ESIP Information Quality Cluster

Hampapuram Ramapiyana,1,2 (Hampapuram.Ramapiyan@ssaihq.com), Ge Peng3,4 (Ge.Peng@noaa.gov) and David Moron5 (David.F.Moron@jpl.nasa.gov)

1 NASA Goddard Space Flight Center, 2 Science Systems and Applications, Inc., 3 NOAA’s Cooperative Institute for Climate and Satellites, North Carolina (CICS-NC), 4 NOAA’s National Centers for Environmental Information (NCEI), 5 Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA

Objectives

- Bring together people from various disciplines to assess aspects of quality of Earth science data
- Establish and publish baseline of standards and best practices for data quality for adoption by inter-agency and international data providers
- Become an authoritative and responsive resource of information and guidance to data providers on how to best implement certain data quality standards and best practices for their datasets
- Build framework for consistent capture, harmonization, and presentation of data quality for the purposes of climate change studies, Earth science and applications
- Objectives evolve with participant inputs

Background

- **ESIP work in years past**
  - IQ Cluster kick-off Meeting – Jan 6, 2011
  - Data Quality Session – Santa Fe Sumer Meeting – July 14, 2011
  - Led by Greg Leptoukh (NASA GSFC), who passed away on January 12, 2012
  - Data/Information Quality Birds of a Feather Session – Winter Meeting – January 2014, led by Carol Meyr
  - Information Quality Cluster session – Summer Meeting – July 2014, led by Gilberto Vicente
  - All focused on identifying challenges, use cases, representation of DQ/IQ to help users

- **Other relevant activities**
  - NASA Earth Science Data System Working Groups (ESDSWG) – Metrics Planning and Reporting WG (Product Quality Checklists) – 2010-2012
  - NASA ESDSWG Data Quality WG (Recommendations) – 2014-present
  - NOAA Data and Stewardship Maturity Matrices – 2008 – present
  - EUMETSAT CORE-CLIMAX System Maturity Matrix (e.g., http://presentations.copernicus.org/EGU2015-10158_presentation.pdf - 2014)
  - CEOS Essential Climate Variables (ECV) Inventory Questions
  - GEOSS Data Quality Guidelines
  - Quality Assurance framework for Earth Observation (QA4EO)
  - NCAR Community Contribution Pages

Aspects of Information Quality – Key Defining Questions

- **Science Data Quality**
  - How accurate, precise and valid are the data?
  - How well have the error sources and uncertainties been characterized and documented?
- **Product Quality**
  - Has science quality been assessed and well documented?
  - How well have quality procedures and methods been defined, implemented, and documented?
  - How complete are metadata and documentation?
- **Stewardship Quality**
  - How well are data being managed and preserved by an archive or repository?
  - How well are science and product quality information being documented and captured in metadata?
  - How easy is it for users to find, get, understand, trust, and use data?
  - Does archive have people who understand the data available to help users?

IQ Cluster- Suggested Activities

- Coordinate use case studies with broad and diverse applications, collaborating with the ESIP Data Stewardship Committee and various national and international programs
- Identify additional needs for consistently capturing, describing, and conveying quality information
- Establish and provide community-wide guidance on roles and responsibilities of key players and stakeholders including users and management
- Prototype conveying quality information to users using approach proposed by Vicente (Summer 2014)
- Evaluate NASA ESDSWG DQWG recommendations and propose possible implementations
- Establish a baseline of standards and best practices for data quality, collaborating with the ESIP Documentation Cluster and Earth Science agencies.
- Engage data provider, data managers, and data user communities as resources to improve our standards and best practices.