

3rd GEO ●●●  
**DATA**  
**PROVIDERS**  
WORKSHOP

**DATA PROVIDERS  
MEET USERS**

**FRASCATI, ITALY  
2-4 MAY 2018**



# GEOSS Data Management Principles: Importance and Implementation

Alex de Sherbinin / Associate Director / CIESIN, Columbia University

Gregory Giuliani / Lecturer / University of Geneva

Joan Maso / Researcher and GIS developer / CREA



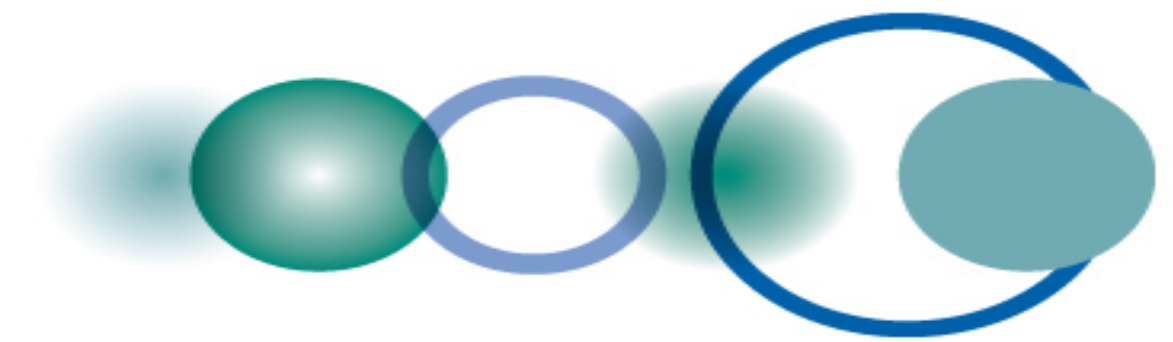
# Why?

*“Ensure that data and information of **different origin and type** are comparable and compatible, facilitating their **integration into models** and the development of applications to derive **decision support tools**.”*

GEO Strategic Plan 2016-2025

- Interoperability
- Quality (Fitness for use)
- Trustworthiness





## **GEOSS Data Management Principles (DMPs)**

- Build on GEOSS Data Sharing Principles
- Set standard for **good data (management) practices**

