GROUP ON EARTH OBSERVATIONS

GEO-XI

13-14 November 2014

Towards Data Management Principles

Document 7

For consultation.

Towards GEO Data Management Principles

(Report of the Data Management Principles Task Force)

1 BACKGROUND

The Group on Earth Observations (GEO) since its inception in early 2000s has been working to facilitate 'data management approaches that encompass a broad perspective of the observation data life cycle'. Several aspects of data management, e.g., access, documentation, data quality and interoperability, have been addressed by some working teams; GEO-X Plenary endorsed the establishment of a Data Management Task Force (DMP-TF) to identify Life-Cycle Data Management Principles.

Terms of Reference (see Appendix 1) have been drafted by the Infrastructure Implementation Board (IIB) and the Data Sharing Working Group (DSWG), as well as 19 individuals nominated by 16 GEO Members and Participating Organizations (see Appendix 2). Administrative support provided by the GEO Secretariat also helped the DMP-TF complete its preparation and kick off its work on 16 May 2014.

2 RATIONALE

The DMP-TF has been set-up to support GEO in the definition and endorsement of common GEOSS Data Management Principles.

The three principles initially drafted by the Infrastructure Implementation Board (IIB) are: (i) Ensure long time data preservation and distribution; (ii) Ensure the quality information of Earth observation (EO) data; and (iii) Answer EO user needs. These principles have been subsequently reviewed by the DMP-TF which has assessed their feasibility and proposed a new formulation.

The DMP-TF worked with GEO collectively, including the Implementation Boards, the Data Sharing Working Group (DSWG) and the Implementation Plan Working Group (IPWG). The work of the DMP-TF concentrated on developing principles for:

- Ensuring data are properly managed (including data citation), accessible, archived and long term preserved (when appropriate);
- Ensuring data are properly documented (metadata), quality controlled and quality assessed, delivered, and updated in ways to facilitate access and re-use of information made available through the GEOSS Common Infrastructure (GCI);
- Facilitating the link between user needs and data availability, especially with regard to the needs of users from developing countries (e.g. by identifying existing sources of requirements already approved by the relevant user community);
- Facilitating interoperability of GEOSS data resources by promoting a progressive harmonization/standardization of content (data models, thesauri, coding list) and dissemination and usage rights in order to facilitate their re-use at global or regional scales.

The DMP-TF started operations in mid-May with the ambition to deliver "draft principles for adoption" at the GEO Plenary. The time at disposal was very limited also considering the summer break and the need for a broad consultation with the GEO Community on the DMP-TF initial proposal.

At the 31st Executive Committee meeting, the original plan was reviewed and the TF was asked to produce a report for consultation aiming at moving steps along agreed principles to be adopted in 2015 together with the new GEOSS 10-Year Implementation Plan. This report will allow GEO Members and Participating Organizations to launch an internal consultation on the proposed principles and meanwhile will allow the TF to make some progress on possible implementation guidelines.

3 TASK FORCE ACTIVITIES

3.1 Roadmap

The TF operated in a very short period accordingly to the following roadmap agreed among TF members:

- Kick-off 16 May
- Collect existing best practices and "formal" procedures Continuous
- 1st Review, Identify similarities and differences 30-May
- Identify Categories for principles 3 June
- Propose text for each category 6 June
- Discuss and Assess feasibility for principles 10 June
- Consolidation of draft version to Executive Committee 16-June
- Send Present preliminary results to Executive Committee 16 June
- Start "internal" consultation with GEO community End June
- Review the draft to align with GEO 2025 vision End June
- Present preliminary results to Executive Committee 9July
- Start "internal" and "External" consultation 1 August
- Review principles according to comments received mid-September
- Produce the Report for the Plenary 26 September

3.2 Communication

3.2.1 Teleconference

At the moment of this report the TF already met 7 times (by teleconference only):

- 2014-05-16 Kick-off teleconference;
- 2014-05-27 2nd teleconference;
- 2014-06-10 3rd teleconference;
- 2014-06-23 4th teleconference;
- 2014-07-17 5th teleconference;
- 2014-09-04 6th teleconference;
- 2014-09-19 7th teleconference.

At the kick-off meeting, Alessandro Annoni (representing the European Commission and also the IIB) and David Halpern (representing COSPAR and also a co-chair of DSWG) were appointed by the GEO

Secretariat as ad-hoc co-chairs. They agreed to serve as co-chairs and encouraged other TF members to become additional co-chairs, setting a deadline for self-nominators to inform the GEO Secretariat. At the end of the consultation process, no additional candidatures were received.

Nearly all members of the TF actively participated in the work. An attempt was done by the Secretariat to contact the few members who have yet to contribute to TF activities about how to increase their involvement.

