TC211 for potential adoption into the international standard.

4.3 Baseline Schema

The initial NASA ISO 19115 conventions need a comprehensive starting point for the early adopters of the NASA Decadal Survey Missions. A baseline schema would serve this purpose by identifying the "core" collection and granule metadata fields that will be utilized to translate metadata between ISO 19115 and other data models. The baseline schema should specify one set of classes and attributes that would be required of all missions, and a second set of classes and attributes that would be optional depending on the characteristics of each individual mission.

4.4 Automatic Metadata Synchronization

The primary advantage of adopting ISO 19115 and the NASA ISO 19115 conventions is the promotion of interoperability among applications that are able to communicate using the same standard. Collection-level metadata within EOSDIS typically exist in three places, represented by three different data models. Once a common metadata standard based on ISO 19115 is adopted and implemented amongst these entities, a user should be able to get the same metadata record from any of the parties, eliminating the current uncertainty experienced by users receiving varying results when using these systems.

4.5 Community Engagement

Involvement of the stakeholder community, both within and outside of NASA, will be essential to ensure that the adoption of ISO 19115 across the NASA Earth Science Enterprise is successful. Harmonization of NASA's efforts with those of the greater Earth Science community will reduce the effort required of NASA stakeholders and maximize the interoperability of data products across all providers, domestic and international. The experience of implementing the NASA ISO 19115 conventions and the evolution of its use will also be well served by this harmonization. To evolve the NASA ISO 19115 conventions, possibly towards an extension or profile of ISO 19115, NASA will need to study the ISO metadata implementations of other agencies and organizations to learn from their experiences.

4.6 User Support

The adoption of ISO 19115 by NASA Earth science data systems will require easily accessible documentation as well as tools to help translate, create and validate metadata that complies with the NASA conventions. NASA must ensure that the engineers and technicians responsible for the design and implementation of systems supporting ISO 19115 have the necessary training and skills. The development of generic, modular tools will significantly reduce the effort required to generate and to access ISO 19115 metadata. The required documentation and tools should be maintained and made accessible at a website dedicated to community sharing.

5. CONCLUSIONS

A study of the metadata needs and practices among the heterogeneous data systems supporting NASA Earth science argues for the adoption of the ISO 19115 metadata standard with the development of specific guidelines for implementation. These guidelines will evolve as various disciplines and groups gain experience with the standard and may eventually lead to a formal profile or extension. Each stakeholder must develop their own implementation plan which takes into account their specific requirements, schedule and available resources. Increased data interoperability, reusable authoring and validation tools, reduced duplication and easier synchronization of metadata holdings will be some of the benefits realized by the producers, stewards and consumers of NASA's data products.

6. REFERENCES

European Commission Joint Research Centre, 2010, INSPIRE Metadata Implementing Rules: Technical Guidelines based on EN ISO 19115 and EN ISO 19119 (Version 1.2).

INCITS 453-2009, Information Technology - North American Profile of ISO 19115:2003 - Geographic Information - Metadata, 2009.

ANZLIC the Spatial Information Council, 2007, ANZLIC Metadata Profile: An Australian/New Zealand Profile of AS/NZS ISO 19115:2005, Geographic information — Metadata (implemented using ISO/TS 19139:2007, Geographic information — Metadata — XML schema implementation), ISBN: 978-0-646-46940-9

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the contributions that MENDS team members made to the research and analysis described. We are especially indebted to Ed Armstrong, Helen Conover and Chris Lynnes who provided valuable data center perspectives, Ted Habermann who brought in-depth NOAA experience with ISO 19115, and Andrew Mitchell who organized and provided direction for the team.