GEOSPATIAL STANDARDS AND THE KNOWLEDGE GENERATION LIFECYCLE

Siri Jodha S. Khalsa[1] *and Rahul Ramachandran*[2]

[1]University of Colorado, NSIDC
[2]NASA/MSFC

1. INTRODUCTION

Standards play an essential role at each stage in the sequence of processes by which knowledge is generated from geoscience observations, simulations and analysis. This paper provides an introduction to the field of informatics and the knowledge generation lifecycle in the context of the geosciences. In addition we discuss how the newly formed Earth Science Informatics Technical Committee is helping to advance the application of standards and best practices to make data and data systems more usable and interoperable.

2. THE GRSS EARTH SCIENCE INFORMATICS TECHNICAL COMMITTEE

Informatics is the science and technology of applying computers and computational methods to the systematic analysis, management, interchange, and representation of data, information, and knowledge. The mission of the IEEE GRSS is "to advance science and technology in geoscience, remote sensing and related fields," and the society's fields of interest include the "techniques of science and engineering as they apply to the remote sensing of the Earth", as well as the "processing, interpretation and dissemination" of information derived from remote sensing and other techniques. For these reasons the GRSS Earth Science Informatics Technical Committee (ESI TC) was formed to advance the application of informatics to the geosciences and remote sensing. One of key elements of the ESI TC mission is to help develop and employ standards and best practices that are needed to make both data and data systems usable and interoperable.

3. THE KNOWLEDGE GENERATION LIFECYCLE

The scope of the ESI TC can be better understood by considering the knowledge generation lifecycle, shown schematically at a high level in Figure 1. This lifecycle depicts the sequence of processes involved in knowledge generation and is useful in identifying where data and information can be enhanced or even lost. Standards play important roles at each stage of the knowledge generation lifecycle and some relevant categories of standards are given in bulleted lists at each stage to illustrate this fact.

The scope of the original DAD TC was essentially limited to the data lifecycle, shown by the inner cycle of

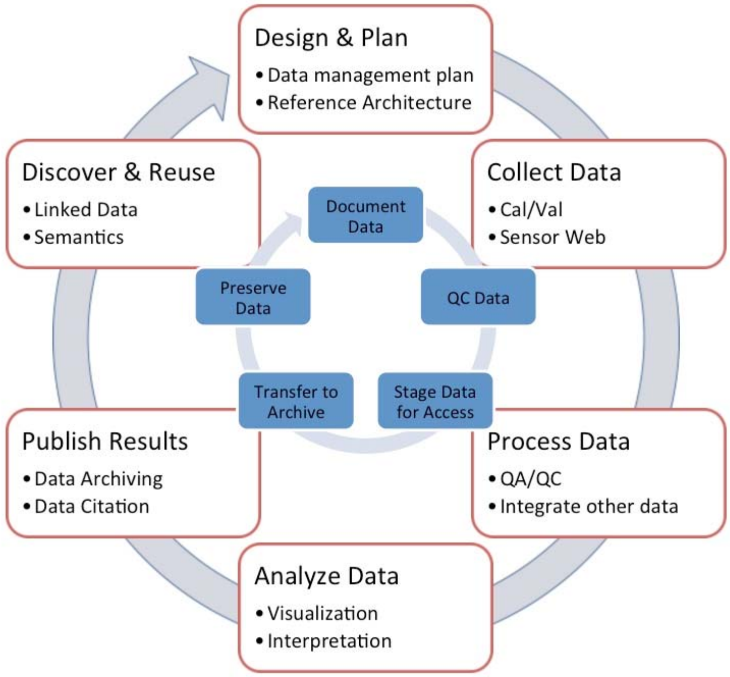

Fig. 1. The Research Knowledge Generation Lifecycle. The inner cycle is the foundational data lifecycle, which is an integral aspect of the outer knowledge generation lifecycle. Example categories of standards that apply in each phase of the knowledge management lifecycle are shown.

5.1. Lifecycle Phases

In the Design and Plan phase of the lifecycle it is important to consider how data will be acquired, evaluated, transferred, stored and documented. These activities are best captured in a data management plan, which is now a requirement of awards made by many agencies. While various agencies and organizations have developed guidelines and templates for writing data management plans, there has yet to be developed an international standard for this.

A reference architecture can be helpful in designing the systems that will realize project goals in a way that makes the components and interfaces of that system more reusable and interoperable with other systems. Reference architectures represent abstract solutions implementing the concepts and relationships identified in a reference model, for which there are several standards such as OSI [1], OAIS [2] and RM-ODP [3].

Research projects often Collect Data from a suite of sensors, which must be controlled, calibrated and monitored. Traceability to reference standards is a fundamental requirement for producing accurate and reliable data. There are also information standards specifying how to calibrate and document instrument performance.