3.2.2 Document Management / Collaborative Platform (WIKI space)

In addition to using the regular teleconference and e-mail communications, the TF agreed to use a WIKI system hosted by EC/JRC as a collaborative platform for collecting relevant materials, sharing the ideas of Principles and exchanging opinions. This tool allows record keeping of all documents collected, minutes of the meetings (including comments, actions and decisions) thereby ensuring a better understanding/documenting of the reasons for certain deliberations.

The WIKI could be made accessible to the IPWG and the GEO Community for their reference, .

3.3 Drafting Data Management Principles (DMPs)

The first activity of the DMP-TF has been an inventory of Data Management Principles adopted in different initiatives and regions. Each member of the TF has contributed by identifying and submitting reference materials and producing a short summary describing the underlying principles. Among others, the following initiatives have been particularly considered by the TF:

- Australian Government - Bureau of Meteorology;
- CEOS contribution;
- COSPAR Task Group on GEO;
- e-IRG Data Management Task Force;
- ESA contribution;
- EuroSDR (European National Mapping Agencies);
- INSPIRE (Infrastructure for Spatial Information in the European Community);
- WMO Resolutions;
- US Open Data Policy;
- US Strategy for Civil Earth Observations ("EO Strategy");
- Various examples in UK.

A Working Group comprised of Mirko Albani, Siri Jodha Khalsa, Françoise Genova, and Jeff de la Beaujardière (chair) agreed to assist the development of GEO Data Management Principles by defining high-level topics that could be used to categorize the various principles from other organizations submitted by GEO DMP TF members.

This Working Group produced 3 lists:

1. Data Qualifiers (define the characteristic of the data: open, discoverable, etc.);
2. Data lifecycle (define what should be done in the different phases of the data management cycle: data collection, quality control, etc.);
3. Grouping by categories (aggregate principles according to their nature: technical, legal, etc.).

The TF agreed that a good starting point was to further develop List 1 but meanwhile recognized the value of List 2 (that could be used in the near future to produce guidelines and best practices) as well

as List 3 (that is the most effective for communication purposes)At the time of the report to the 31st Executive Committee, the TF had established a list of 11 principles. After the Executive Committee consultation, the TF finally divided the 11 principles into 5 categories:

- Discoverability;
- Accessibility;
- Usability;
- Preservation;
- Curation.

3.4 Consultation

The TF based its proposal on the analysis of a number of existing initiatives. However the TF considered necessary to launch a broad consultation internally (through the Executive Committee) and externally (> 50 selected organizations/initiatives) to collect additional feedback on its initial proposal.

3.4.1 Executive Committee Consultation

The DMP-TF delivered a report on Data Management Principles to the 31st Executive Committee for consultation. The report was presented by Alessandro Annoni who, on behalf of the TF, asked guidance from the Executive Committee in particular to clarify the following:

GEO is based on voluntary contributions. A proper Data Management Framework requires both professionalism and resources to ensure quality data. The TF can be very prescriptive, or propose a lighter approach. Both have advantages and negative consequences. We believe that we should not compromise the level of ambition, but can adapt to different situations in the implementation phase. Is this a correct approach or alternative solutions are recommended?

The level of ambition is tightly connected to the future vision of GEO. In particular the word "robust information system" requires strict adherences to data management principles that will be proposed. If the Executive Committee has any specific recommendations, please identify them.

The GEO community is quite heterogeneous, so principles that can be easily respected by some communities of practice (space agencies, meteorological services, etc.) could be very challenging for other organisations. The Executive Committee could advice the TF if a single approach should be adopted or if principles should be adapted to different categories.

The TF noted an increasing importance of Big Data (from social media), crowdsourcing information, sensor web, etc. that introduces new data streams in which data collection is not strictly controlled. These data also generate new demand for data archiving and preservation. Should EO registered resources only or also data "not controlled" by GEO, but of potential high interest for the Societal Benefis Areas, be included?

Professional data management require resources. If the adoption of the Data Sharing Principles (and GEOSS DataCORE in particular) only speaks about lost revenues (for data not previously accessible) or vice versa a professional life cycle data management will have additional costs (if not already in place). At which level should GEO propose principles that could have a potential financial impact?

All Executive Committee representatives were invited to send written answers to questions posed in the presentation and directly to the DMP-TF, with copy to Executive Committee and the GEO Secretariat. The DMP-TF made a formal request (Appendix3) through GEO Secretariat regarding the subject on 1st August. At the moment of this report, 2 responses were received from Gabon and EC (Appendix 5).

3.4.2 External consultation

On 1st August, the GEO Secretariat sent a request (Appendix 4) for consultation on Data Management Principles to more than 50 entities, including related organizations, projects, as well as GEO Implementation Boards, and Working Groups. The consultation asked participants to list the top 5 primary objectives of good Data Management and to rate the importance of the specific data management principles and judge their feasibility of implementation.

By 26th September, 26 responses had been received. The analysis of consultation results is listed in the following section.

4 PROPOSED DATA MANAGEMENT PRINCIPLES[1]

Each Earth observation is unique because the observation occupies a specific location and time in an environment that is continuously changing. No two Earth environmental observations are the same, making each Earth observation an irreplaceable asset to understand the past, describe the present, and forecast the future of the global integrated Earth system.