### **GEOSS DMPs Approved in April 2015:**

- Long version
- Condensed version

### **GEOSS DMPs Implementation Guidelines**

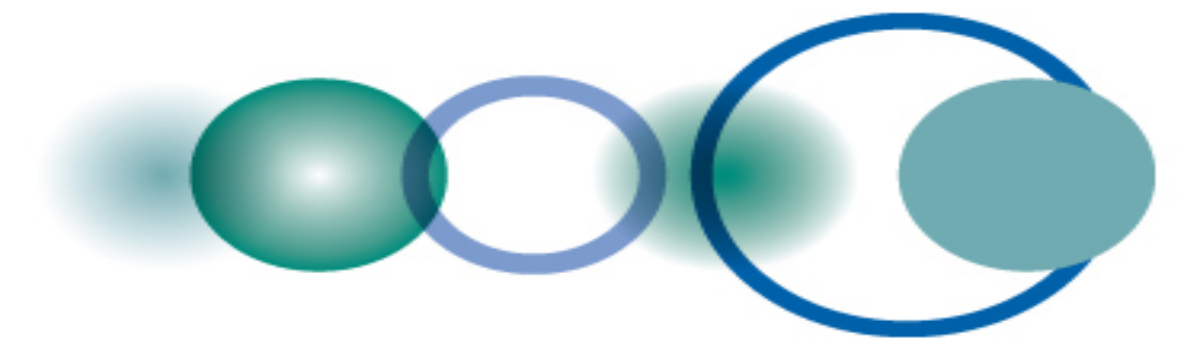
Endorsed in November 2015



# **GEOSS DMPs**

## **10 Principles in 5 categories**

1. Discoverability
2. Accessibility
3. Usability
4. Preservation
5. Curation



## GEOSS DMPs condensed

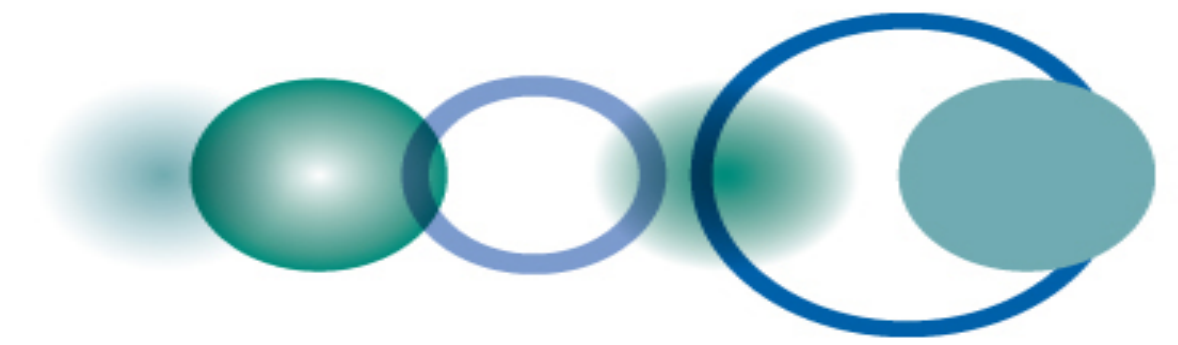
**Earth observations** will be **catalogued** or otherwise advertised on the internet so that they can be **discovered**, and will be **accessible** online using **open-standard encodings and services**. Data and services will be comprehensively **documented** using international or community-approved standards, and to the extent possible, peer-reviewed publications, so that users can understand and make use of the data. Metadata will include **access and use conditions**, the **results of quality control procedures**, and **provenance statements** indicating the origin and processing history of the dataset or product. Data and associated metadata will be **protected from loss** and periodically **verified to ensure integrity, authenticity and readability**. **Corrections and updates** to data and metadata records will be performed as required. Finally, **persistent identifiers** will be assigned to data so that they can be tracked and cited and data providers can be acknowledged.



## Implementation of DMPs

*“A **priority** mission for GEO is to encourage the implementation of the Principles (DMPs) by organizations contributing to GEOSs”*

*GEO Strategic Plan 2016-2025*



# **GEOSS DMPs Implementation Guidelines**

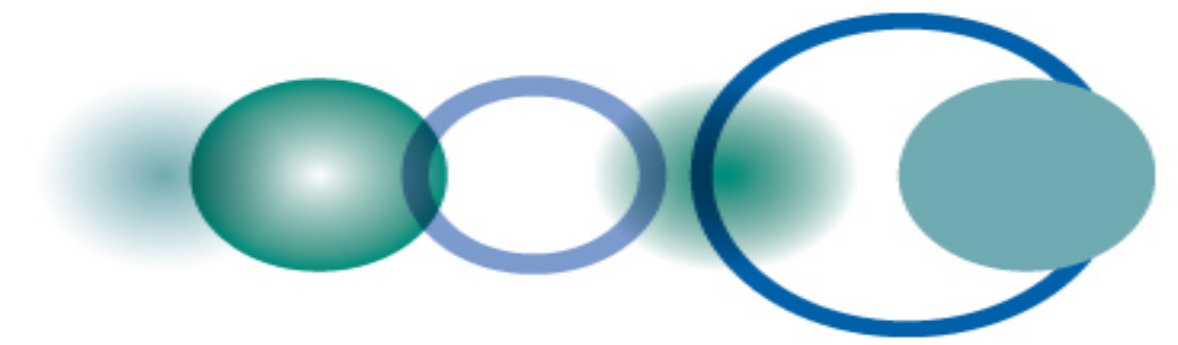
For each DMP:

- Terms used to describe the principle and its implementation
- Explanation of the DMP
- Guidance on implementation with examples

**November 2015**

[www.earthobservations.org/documents/geo\\_xii/GEO-XII\\_10\\_Data%20Management%20Principles%20Implementation%20Guidelines.pdf](http://www.earthobservations.org/documents/geo_xii/GEO-XII_10_Data%20Management%20Principles%20Implementation%20Guidelines.pdf)





# Implementation guidelines uptake?

## Monitoring DMPs implementation

### 1. GEOS Data Providers

- a) Data Repositories Certification
- b) Status checker...

