Towards Data Management Principles

Alessandro Annoni
European Commission
Joint Research Centre
GEO-XI Plenary
13-14 November 2014
Geneva, Switzerland

Towards Open Data in the Developing World: GEO Data Sharing Principles and Data Management Guidelines

DSWG and DMP TF Side Event at GEO-XI Plenary Session

GROUP ON EARTH OBSERVATIONS

Data Management (DM) in GEO

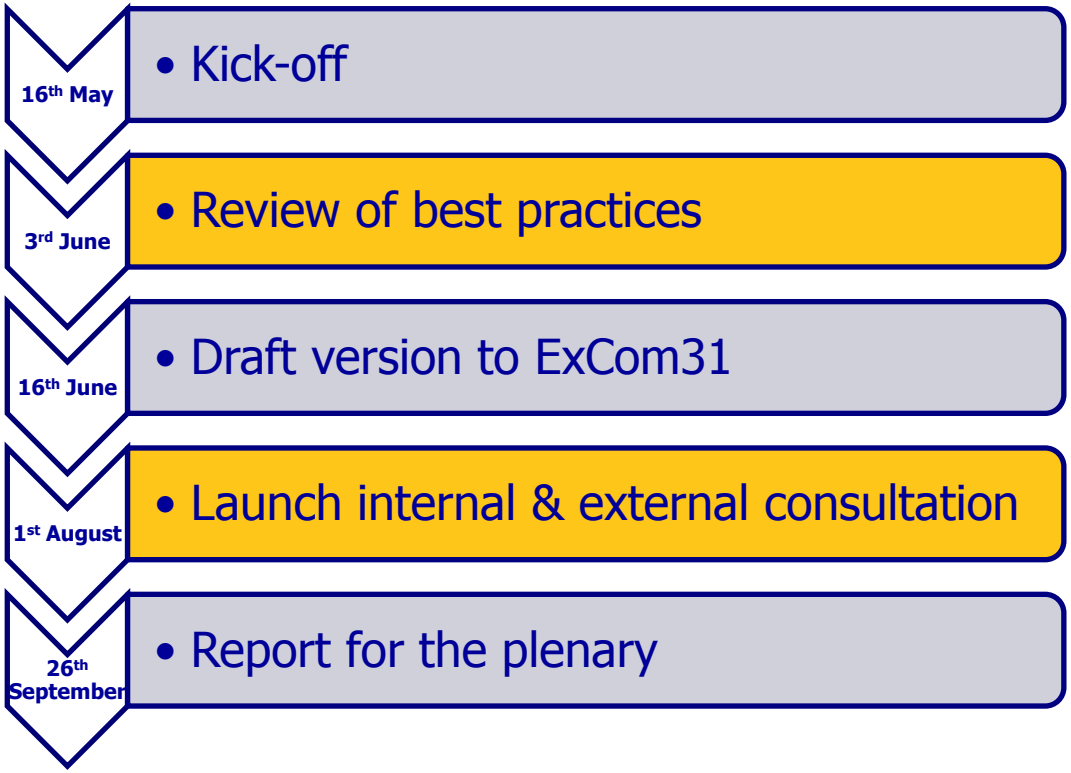
- A fundamental recognised aspect for GEO
  - "several" issues already addressed by working teams, e.g., access, documentation, data quality and interoperability
- But clear need to
  - ensure that "all relevant" aspects of DM are properly addressed
  - raise awareness across GEO community
  - promote adoption of "common" data management principles (as done in the past for "data sharing")
- Why? Users' needs can only be addressed if:
  - access to data is ensured (e.g. GCI)
  - those data are of known quality and lineage
  - and are properly managed

GROUP ON EARTH OBSERVATIONS

TF MEMBERSHIP

- Good geographical coverage
- Very good and complementary expertise
- High level of commitment to deliver in short time.

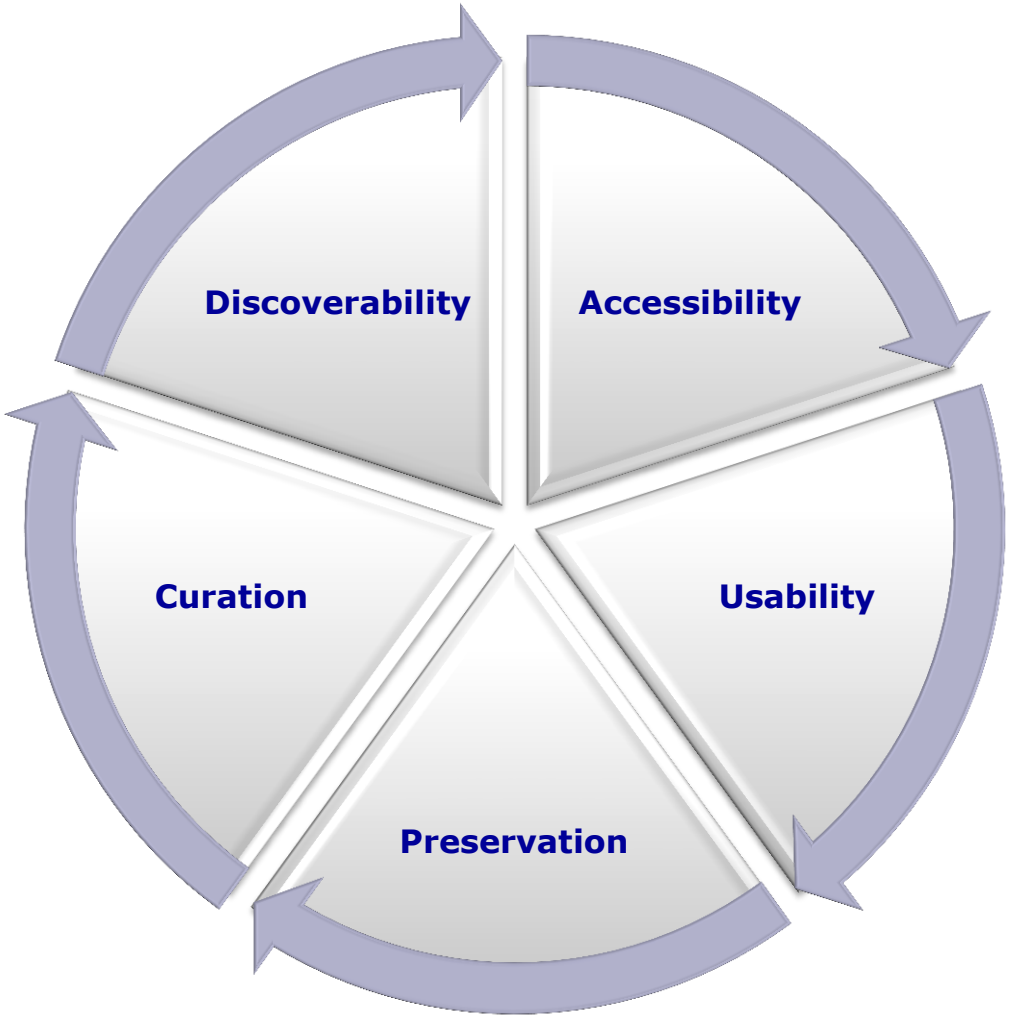
Roadmap (… in only 4 months !! ..)



