GEOSS Data Management & Data Sharing Principles and TRUST: Implications for Data Quality

Dr. Robert R. Downs

NASA Socioeconomic Data and Applications Center (SEDAC)

Center for International Earth Science Information Network (CIESIN)

The Earth Institute, Columbia University

Prepared for presentation to the Pre-ESIP Workshop:

Developing Community Guidelines for Consistently Curating and Representing Dataset Quality Information

13 July 2020

Additional Principles to Facilitate Data Use, with Implications for Data Quality

- GEOSS Data Sharing Principles. 2015. Group on Earth Observations (GEO) http://earthobservations.org/open_eo_data.php
- GEOSS Data Management Principles. 2015. Group on Earth Observations (GEO) http://earthobservations.org/open_eo_data.php
- TRUST Principles for Digital Repositories
  - Lin, D., Crabtree, J., Dillo, I. et al. The TRUST Principles for digital repositories. Sci Data 7, 144 (2020). https://doi.org/10.1038/s41597-020-0486-7

GEOSS Data Management Principles: Data Quality Implications

Discoverability

- DMP-1: Metadata for Discovery

** Describe data quality in metadata*

Usability

- DMP-3: Data Encoding

** Encode data quality information*

- DMP-4: Data Documentation

** Describe data quality in documentation*

- DMP-5: Data Traceability

** Capture data quality information early*

- DMP-6: Data Quality-Control

** Assess data quality and record the results*

Accessibility

- DMP-2: Online Access

** Place data quality information online*

Preservation

- DMP-7: Data Preservation

** Preserve data quality information with data*

- DMP-8: Data and Metadata Verification

** Verify data quality assessments and records*

Curation

- DMP-9: Data Review and Reprocessing

** Review and reassess data quality routinely*

- DMP-10: Persistent & Resolvable Identifiers

** Provide persistent access to data quality information*

GEO GROUP ON EARTH OBSERVATIONS

Source: GEOSS Data Management Principles. 2015. Group on Earth Observations (GEO) http://earthobservations.org/open_eo_data.php

GEOSS Data Sharing Principles: Data Quality Implications

GEOSS Data Sharing Principle 1.

Data, metadata and products will be shared as Open Data by default, by making them available as part of the GEOSS Data Collection of Open Resources for Everyone (Data-CORE) without charge or restrictions on reuse, subject to the conditions of registration and attribution when the data are reused;

** Describe data quality in metadata that is freely distributed with open data.*

GEOSS Data Sharing Principle 2.

Where international instruments, national policies or legislation preclude the sharing of data as Open Data, data should be made available with minimal restrictions on use and at no more than the cost of reproduction and distribution;

** Provide data quality information with minimal restrictions and limited costs.*

GEOSS Data Sharing Principle 3.

All shared data, products and metadata will be made available with minimum time delay.

** Conduct and share the results of data quality reviews in a timely manner.*

Derived from: GEOSS Data Sharing Principles Implementation Guidelines 2016-2025.
http://earthobservations.org/open_eo_data.php

TRUST Principles for Digital Repositories: Data Quality Implications

- **Transparency**
  - To be transparent about specific repository services and data holdings that are verifiable by publicly accessible evidence.

  * Disseminate the results of data quality assessments with the data.
- **Responsibility**
  - To be responsible for ensuring the authenticity and integrity of data holdings and for the reliability and persistence of its service.

  * Ensure that data quality is assessed and described for prospective users.
- **User Focus**
  - To ensure that the data management norms and expectations of target user communities are met.

  * Foster usability by enabling users to understand and assess data quality.
- **Sustainability**
  - To sustain services and preserve data holdings for the long-term.

  * Provide consistent data lifecycle capabilities for assessing data quality.
- **Technology**
  - To provide infrastructure and capabilities to support secure, persistent, and reliable services.

  * Develop and maintain capabilities to assess and describe data quality.

Lin, D., Crabtree, J., Dillo, I. et al. The TRUST Principles for digital repositories. Sci Data 7, 144 (2020). https://doi.org/10.1038/s41597-020-0486-7

Summary of Implications for Data Quality

- Data Management
  - Facilitate discovery, access, usability, preservation, and curation of data quality information.
- Data Sharing
  - Provide data quality information with open data, minimal restrictions, limited costs, in a timely manner.
- Digital Repositories
  - Employ transparency, responsibility, user focus, sustainability, and technology to provide data quality information with data.

GEO GROUP ON EARTH OBSERVATIONS