### 2. Dataset/Collection

- a) Data fitness for use: Certification, DMP Labels...
- b) User feedback...





# From Principles to Implementation

DOI 10.5281/zenodo.34354

**WORLD DATA SYSTEM**

## WDS Data Sharing Principles

### Purpose

The International Council for Science – [World Data System](#) (ICSU-WDS) aims to provide equitable access to quality-assured scientific data, data services, products and information towards long term data stewardship. Furthermore, ICSU-WDS is committed to foster agreed-upon data standards and conventions, and providing mechanisms to facilitate access to data. As the leading international, multidisciplinary organization in the provision of data, ICSU-WDS has adopted Data Sharing Principles to advance its goals.

The Principles express core ethical commitments that are operationalized in [WDS Certification](#) and Network Members, it is anticipated that existing organizational policies align with them, but are encouraged to do so. The Principles embody the spirit of 'open science' for diverse communities of data producers and data users, and thus could be adopted for science for the public good.

2009-15



**e-INFRASTRUCTURE DATA MANAGEMENT PROJECT** **BELMONT FORUM**

Home About Activities Documents Announcements

Activities » Data Principles

## Data Principles

The following data principles are recommended for adoption by the Belmont Forum e-Infrastructure and Data Management CRA.

Data should be:

- **Discoverable** through catalogues and search engines, with data access and use conditions, including licenses, clearly indicated. Data should have appropriate persistent, unique and resolvable identifiers.
- **Accessible** by default, and made available with minimum time delay, except where international and national policies or legislation preclude the sharing of data as Open Data. Data sources should always be cited.
- **Understandable and interoperable** in a way that allows researchers, including those outside the discipline of origin, to use them. Preference should be given to non-proprietary international and community standards via data e-infrastructures that facilitate access, use and interpretation of data. Data must also be reusable and thus require proper contextual information and metadata, including provenance, quality and uncertainty indicators. Provision should be made for multiple languages.
- **Manageable** and protected from loss for future use in sustainable, trustworthy repositories with data management policies and plans for all data at the project and institutional levels. Metrics should be exploited to facilitate the ability to measure return on investment, and can be used to implement incentive schemes for researchers, as well as provide measures of data quality.
- **Supported** by a highly skilled workforce and a broad-based training and education curriculum as an integral part of research programs.

2015

**F**indable **A**ccessible **I**nteroperable **R**eusable

DOI: 10.1038/sdata.2016.18

www.nature.com/scientificdata

## SCIENTIFIC DATA

**OPEN**

SUBJECT CATEGORIES  
» Research data  
» Publication characteristics

### Comment: The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson *et al.*<sup>#</sup>

Received: 10 December 2015  
Accepted: 12 February 2016  
Published: 15 March 2016

