

SIRI JODHA S. KHALSA, University of Colorado, NSIDC
RAHUL RAMACHANDRAN, NASA/MSFC

Earth Science Informatics Comes of Age

I. THE EMERGENCE OF EARTH SCIENCE INFORMATICS

The volume and complexity of Earth science data have steadily increased, placing ever-greater demands on researchers, software developers and data managers tasked with handling such data. Additional demands arise from requirements being levied by funding agencies and governments to better manage, preserve and provide open access to data. Fortunately, over the past 10–15 years significant advances in information technology, such as increased processing power, advanced programming languages, more sophisticated and practical standards, and near-ubiquitous internet access have made the jobs of those acquiring, processing, distributing and archiving data easier. These advances have also led to an increasing number of individuals entering the field of informatics as it applies to Geoscience and Remote Sensing. 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. Informatics also encompasses the use of computers and computational methods to support decision-making and other applications for societal benefits.

II. THE GRSS ESI TC

The mission of the IEEE GRSS is “to advance science and technology in geoscience, remote sensing and related fields...” with the society’s fields of interest being “the theory, concepts, and techniques of science and engineering as they apply to the remote sensing of the Earth, oceans, atmosphere, and space, as well as the processing, interpretation and dissemination of this information.” Both the mission statement and the fields of interest of the IEEE GRSS clearly encompass Earth Science

Informatics (ESI). A large number of IEEE GRSS members work in the ESI area and at each IGARSS there are ESI related regular and invited sessions on topics such as GIS, semantic web, data provenance, sensor web, GEOSS, standards, data processing, data management, and decision support. However, until recently GRSS had yet to set up an ESI related technical committee.

Given the rapid growth in informatics, GRSS decided to expand the original mission of its existing Data Archiving and Distribution Technical Committee (the DAD TC) and rename it the Earth Science Informatics Technical Committee (ESI TC), focusing on advancing the application of informatics to the geosciences and remote sensing.

By establishing an ESI Technical Committee at GRSS we provide a home to GRSS ESI professionals, enabling them to exchange information and knowledge while setting a research agenda and making GRSS more visible in the broader ESI community. We aim to provide technology advice to major national and international ESI initiatives. An ESI TC also helps GRSS attract more ESI professionals to the GRSS.

III. 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 listed 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 Figure 1. The data lifecycle is part of the more comprehensive knowledge generation lifecycle, and could be said to underpin it.