The value of each Earth observation is maximized through data life-cycle management, including the following foundational elements:

- Discoverability;
- Accessibility;
- Usability;
- Preservation;
- Curation.

Discoverability

1. Data and all associated metadata will be discoverable through catalogues and search engines, and data 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.

Usability

4. Data should be structured using encodings that are widely accepted in the target user community and 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 formal structured 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 shall be indicated in metadata; data made available in advance of quality control will be flagged in metadata as unchecked.

[1] The principles have been slightly modified taking into account consultation comments.

Preservation

8. Data will be protected from loss and preserved for future use; preservation planning will be for the long term and include guidelines for loss prevention, retention schedules, and disposal or transfer procedures.
9. Data and associated metadata held in data management systems will be periodically verified to ensure integrity, authenticity and readability.

Curation

10. Data will be managed to perform corrections and updates in accordance with reviews, and to enable reprocessing as appropriate; where applicable this shall follow established and agreed procedures.
11. Data will be assigned appropriate persistent, resolvable identifiers to enable documents to cite the data on which they are based and to enable data providers to receive acknowledgement of use of their data.

5 RESULTS OF THE EXTERNAL CONSULTATION

About 50% of institutions, organizations and projects contacted have responded to the request for comments on the draft Data Management Principles for GEO. Considering the timing of the request (end of summer break) this can be considered a *great* success. Also the comments received indicate that this discussion is timely, and GEOs efforts are appreciated.

A brief summary of responses is presented below (a more detailed digest as well as the individual questionnaires are provided in Appendix 6).

5.1 General comments and observations

Overall, the draft Data Management Principles elaborated by the GEO DMP-TF have been met with agreement, with no negative response or rejection of the proposed Principles. Comments received questioned some details and requested some clarifications, and often addressed issues of implementation rather than the principles themselves. It is evident that similar discussions and activities are underway 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. A further issue common to various responses was the question of resources required to implement data management principles, as well as where to find those resources. While this issue was deliberately not included in the formulation of the principles so far, it is clear that any further elaboration of implementation guidelines should address the resources issue. The DMP-TF suggests that it should be clearly stated that proper Data Management is a whole life-cycle undertaking, and that adequate provisions need to be made at every stage of the data life-cycle, starting already at the data acquisition and generation process.

5.2 Section A: Identify your 5 top primary objectives of good Data Management

The majority of the objectives can be grouped in 6 categories:

- Metadata, provenance and workflow documentation;
- Secure long term preservation;
- Data curation and data management governance;
- Ease of discovery and access;
- Interoperability and standardization;

- Data quality and quality control documentation.

In addition to the major recurring categories, statements also referred to other issues (e.g. availability of software to visualize and process data).

People were asked to answer this question before reading the draft GEO Data Management Principles, and while we have no control as to whether they actually did this, it is reassuring that the major recurring categories are also included in the draft GEO Data Management Principles.

5.3 Section B: Specific Questions

Are the proposed principles in line with the principles of your community / organization?

All responses noted that the draft GEO Data Management Principles are in line with their own (between 'mainly / in principle' and 'well / fully'). Responders acknowledged however that issues with implementing the principles remain, mainly with regard to resources.

Answers also indicate a good agreement of the draft GEO Principles with other organizations principles (some references are provided). As already noted above, there may be the opportunity for GEO to either join or coordinate with existing initiatives, or even to take a role in an across-the-board coordination effort.

Specific issues raised in response to this question include the need to clarify what is meant by data quality (GEO data management principle #7), and that embargo periods on data are a fact in the academic domain that will have to be respected.

Would the proposed principles be agreeable / applicable in your community?

Again the general response to this question is positive, with some organizations indicating that more thought is needed and/or details need to be reviewed (data quality). EUDAT indicates that they would be happy to adapt / support the GEO Data Management Principles rather than their own (indirectly this attitude can also be seen in other responses). One noteworthy remark reflects on the limited means of the data custodian to enforce or guarantee some of the principles, for which compliance is largely on the side of the data producer.

Is something missing, and if so what?

11 out of 26 responses find nothing (major) missing. Many of the comments to this question go already towards implementation issues. It is obvious that the border between formulation of a general principle and more detailed guidelines on how to accomplish them is not well defined. While the DMP-TF still believes in the value of a rather concise formulation of principles, extended versions of the principles should be included in any implementation guidelines (or even provided as a separate document). Specific comments mention the following issues: Data management to be aligned with national priorities and relevant international practices; data management to follow agreed standards; traceability; explicit inclusion of non-digital data; licensing and terms of use. Also, consideration of the strategic long-term goals for a data set, thereby providing the guiding principles for the data management, which might vary data set to data set.

Is the formulation of the principles clear (enough)?