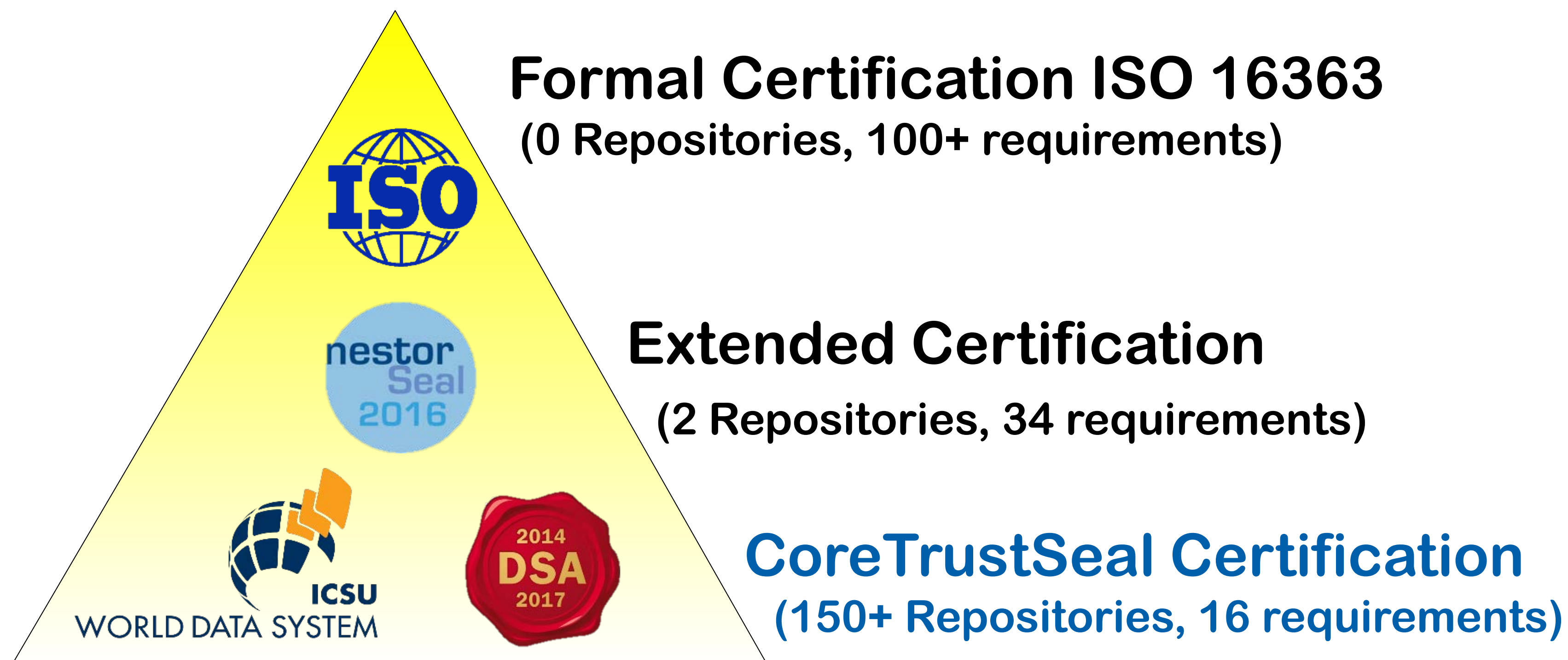
There is an urgent need to improve the infrastructure supporting the reuse of scholarly data. A diverse set of stakeholders—representing academia, industry, funding agencies, and scholarly publishers—have come together to design and jointly endorse a concise and measurable set of principles that we refer to as the FAIR Data Principles. The intent is that these may act as a guideline for those wishing to enhance the reusability of their data holdings. Distinct from peer initiatives that focus on the human scholar, the FAIR Principles put specific emphasis on enhancing the ability of machines to automatically find and use the data, in addition to supporting its reuse by individuals. This Comment is the first formal publication of the FAIR Principles, and includes the rationale behind them, and some exemplar implementations in the community.

2016





## Practical Implementation



**Trustworthy Data  
Repositories (TDRs)**





# Core Trustworthy Data Repository (TDR) certification

## Catalogue of requirements (16): Context

1. Organizational infrastructure (6)
2. Digital object management (8)
3. Technology (2)

## Applicant feedback

## Certification procedures :

1. Self-assessment: documented  
public evidence + compliance level
2. Peer-review (2 reviewers)
3. Renewal (3 years)

DOI 10.5281/zenodo.168411



2009-15

2017-forward







# CORE TDR Requirements

## Context

Repository type, designated community, Level of curation performed ,  
Outsource partners, Impact

## Organizational infrastructure

- **R1.** DR has **an explicit mission** to provide **access to and preserve data** in its domain.
- **R2.** DR maintains all applicable **licenses** covering data access & use & monitors compliance
- **R3.** DR has a **continuity plan** to ensure ongoing access to and preservation of its holdings.
- **R4.** DR ensures that data are created, curated, accessed, and used in compliance with **disciplinary and ethical norms**.
- **R5.** DR has **adequate funding** and sufficient numbers of **qualified staff** managed through a clear system of governance
- **R6.** repository adopts mechanism(s) to secure ongoing **expert guidance** and feedback (either in-house, or external, scientific guidance).





