Information Quality Cluster and Usability

ESIP Summer Meeting

July 20, 2018

H. K. “Rama” Ramapriyan – SSAI/GSFC
Hampapuram.Ramapriyan@ssaihq.com
David Moroni – JPL/CalTech David.F.Moroni@jpl.nasa.gov
Ge Peng – CICS-NC/NCEI Ge.Peng@noaa.gov

This work was a result of the authors' participation in the ESIP IQC. Ramapriyan's work was supported by NASA under a contract with Science Systems and Applications, Inc. Peng's work was supported by NOAA under a grant with CICS-NC. Moroni's work was supported by NASA under a contract with the Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA. U.S. government sponsorship is acknowledged.

Topics

- **Importance of Information Quality**
- **Information Quality Cluster**
- **Information Quality & Data Lifecycle**
- **Typical Issues Revealed via Use Cases**
- **Information Quality and Usability**

Importance of Information Quality

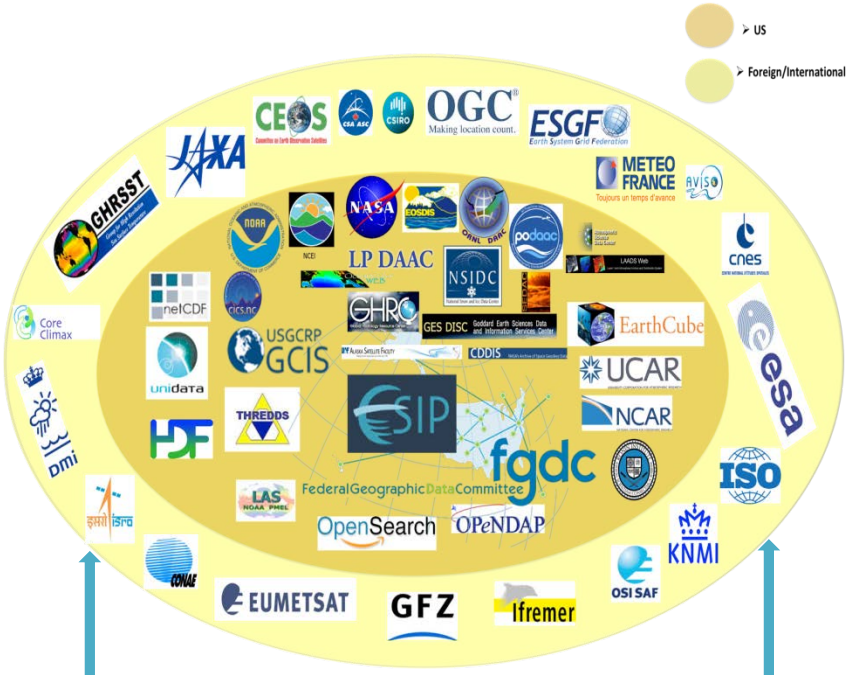
- **Trustworthiness of data depends on information quality**
- **Section 515 of the Treasury and General Government Appropriations Act for Fiscal Year 2001 (Public Law 106-554; H.R. 5658) directed the Office of Management and Budget (OMB) to issue government-wide information quality guidelines.**
- **OMB's final guidelines, entitled "Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by Federal Agencies," (February 22, 2002) call for agencies to develop their own IQ guidelines**
- **Agencies have developed information quality guidelines**
  - Department of Energy Information Quality Guidelines
  - Guidelines for Ensuring and Maximizing the Quality, Objectivity, Utility, and Integrity of Information Disseminated by the Environmental Protection Agency
  - NASA Guidelines for Ensuring the Quality of Information
  - NOAA Information Quality Guidelines
  - NSF Information Quality Guidelines
  - USGS Information Quality Guidelines

From NASA's Guidelines

- **NASA's information from its missions and programs is used by:**
  - **Government and national and international policymakers to enable sound and better public policy**
  - **NASA's scientists and others cooperating with NASA to pursue their important work**
  - **The media in describing to the public the importance and advances of research**
  - **The educational community to educate a new generation of citizens in science, math, and engineering**
  - **Members of the public to enable them to be knowledgeable and inspired about NASA's goals and accomplishments.**