The majority of responses answers yes to this question. Some suggestions where to include more detail or be more specific are provided. Again comments go towards implementation (ensure that principles are actionable, provide examples how to implement). Specific comments on formulations were considered in the revised version of the Data Management Principles proposed in this document.

Any other comments or suggestions?

In response to this question, some general issues are taken up again: there are various similar efforts ongoing in different organisms / communities, and coordination or participation should be considered;

importance of the governance of the data management process; implementation needs resources; compliance criteria should be provided.

One response emphasizes the need for training and capacity building for data management, a statement that the DMP-TF fully endorses.

5.4 Section C: Implementation and feasibility rating

A majority of responses assign high importance to most of the Principles, almost no response assigns low importance to any of Principles. A significant number of responses assign medium importance to Principles 7 (data quality control), 9 (verification), and 10 (corrections and updates).

The responses to this rating are summarized in the table below (detailed comments are available in Appendix 6).

Principle	Importance			Feasibility		
	L	M	H	L	M	H
1	0%	17%	83%	0%	29%	71%
2	0%	21%	79%	17%	48%	35%
3	0%	22%	78%	9%	73%	18%
4	0%	26%	74%	9%	57%	35%
5	4%	13%	83%	17%	61%	22%
6	4%	21%	75%	13%	65%	22%
7	0%	29%	71%	26%	57%	17%
8	4%	13%	83%	22%	30%	48%
9	13%	30%	57%	23%	45%	32%
10	13%	30%	57%	21%	58%	21%
11	5%	27%	68%	9%	52%	39%

With regard to feasibility of implementation, the clear majority of responses indicates 'medium' (no principle obstacles, but requires effort and may have resource issues) for all Principles except 1 (discoverability) and 8 (preservation), rated as 'high'. Also most of the remarks received to this question concern the feasibility (resources, political and legal issues, identifying responsibilities).

6 IMPLEMENTATION STRATEGY DIRECTIONS

The TF has not been mandated to work on possible implementation of the proposed GEO Data Management principles.

In fact, according to the TOR *"Pending GEO-XI Plenary's approval of the GEOSS Data Management Principles, DMP-TF Terms of Reference may be extended and modified to include a mandate to develop Implementation Guidelines for GEOSS Data Management Principles"*, a revision of the mandate of the TF is needed if and when the Principles will be endorsed by the Plenary.

In reality, it became clear during the consultation with the Executive Committee and external organizations/initiatives that it could be difficult to adopt principles in absence of preliminary orientations and clarification about their implementation.

Appendix 6 already includes a list of interesting comments related to the feasibility to implement proposed principles. Recurring items are related to resources needed for a proper life cycle data management, the diversity and heterogeneity of organizational structures and practices (and the need to respect them), the need to assess cost/benefit considerations and so on.

In addition, the process to define the new GEO implementation plan is ongoing and the IPWG is still largely debating if GEOSS should be or not a robust information systems (as envisaged by the vision paper adopted at the last plenary). For the TF, such a system requires to pay attention to the content and for this reason we strongly believe that a progressive harmonization toward professional data management practices will be essential. In particular if GEO is to engage with the Private Sector in the near future, some guarantees should be provided about the real "usability" of data made available through GEOSS.

Also the evolution of the work on Data Sharing Principles should be considered. In fact, some of the proposed Data Management Principles partially overlap (e.g. data accessibility). The TF recommends that a full convergence between the two groups be established after the Plenary (e.g. joint working group to further elaborate the few principles of common interest).

The TF has not formally discussed about implementation guidelines but some "principles" already emerged when discussing the comments received. These principles could be the initial basis to further develop preliminary guidelines for implementation:

- Data management principles are not legally binding. The benefits in investing in related practices should be clearly communicated. They will strongly depend on the future of GEOSS and the real intention to enlarge user community and usability of the entire system (e.g. a lot of freely accessible data of very low quality will not necessary increase the user basis);
- When data management practices already exist, they should be respected. GEO should assess if the practices cover all principles and if not suggest corrective actions;
- Adoption and compliance with the DM principles may be gradual. Data providers should have the opportunity to decide at which level they can afford the costs to be fully compliant;
- For facilitating usability of the GEOSS and to avoid possible dissatisfaction about GEOSS data, it will be extremely important for data providers to declare the level of compliance with proposed principles (e.g. as done for Data Sharing with GEOSS DataCORE, different labels can be introduced in GEOSS to inform users about the level of compliance on DM practices);
- Some areas (e.g. information collected by citizen observatories) require further investigation to assess how they could be made compliant with proposed DMPs;
- Considering the large number of initiatives running (including existing "certification" schemas and new working groups in ISO), GEO should further reflect if it wants to play a leading role in this area acting as facilitator platform (as requested by several organizations in the consultation phase). In this case, the future discussion about DMP implementation should be enlarged also to communities not yet active in GEO.

7 NEXT STEPS AND GEO PLENARY DECISION

At the GEO Plenary, this report will be discussed.