# CORE TDR Requirements

## Digital object management

- R7. DR guarantees the **integrity and authenticity** of the data.
- R8. DR accepts data & metadata based on **defined criteria to ensure relevance and understandability** for users.
- R9. DR applies **documented processes and procedures** in managing archival storage of the data.
- R10. DR assumes responsibility for **long-term preservation** and manages this function in a planned and documented way.
- R11. DR has appropriate expertise to address **technical data and metadata quality** sufficient to make quality evaluations.
- R12. Archiving takes place according to **defined workflows** from ingest to dissemination.
- R13. DR enables users to **discover the data and refer to them in a persistent way** through proper citation.
- R14. DR enables reuse of the data over time, ensuring that **appropriate metadata** support the understanding and use of the data.







# CORE TDR Requirements

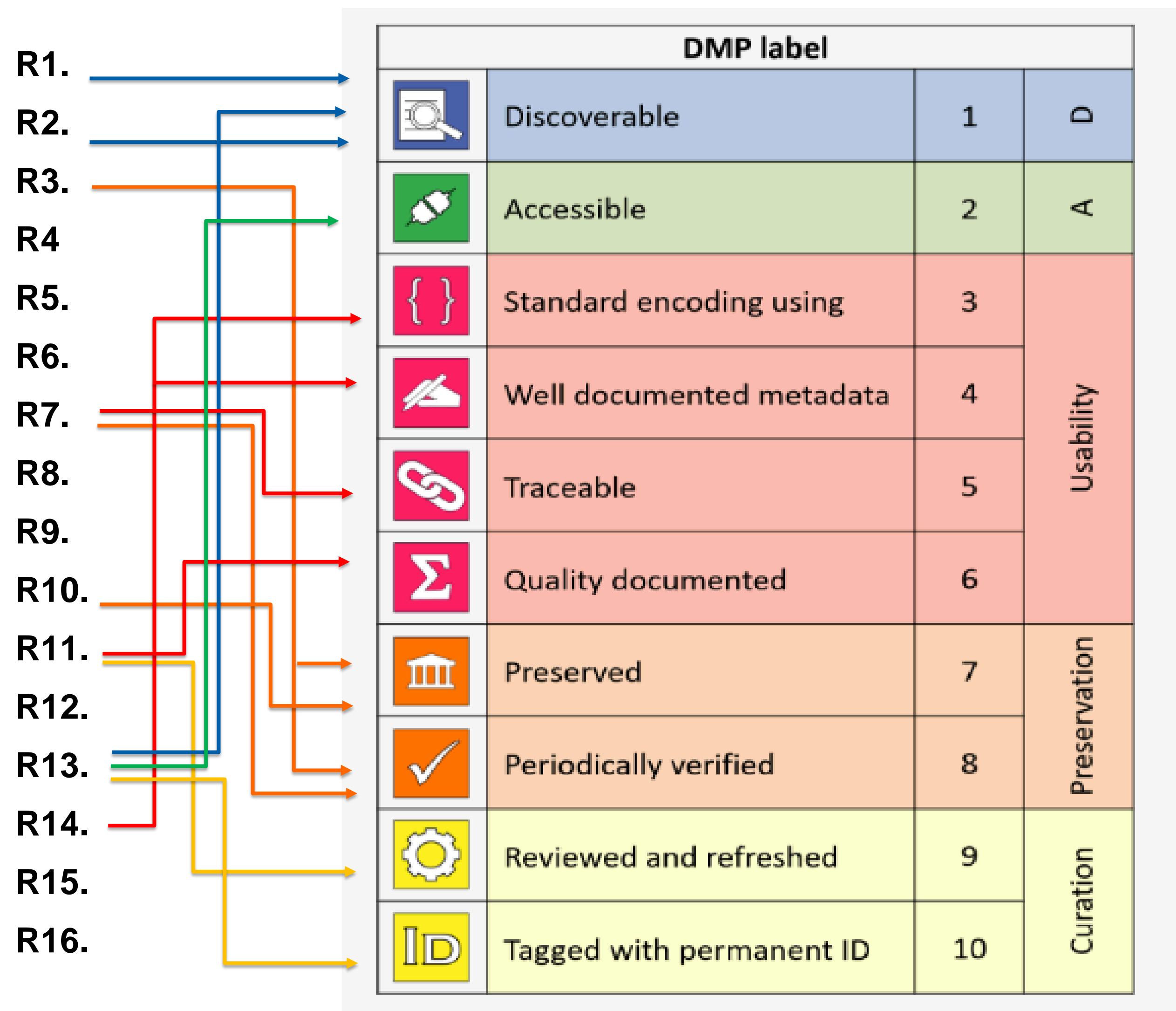
## Technical infrastructure

- **R15.** DR functions on **well-supported operating systems and other core infrastructural software** and is using hardware and software technologies appropriate to the services it provides to its Designated Community.
- **R16.** The technical infrastructure of the repository provides for **protection** of the facility and its data, products, services, and users.