The Process Data phase of the lifecycle includes the many steps needed to harmonize and integrate data streams and otherwise prepare it for analysis. Conformance to standards can be of great benefit in this task. Standards are also useful in applying and documenting the outcomes of quality assurance steps.

T he ability to apply tools and algorithms in the Analyze Data phase is enhanced by the use of standards, such as for encoding, geospatial referencing and portrayal. The reuse of software and procedures is also facilitated by the use of standards.

It is coming to be recognized that it is important to preserve all the outputs of the research process, not just publications. Reproducibility and traceability demand that the data behind the publication be documented, preserved and made available. Placing data into a trusted repository, assigning persistent identifiers to data and referring to those PIDs in the publications is now considered an essential part of the Publish Results phase.

Finally, data must be discoverable and accessible so that future research can build upon those results. The traditional approach to Discovery and Reuse, i.e. placing the data in an archive and populating a metadata catalog, is being extended through linked data and semantic technologies. Of particular importance is the ability for data to be used by disciplines and in contexts other than those in which the data were generated. Mediation and brokering technologies are beginning to be applied to meet this challenge [4].

This talk will describe the knowledge generation lifecycle - the sequence of processes involved in generating knowledge from geoscience observations, simulations and analysis, and discuss some of the important and popular standards involved in each stage of the knowledge generation lifecycle.

4. STANDARDS DEVELOPMENT AND USAGE

One of key elements of the ESI TC mission is to help develop and employ standards and best practices that are needed to make both data and data systems usable and interoperable. The GRSS ESI TC is pursuing this objective through participation in, and collaboration with the Open Geospatial Consortium, OGC [5] and Technical Committee 211 of the International Organization for Standardization, ISO TC211 [6] and the IEEE Standards Association, IEEE-SA [7].

4.1. OGC Standards

The Open Geospatial Consortium develops geospatial standards that are in widespread use within the geoscience community. Among the more commonly known standards and specification that the OGC has developed are:

- CSW - Catalog Service for the Web
- GML - Geography Markup Language
- SOS - Sensor Observation Service
- SensorML - Sensor Model Language
- W*S - Suite of Web Services; and
- GeoSPARQL - for representation and querying of geospatial data for the Semantic Web.

A recently signed MOU between GRSS and OGC will enhance the cooperation and provide support to the GRSS Earth Science Informatics (ESI) Technical Committee. GRSS will provide support to the OGC Earth Systems Science (ESS) Domain Working Group (DWG), contributing to the ESS discussions based on GRSS developments and recommending GRSS related presentations at OGC ESS meetings. GRSS and OGC agree to jointly support presentations, journal articles and other related outreach to highlight the applicability and benefits of geoscience interoperability. OGC and GRSS will work to involve other relevant standards consortia and professional organizations in the development and advancement of geoscience interoperability.

4.2. ISO/TC211 Standards

ISO/TC211 develops international standards for geographic information, addressing the methods, tools and services for management and interoperability of geospatial data. Among the standards that are relevant to the routine activities of many GRSS members are:

- ISO 19115 -- Metadata
- ISO 19119 -- Services
- ISO 19130 -- Imagery sensor models for geopositioning
- ISO 19139 -- Metadata -- XML schema implementation
- ISO 19157 -- Data Quality; and
- ISO 19159 -- Calibration and validation of remote sensing imagery sensors and data.

Currently under development are standards for representing concepts that support the interpretation of, and reasoning with geographic information (ISO 19150), and a common content model for imagery formats (ISO 19163).

GRSS established a liaison relationship with ISO/TC211 in 2004 and has since made regular presentations to its Plenary on GRSS' activities, and has had regular representation on its projects and committees.

4.1. IEEE Standards Association

The IEEE Standards Association facilitates standards development and standards related collaboration to advance global technologies. The IEEE-SA has overseen the development of many of standards that are at the heart of the information infrastructure. GRSS interfaces with the IEEE-SA through Standards Coordinating Committee 40 (SCC40 – Earth Observations).

11. REFERENCES

[1] ISO/IEC, "Open System Interconnection (OSI), ISO/IEC 7498-1," ISO/IEC Standard, 1994.
[2] CCSDS, "Reference Model for an Open Archival Information System (OAIS)," CCSDS Standard, 2012.
[3] ISO/IEC, "ISO/IEC 10746-1: Information technology--Open Distributed Processing--Reference model: Overview," ISO/IEC Standard, 1998.
[4] Nativi, S., M. Craglia, and J. Pearlman. "The brokering approach for multidisciplinary interoperability: a position paper." International Journal of Spatial Data Infrastructures Research 7 (2012): 1-15.
[5] http://www.opengeospatial.org/
[6] http://www.isotc211.org/
[7] http://standards.ieee.org/