We expect that GEO Members and Participating Organizations will launch an internal consultation to provide feedback on the proposed Principles and as well express their view about future work.

The TF has been set-up to work until the GEO-XI Plenary and does not cover implementation guidelines.

Following the 31st Executive Committee deliberation, the TF has been mandated to propose DMP for consultation only.

At a minimum, the TF should continue to operate to assess the comments received from the next Plenary on the proposed Principles and present a revised proposal (if appropriate) at the following Executive Committee.

At this stage, the TF can be disbanded or its TOR be revised to carry out future work.

The GEO Plenary should now decide what to do with the proposal submitted by the TF and recommend a way to address the recurring comments about implementation guidelines.

The Plenary could also make suggestions to streamline future work (including recommendation for collaboration with IPWG, DSWG). If a reflection on implementation is recommended, then the mandate and duration of the DMP-TF should be revised.

The GEO Plenary could also (i)advise about the role that GEO could play on a specific subject (e.g. leading role as facilitator platform); (ii) give clear instructions about dealing with organisations/initiatives not acting under the GEO umbrella; and (iii) recommend if and how the view (and requirements) of the private sector should be included.

APPENDIX 1

Data Management Principles Task Force
Terms of Reference

PURPOSE

The GEO Data Management Principles Task Force (hereafter referenced as "DMP-TF") will support the Group on Earth Observation (GEO) in the definition and endorsement of common GEOSS Data Management Principles.

The following principles have been initially drafted by the GEO Infrastructure Implementation Board. It will be the responsibility of the DMP-TF to review the principles, assess feasibility and eventually propose a new formulation:

- Ensure long time data preservation and distribution;
- Ensure the quality information of EO data;
- Answer the EO user need.

OBJECTIVES

The DMP TF is convened to work with GEO collectively, including the Implementation Boards, the Data Sharing Working Group (DSWG).

The work of the DMP-TF will concentrate on developing Principles for:

- Identifying measures to ensure that data are properly managed (including data citation), accessible, archived and long term preserved (when appropriate);
- Identifying measures to ensure that data are properly documented (metadata), quality controlled and quality assessed so facilitating access and re-use of information made available through the GEOSS Common Infrastructure (GCI);
- Facilitating the link between user demands and data availability (e.g. by identifying existing sources of requirements already approved by the relevant user community);
- Facilitating interoperability of GEOSS data resources by promoting a progressive harmonisation/standardisation of content (data models, thesauri, coding list) in order to facilitate their re-use at global or regional scales.

DMP-TF will also participate in GEO relevant events and will make a recommendation to GEO-XI Plenary regarding next steps (e.g. development of GEOSS Data Management Principles Implementation Guidelines).

MEMBERSHIP AND WORKING ARRANGEMENTS

- The DMP-TF was created by, acts on requests from, and reports to, the GEO Plenary;
- The DMP-TF will provide reports to, and request the guidance of, the Executive Committee, as required;

- The DMP-TF shall be composed of individuals nominated by GEO Members and Participating Organisations, in particular those with relevant expertise, with administrative support provided by the GEO Secretariat;
- Co-chairs of the DMP-TF shall be decided by the DMP-TF. There should be a minimum of two co- chairs with representation from both GEO Members and Participating Organizations;
- The DMP-TF will meet at such times and places as determined by its members and work mainly through teleconferences and e-mail;
- The DMP-TF will coordinate its work with that of the Implementation Boards and Data Sharing Working Group as appropriate;
- The DMP-TF will coordinate with other international and regional initiatives and organisations working on data management issues;
- Specific research work or investigations may be assigned by the DMP-TF to sub-teams of individual experts to address identified issues at the appropriate level of detail and professional expertise.

DURATION

- These Terms of Reference shall be effective from March 2014 through the GEO-XI Plenary in 2015;
- Pending GEO-XI Plenary's approval of the GEOSS Data Management Principles, DMP-TF Terms of Reference may be extended and modified to include a mandate to develop Implementation Guidelines for GEOSS Data Management Principles.