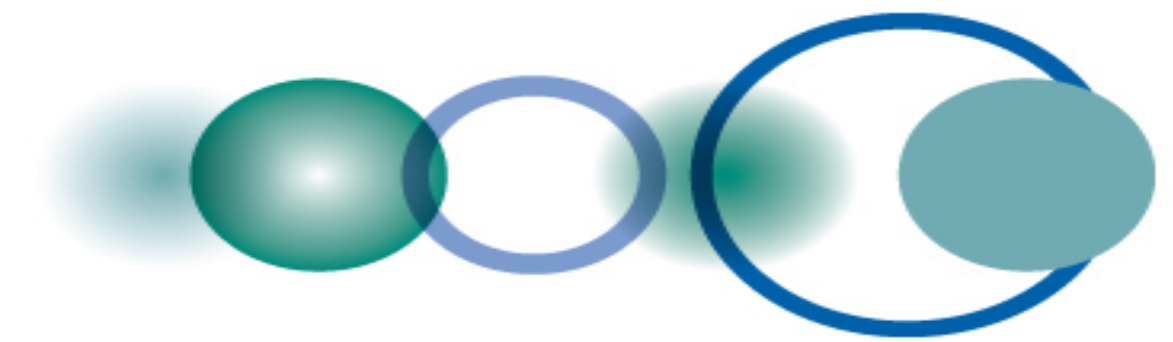




# Mapping CORE TDR to GEOSS DMPs

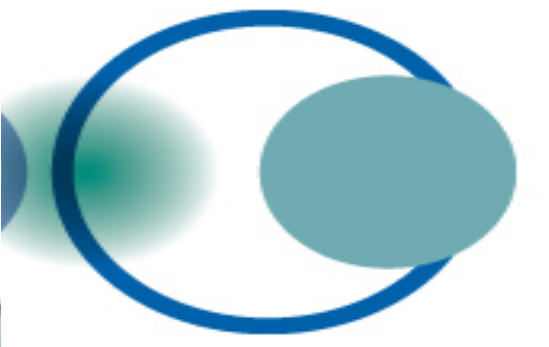






# Myths about Core TDR Certification

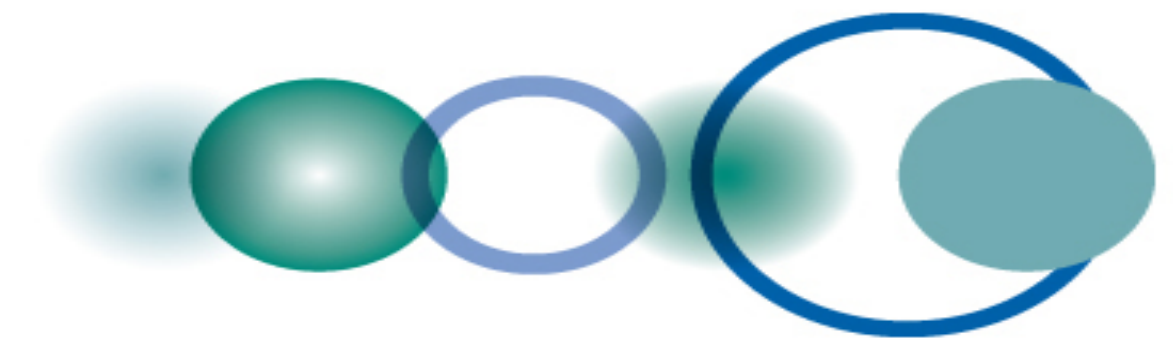
- Costly !
- Pass or fail !
- Difficult for small data providers...
- Time consuming!
- Not yet recognized?



**Thank you!**







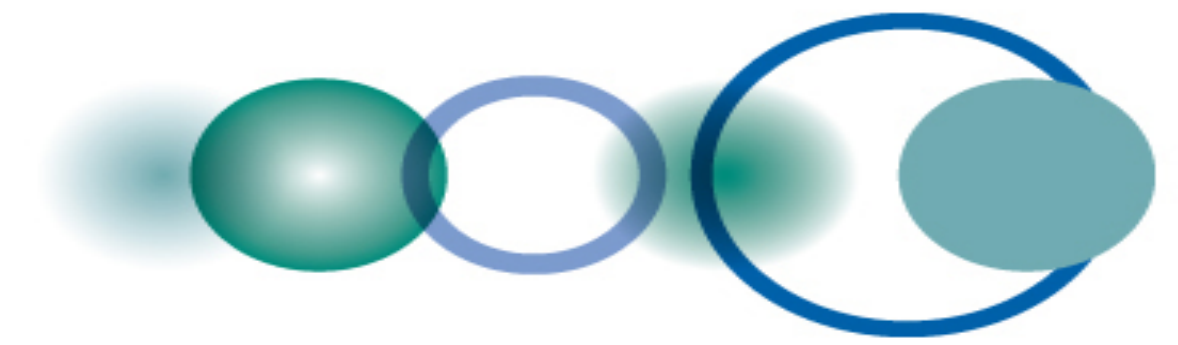
## GEOSS DMPs in detail

### Discoverability

- DMP-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

- DMP-2: Data will be **accessible via online services**, including, at minimum, direct