Information Quality Cluster

VISION: *Become an authoritative and responsive resource of information and guidance to data providers on how best to implement data quality standards and best practices*



Scope of Mutual Influence and Sources of Domain Knowledge

- Use case analyses - **http://wiki.esipfed.org/index.php/File:Summary_of_Evaluations_of_Use_Cases.pdf**
- Fostering Collaborations
  - Invited speakers from US and abroad at ESIP meetings and Monthly telecons
- AGU sessions on information quality
- Publications and Presentations
- References – extensive list maintained
- White paper on Uncertainty – in preparation

Quality Aspects:

Science, Product, Stewardship, Service

Want to learn more?

Contact Info at:
http://wiki.esipfed.org/index.php/Information_Quality

Information Quality

- **Science quality**
  - **Accuracy, precision, uncertainty, validity and suitability for use (fitness for purpose) in various applications**
- **Product quality**
  - **how well the scientific quality is assessed and documented**
  - **Completeness of metadata and documentation, provenance and context, etc.**
- **Stewardship quality**
  - **how well data are being managed, preserved, and cared for by an archive or repository**
- **Service Quality**
  - **how easy it is for users to find, get, understand, trust, and use data**
  - **whether archive has people who understand the data available to help users.**

Information Quality is a combination of all of the above

Information Quality & Data Product Lifecycle

Information Quality Aspect	Life Cycle Stage	Responsible Group
Science	Define, develop, and validate	Science Team
Product	Produce, assess and deliver (to an archive or data distributor)	Science Team
Stewardship	Maintain, preserve and disseminate	Archive/Distributor
Service	Enable use, provide support and service	Archive/Distributor

See Ramapriyan, Peng, Moroni and Shie, "Ensuring and Improving Information Quality for Earth Science Data and Products", D-Lib Magazine, July/August 2017, https://doi.org/10.1045/july2017-ramapriyan

Typical Issues Revealed via Use Cases

"usability issues"

Key issue(s)
Multiple inputs contributing to the product, derived from diverse sources. Some inputs may be of unknown/ undocumented data quality.
Not all types of users need the same level of detail in documentation.
User needs to identify and retrieve dataset(s) of interest from predefined collections based on the metadata/documentation provided.
Varying quality of data depending on people collecting them.
Guidance regarding how to use already available quality indicators.
Large differences between buoy and satellite-derived data.
Need for a service to apply specific quality filtering levels or flags while extracting data values from a file.
Selecting the most relevant and useful datasets among those containing similar geophysical parameters.
Users need to know error propagation as higher level products are generated.
Use of data outside "normal" spatial coverage area.
Geometric error in land mask.
Guidance to Principal Investigators about proper level of data quality documentation.
Conformance of netCDF or HDF files (granules) to the Climate Forecast (CF) and Attribute Convention for Dataset Discovery (ACDD) metadata models.
Quality flag that marks questionable ice values, rather than filtering out such values.
Need improved identification and characterization of outliers.
Provide sufficient information to users such that they can judge and replicate our products.
Insurance company trying to assess the coastal region that is vulnerable to storm surge finds that only limited types of data available.
Need to know how much of a pixel is comprised of specific spaceborne sensor inputs and /or in situ measurements.
User needs data with spatial resolution under 10 km and maximum data coverage with minimal data dropouts.
Accuracy of product documentation vs. provisional product contents.

Information Quality and Usability

- **Capture, description, discovery, and *usability* of information about data quality in Earth science data products is critical for proper use of data.**
- **Usability**
  - **Relevance**
  - **Completeness**
  - **Clarity**
  - **Easy access**
  - **Timeliness**
- **Usability depends on users – i.e., design should consider target user community**
- **Usability of data vs. Usability of information about data quality – latter is of greater concern to IQC**