APPENDIX 2

Members of Data Management Principles Task Force

No	Name	GEO Affiliation	IIB or DSWG	email
1	Aboubakar Mambimba Ndjoungui (Mr)	Gabon	N/A	abmambimba@gmail.com
2	Florian Haslinger (Dr)	EPOS	N/A	florian.haslinger@sed.ethz.ch
3	Siri Jodha Khalsa (Dr)	IEEE	IIB	sjsk@nsidc.org
4	David Halpern (Prof)	COSPAR	DSWG	David.Halpern@jpl.nasa.gov
5	Alessandro Annoni (Dr)	EC/JRC	IIB	alessandro.annoni@jrc.ec.europa.eu
6	Ryosuke Shibasaki (Prof)	Japan	IIB	shiba@csis.u-tokyo.ac.jp
7	Simon Hodson (Dr)	ICSU (CODATA)	DSWG	execdir@codata.org
8	Mirko Albani (Mr)	ESA	IIB	Mirko.albani@esa.int
9	Ivan Deloatch (Mr)	US	IIB	Ideloatch@usgs.gov
10	Jeff de La Beaujardiere (Dr)	US	N/A	jeff.delabeaujardiere@noaa.gov
11	Garry Baker (Dr)	UK	N/A	grba@bgs.ac.uk
12	Françoise GENOVA (Ms)	France	N/A	francoise.genova@astro.unistra.fr
13	Richard Moreno (Mr)	CEOS	N/A	richard.moreno@cnes.fr
14	Tony Boston (Mr)	Australia	N/A	T.Boston@bom.gov.au
15	Peiliang Shi (Mr)	WMO	N/A	pshi@wmo.int
16	Sergio Ojeda[2] (Mr)	Mexico	N/A	Sergio.Ojeda@Inegi.Org.Mx
17	PARK Chan-Ho (Dr)	Republic of Korea	N/A	ddony@korea.kr

[2] Mr Sergio Ojeda replaced Mr. Reynaldo Mondragon as Mexico's representative due to change in work assignment.

18	CHOI Won Young (Dr)	Republic of Korea	N/A	nnlwchoi@korea.kr
19	KIM Youn-gi (Mr)	Republic of Korea	N/A	99kyg@korea.kr

GEO Secretariat: Osamu Ochiai, oochiai@geosec.org, Wenbo Chu, wchu@geosec.org

APPENDIX 3

Executive Committee Consultation on Open Issues Relevant for the Drafting of Data Management Principles

1 DRAFT DATA MANAGEMENT PRINCIPLES

The current version, presented at the 31st Executive Committee meeting, differs from the document submitted to Executive Committee because it reflects all recent developments.

Each Earth observation is unique because the observation occupies a specific location and time in an environment that is continuously changing. No two Earth environmental observations are the same, making each Earth observation an irreplaceable asset to understand the past, describe the present, and forecast the future of the global integrated Earth system.

The value of each Earth observation is maximized through data life-cycle management, articulated in five following foundational elements and twelve principles:

Discoverability

1. Data and all associated metadata will be discoverable through catalogues and search engines, and data access and use conditions 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.

Usability

4. Data will be structured on open-standards and encodings that are harmonized to the greatest extent possible given organizational needs and observing methods.
5. Data will be comprehensively documented, including all elements necessary to access, use, understand, and process, preferably via formal structured metadata based on international standards.
6. Data will include provenance metadata indicating the origin and processing history of raw observations and derived products.
7. Data will be quality-controlled and the results of quality control shall be indicated in metadata; data made available in advance of quality control will be flagged in metadata as unchecked.

Preservation

8. Data will be protected from loss and preserved for future use; preservation planning will be for the long term and include guidelines for loss prevention, retention schedules, and disposal or transfer procedures.
9. Data and associated metadata will be periodically verified to ensure integrity, authenticity and readability.

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 persistent, resolvable identifiers to enable documents to cite the data on which they are based and to enable data providers to receive acknowledgement of use of their data.

2 QUESTIONS FOR THE EXECUTIVE COMMITTEE

1. Data Management requires professional attitude and resources. The DMP-TF can be very prescriptive or propose a lighter approach. The DMP-TF believes that **GEO should not compromise on the level of ambition but can adapt to different situations in the implementation phase**. *Is this a correct approach or alternative solutions are recommended?*
2. Professional data management require resources. *At which level GEO can propose principles that could have a potential high financial impact?*
3. The level of ambition is tightly connected to the future vision of GEO. In particular a **"robust information system" requires to strictly respecting the data management principles** that will be proposed. *Is the Executive Committee able to provide a clear definition of what such system should be in 2025?*
4. The GEO community is heterogeneous. The Principles could be very challenging for some organisations. *The Executive Committee should advise the* DMP-TF **if a single approach should be adopted or if principles should be adapted to different categories.**
5. The TF noted an increasing importance of **Big Data** (from social media), crowdsourcing information, sensor web … that introduce new data stream in which **data collection is not strictly controlled.** These data also generate new demand for data archiving and preservation. *How to apply principles to data "not controlled" by GEO but of potential high interest for the SBAs?*

APPENDIX 4

External Consultation Document
On GEO Data Management Principles

1 WHAT IS GEO

GEO is an international partnership of governments and organizations established in 2005 with an initial ten-year mandate designed to link the various Earth Observation endeavors worldwide into a "Global Earth Observation System of Systems" (GEOSS). GEO fosters coordination of Earth Observations amongst its partners and seeks to promote free and open access to data and information. Through GEOSS, the GEO Members and Participating Organizations contribute "to realize a future wherein decisions and actions for the benefit of humankind are informed by coordinated, comprehensive, and sustained Earth observations and information."