download but preferably user-customizable services for visualization and computation.

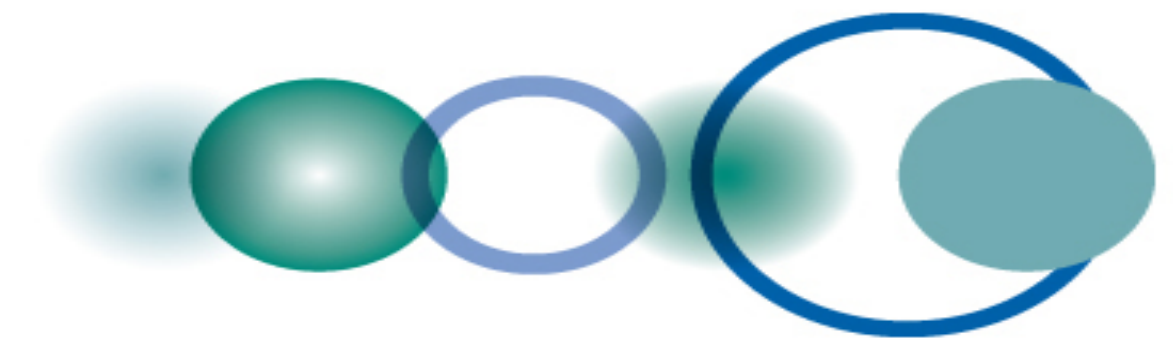


# GEOS DMPs in detail

## Usability

- DMP-3: 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.
- DMP-4: Data will be **comprehensively documented**, including all elements necessary to access, use, understand, and process, preferably via formal structured metadata based on international or community-approved standards. To the extent possible, data will also be described in peer-reviewed publications referenced in the metadata record.