GROUP ON EARTH OBSERVATIONS

Principles (overview)

The value of each EO is maximized through data life-cycle management, including five foundational elements



GROUP ON EARTH OBSERVATIONS

Discoverability

1. Data and metadata will be discoverable through catalogues and search engines, access and use conditions, including licenses, will be clearly indicated.

Accessibility

2. Data will be openly accessible with minimum delay and cost.
3. Data will be accessible via online services, including, at minimum, direct download but preferably user-customizable services for visualization and computation.

GROUP ON EARTH OBSERVATIONS

Usability

4. Data should be **structured** using widely accepted encodings aligned with organizational needs and observing methods, with preference given to non-proprietary international standards.
5. Data will be **comprehensively documented**, including all elements necessary to access, use, understand, and process, preferably via metadata based on international standards.
6. Data will **include provenance metadata** indicating the origin and processing history of raw observations and derived products, to ensure full traceability of the product chain.
7. Data will be **quality-controlled** and the results of quality control be indicated in metadata

GROUP ON EARTH OBSERVATIONS

Preservation

8. Data will be protected from loss and preserved for future use.
9. Data and associated metadata will be periodically verified to ensure integrity, authenticity and readability.

GROUP ON EARTH OBSERVATIONS

Curation

10. Data will be managed to perform corrections and updates in accordance with reviews, and to enable reprocessing as appropriate
11. Data will be assigned appropriate persistent, resolvable identifiers to enable data citation to enable providers to be acknowledged about use of their data.

GROUP ON EARTH OBSERVATIONS

External Consultation

- > 50 institutions, organizations and projects contacted
  - about 50% of responded ! (*Considering the timing of the request - end of summer break- this can be considered a great success)*
- comments indicate that discussion is timely, and GEOs efforts appreciated!
  - similar discussions and activities are under way in various parts of the community, and comments include suggestions that GEO may either participate in these activities, or perhaps even take an active role in overall coordination
- overall, the Principles meet with agreement, and no negative response
- comments often addressed issues of the implementation rather than the principles themselves
  - common to various responses is the question about resources required to implement data management principles, as well as where to find those resources.

GROUP ON EARTH OBSERVATIONS

External Consultation

n.		Importance			Feasibility		
		L	M	H	L	M	H
1	Data and metadata discoverable	0%	17%	83%	0%	29%	71%
2	Data openly accessible	0%	21%	79%	17%	48%	35%
3	Data accessible via online services	0%	22%	78%	9%	73%	18%
4	Data structured using widely accepted encodings with preference to non-proprietary international standards	0%	26%	74%	9%	57%	35%
5	Data comprehensively documented	4%	13%	83%	17%	61%	22%
6	Data include provenance metadata	4%	21%	75%	13%	65%	22%
7	Data quality-controlled	0%	29%	71%	26%	57%	17%
8	Data protected and preserved	4%	13%	83%	22%	30%	48%
9	Data and associated metadata periodically verified	13%	30%	57%	23%	45%	32%
10	Data managed to perform corrections and updates	13%	30%	57%	21%	58%	21%
11	Data with persistent identifiers	5%	27%	68%	9%	52%	39%

GROUP ON EARTH OBSERVATIONS

Conclusions

- Harmonising Data Management practices is important to increase efficiency and facilitate multiple uses of data
- Data management principles and data sharing principles are intrinsically connected
  - Harmonisation of terminology is important
- New initiatives to be monitored and pssible convergence explored
  - e.g. UN Data Revolution report
- Toward implementation guidelines
  - Cost/benefits considerations and feasibility aspects to be taken into account *(common problem with DSWG)*
  - e.g. intelligent solutions to be explored: e.g. introducing a scoring system and/or a qualifier *(e.g. GEOWIQUA guidelines for quality aspects)*
  - Mechanism to collect feedback from users to be included
  - Impact on GEOSS central components (GCI) and contributing systems to be addressed

GROUP ON
EARTH OBSERVATIONS

Thanks for your attention