2 WHY DATA MANAGEMENT PRINCIPLES (CALL FOR ACTION)

The Geneva Ministerial Summit in January 2014 renewed GEOs mandate for 2015-2025 and endorsed the next decade of activities in line with the GEO-X recommendations. The importance of life-cycle data management was specifically addressed by the GEO-X Plenary.

The Data Management Principles Task Force (DMP TF), composed of 19 individuals nominated by GEO Members and Participating Organizations, was established to draft GEO Data Management Principles covering the entire data life cycle from planning to acquisition, quality assurance, documentation, access, archiving, preservation and answering user needs.

The DMP TF has created a draft set of principles through reviewing, comparing and selecting best practice as appropriate from at least 12 sets of Data Management Principles adopted and in use with different initiatives and regions.

On 9 June 2014, the DMP TF reported to the 31st GEO Executive Committee and was requested to consult with related organizations both inside and outside the GEO community.

Please note that:

- The draft Data Management Principles are limited to *data management* and will not generally consider issues of *data sharing*, as GEO Data Sharing Principles are being updated and aligned by a separate working group of GEO. It is, however, acknowledged that data management and data sharing are closely connected, therefore synchronization of the differing sets of principles will be addressed after this external consultation phase;
- The draft Data Management Principles intentionally do not consider issues related to implementation of the principles. Development of an Implementation Plan is part of the scope of the DMP TF and will be addressed in the next stage.

3 QUESTIONS PART 1

(Please answer the following questions **BEFORE** reading the draft GEO DATA MANAGEMENT PRINCIPLES on the following page)

A. PRIMARY OBJECTIVES

In your opinion, which should be the primary objectives (top 5) of good Data Management?

(Consider in particular the context and mission of GEO, if possible from your background and expertise)

01_____

02_____

03_____

04_____

05_____

4 GEO DATA MANAGEMENT PRINCIPLES (DRAFT)

Each Earth observation is unique because the observation occupies a specific location and time in an environment that is continuously changing. No two Earth environmental observations are the same, making each Earth observation an irreplaceable asset to understand the past, describe the present, and forecast the future of the global integrated Earth system.

The value of each Earth observation is maximized through data life-cycle management, including the following foundational elements:

- Discoverability;
- Accessibility;
- Usability;
- Preservation;
- Curation.

Discoverability

1. Data and all associated metadata will be discoverable through catalogues and search engines, and data access and use conditions 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.

Usability

4. Data will be structured on open-standards and encodings that are harmonized to the greatest extent possible given organizational needs and observing methods.
5. Data will be comprehensively documented, including all elements necessary to access, use, understand, and process, preferably via formal structured metadata based on international standards.
6. Data will include provenance metadata indicating the origin and processing history of raw observations and derived products.
7. Data will be quality-controlled and the results of quality control shall be indicated in metadata; data made available in advance of quality control will be flagged in metadata as unchecked.

Preservation

8. Data will be protected from loss and preserved for future use; preservation planning will be for the long term and include guidelines for loss prevention, retention schedules, and disposal or transfer procedures.
9. Data and associated metadata will be periodically verified to ensure integrity, authenticity and readability.

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 persistent, resolvable identifiers to enable documents to cite the data on which they are based and to enable data providers to receive acknowledgement of use of their data.

5 QUESTIONS PART 2

B. SPECIFIC QUESTIONS
(please provide answers on a separate sheet)

- Are the presented principles in line with the principles of your community / your organizations (either formally established ones or informally agreed ones)? If not, please provide details on differences. If your community's principles are documented, please provide URL or citation.
- In the case that your community / your organization does not (yet) have fully established data management principles, do you think that these GEO principles are agreeable / applicable for you?
- Is something missing, and if so what?
- Is the formulation of the principles clear (enough)?
- Do you have any other comments or suggestions?

C. IMPORTANCE AND FEASIBILITY RATING

Please rate the importance of the specific data management principles (Low, Medium, High), and judge their feasibility of implementation / achievement both within GEO and within your own organisation given current funding and priorities (Low: foresee significant problems for implementation (political, resources, others); Medium: no principle obstacles, but will require effort and resources may not be available right away; High: clear path to implementation visible, manageable with current resources, successful examples may exist elsewhere).

Category/Principle	Importance			Feasibility			Remarks
	L	M	H	L	M	H	
Discoverability							
1. Data and all associated metadata will be discoverable through catalogues and search engines, and data access and use conditions 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.							
Usability							
4. Data will be structured on open-standards and encodings that are harmonized to the greatest extent possible given organizational needs and observing methods.							
5. Data will be comprehensively documented, including all elements necessary to access, use, understand, and process, preferably via formal structured metadata based on international standards.							
6. Data will include provenance metadata indicating the origin and processing history of raw observations and derived products.							
7. Data will be quality-controlled and the results of quality control shall be indicated in metadata; data made available in advance of quality control will be flagged in metadata as unchecked.							
Preservation							
8. Data will be protected from loss and preserved for future use; preservation planning will be for the long term and include guidelines for loss prevention, retention schedules, and disposal or transfer procedures.							
9. Data and associated metadata will be periodically verified to ensure integrity, authenticity and readability.							
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 persistent, resolvable identifiers to enable documents to cite the data on which they are based and to enable data providers to receive acknowledgement of use of their data.							