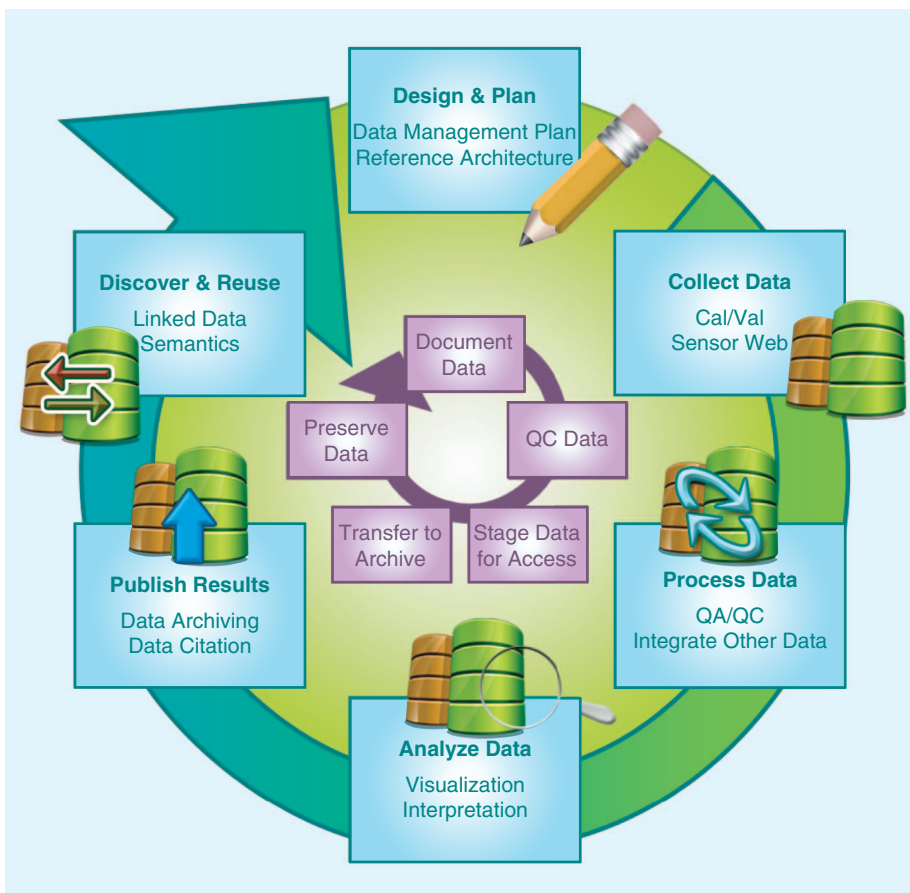


FIGURE 1. The Research Knowledge Generation Lifecycle. The inner cycle is the foundational data lifecycle, which is an integral aspect of the outer knowledge generation lifecycle. Example categories of standards that apply in each phase of the knowledge management lifecycle are shown.

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.

The 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].

IV. 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], Technical Committee 211 of the International Organization for Standardization, ISO TC211 [6], and the IEEE Standards Association, IEEE-SA [7].

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
- GeoSPARQL—for representation and querying of geospatial data for the Semantic Web.

A recently established 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.

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
- 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.

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).

V. OVERVIEW OF THE ESI TC

As stated earlier, the mission of the ESI TC is to advance the application of informatics to the geosciences and remote sensing, and to provide a platform for ESI professionals to collaborate. The fields of interest of the ESI TC include, but are not limited to:

- Data and information policies, stewardship, preservation, provenance and quality
- Knowledge representation, information models for the spatial and temporal relationships between entities in

the Geosciences (e.g., spatial and process ontologies, vocabularies, semantic web)

- Cyberinfrastructures, interoperability, standardization, web service, sensor web and cloud computing
- Improving data discovery and access
- Tools supporting spatial and temporal analyses and their applications including decision support systems, tools and systems to model the Earth system, tools to visualize and analyze geoscience data, information, and knowledge
- Emerging information technologies trends and both their impact and applications in the geosciences.

The ESI TC will be sponsoring two invited sessions at IGARSS 2014. The first session, titled "Implications of Big Data to Remote Sensing," will focus on evaluating different big data technologies that leverage a "shared nothing architecture" and distributed file storage systems to support reliable processing and analysis of satellite imagery. The second session is a joint ESI TC and OGC session titled "Advancing Science through Management of the Geospatial Data Lifecycle". The focus of this session is to explore the role of standards at different stages of the data lifecycle. As science becomes more reliant on information technology, data standards are as vital as uniform standards for weights and measures. In addition to these special sessions, ESI TC will seek to sponsor either a *TGRS* or *JSTARS* special issue focusing on specific Earth Science Informatics topics.

VI. CALL FOR PARTICIPATION

As science and technology progress, the knowledge generation lifecycle evolves, impacting everyone involved including the scientists and engineers who design and operate instruments, processing systems and numerical models, and acquire, validate, analyze, manage and interpret data. GRSS members are thus encouraged to engage with the ESI TC in its mission to bring together those GRSS members interested in advancing the field of informatics. Specific opportunities to contribute include serving as subject matter experts in the development and/or review of standards, presenting ESI related research at IGARSS and submitting papers to the special issue of the Journals. To participate contact the ESI TC chairs Dr. Ramachandran and Yue (ramachr@uah.edu, geopyue@gmail.com) and join the IEEE GRSS Earth Science Informatics group on LinkedIn [8].

REFERENCES

- [1] ISO/IEC *Open System Interconnection*, ISO/IEC Standard 7498-1, 1994.
- [2] *Reference Model for an Open Archival Information System*, CCSDS Standard, 2012.
- [3] ISO/IEC *Information Technology—Open Distributed Processing—Reference Model: Overview*, ISO/IEC Standard 10746-1, 1998.
- [4] S. Nativi, M. Craglia, and J. Pearlman, "The brokering approach for multidisciplinary interoperability: A position paper." *Int. J. Spatial Data Infrastruct. Res.*, vol. 7, pp. 1–15, 2012.
- [5] [Online]. Available: <http://www.opengeospatial.org/>
- [6] [Online]. Available: <http://www.isotc211.org/>
- [7] [Online]. Available: <http://standards.ieee.org/>
- [8] [Online]. Available: <http://www.linkedin.com/groups/IEEE-GRSS-Earth-Science-Informatics-5136161>

GRS