APPENDIX 5

Responses to the GEO DMP TF's Executive Committee Consultation

GABON RESPONSE

On the draft of the DMP, we should add a point on the data policy. This item entitled "Policy data" will have the following sub-items:

- Data under restricted use must be accompanied by a clear license;
- The license must be signed online via geoportals;
- Products of the images must be free of restrictions.

EC RESPONSE

GEO is based on voluntary contributions. A proper Data Management Framework requires both professionalism and resources to ensure quality data. The TF can be very prescriptive, or propose a lighter approach. Both have advantages and negative consequences. We believe that we should not compromise the level of ambition, but can adapt to different situations in the implementation phase. Is this a correct approach or alternative solutions are recommended?

> Data quality is one criterion only – what really matters is that the data quality is documented so that the users know what they have in their hands (sometimes it is better to have 1000 data points of average quality than one item of data of very high quality). So our view is that professionalism and resources are needed to ensure that quality data is properly documented. The level of ambition should take into account of the fact that GEO is not a standardization body, but a voluntary initiative which does not implement data policies, but can actively and widely promote generic principles which apply globally and concern many different data situations. In this respect, what happened with the GEOSS data sharing principles should inspire what could be done with the future GEOSS data management principles.

The level of ambition is tightly connected to the future vision of GEO. In particular the word "robust information system" requires strict adherences to data management principles that will be proposed. If the Executive Committee has any specific recommendations, please identify them.

> The terms "Robust Information System" was introduced by the GEO post-2015 Working Group. The Commission understands that robust means in this context an operational GEO information system (GCI) enabling a variety of GEOSS users to upload and download Earth Observation data, that have been pledged by the GEO Community in a reliable way. GEOSS is characterized by a hybrid nature composed of the GCI belonging to GEO and the systems harvested by the GCI. From case to case, the latter systems might belong to a GEO Member, a GEO Participating Organisation, or an organization (public or private) not belonging formally to the GEO community. A differentiated approach should probably be put in place with respect to the expected degree of implementation of the GEOSS Data Management Principles which would take into consideration of this hybrid nature of the GEOSS. The future GCI should ideally be sustained on the long term, and should be able to evolve taking into account latest technology developments. While some minimum commitment by the GEO community with respect to the GCI should ideally be obtained, we would recommend a ramping up process where the distributed systems harvested by/connected to the GCI could evolve over the time. Future implementation of the DMPs should allow for a

progressive behavior of the Data providers towards higher degrees of fulfillment of the future DMPs (otherwise we never start).

The GEO community is quite heterogeneous, so principles that can be easily respected by some communities of practice (space agencies, meteorological services, etc.) could be very challenging for other organisations. The Executive Committee could advice the TF if a single approach should be adopted or if principles should be adapted to different categories;

There is probably no other ways than adapting to the different categories. The proposed principles could probably be agreed by different communities but implemented to various extents and degrees of fulfillment.

The TF noted an increasing importance of Big Data (from social media), crowdsourcing information, sensor web, etc. that introduce new data streams in which data collection is not strictly controlled. These data also generate new demand for data archiving and preservation. Should EO registered resources only or data "not controlled" by GEO, but of potential high interest for the SBAs be included?

The definition of Big Data is a bit artificial and would require some semantic analysis. Nonetheless more data are generated every day and it is generally well accepted that the term “big data” captures this reality. It is not possible to exclude any data sources in GEOSS as long as they qualify as Earth observations. Data that are documented regarding their quality should probably get higher visibility in the GCI, possibly through ranking or proper tagging (new data management tag(s)).

Professional data management requires resources. If the adoption of the Data Sharing Principles (and GEOSS DataCORE in particular) only speaks about lost revenues (for data not previously accessible) or vice versa a professional life cycle data management will have additional costs (if not already in place). At which level should GEO propose principles that could have a potential financial impact?

The cost for the implementation of management principles is probably proportional to the level of details of those principles. It would be difficult for an initiative as GEO to develop very detailed and prescriptive principles that would be valid across a highly variable landscape of datasets. It is proposed that GEO develops broad management principles that can be implemented a reasonable costs. Future GEOSS DMPs will not be compulsory but will have to be adopted and implemented on a voluntary basis. GEO should not be forcing anyone unwilling to bear very high costs. But if we target a GEOSS delivering data with good quality documentation, we have to benchmark and promote best practices towards the data providers. Those of them who apply the DMP at best should be rewarded somehow through better visibility of their data within the GEOSS.

APPENDIX 6

Responses to the GEO DMP TF's External Consultation

Please see the responses at the following link:

ftp://ftp.earthobservations.org/GEO-XI/Appendix%206%20to%20Doc%207.pdf