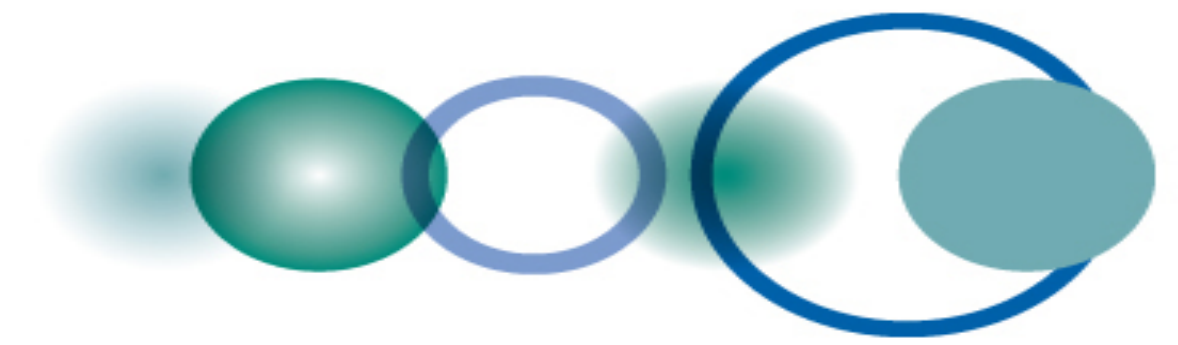




## **GEOSS DMPs in detail**

### **Usability (ctd.)**

- DMP-5: 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.
- DMP-6: 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. of their data.

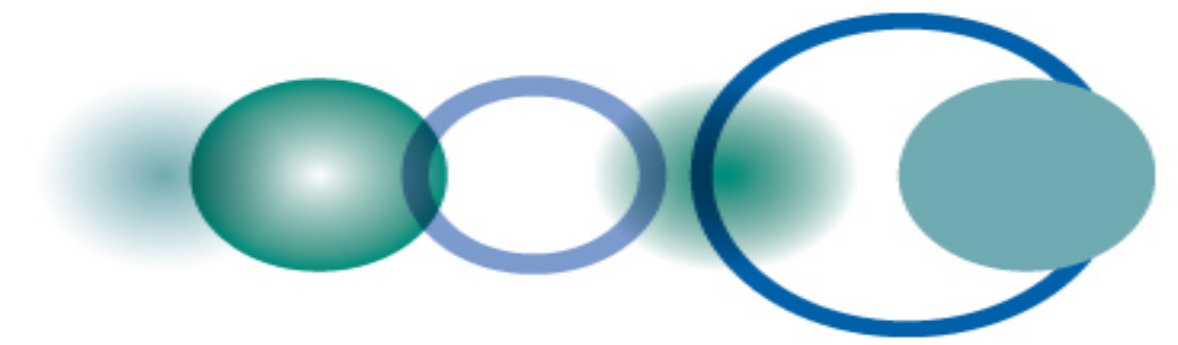


# **GEOSS DMPs in detail**

## **Preservation**

- DMP-7: 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.
- DMP-8: Data and associated metadata held in data management systems will be periodically verified to ensure integrity, authenticity and readability.





# **GEOSS DMPs in detail**

## **Curation**

- DMP-9: 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.
- DMP-10: 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.



## Contact information



DATA PROVIDERS  
MEET USERS

FRASCATI, ITALY  
2-4 MAY 2018

**Alex de Sherbinin** [adesherbinin@ciesin.columbia.edu](mailto:adesherbinin@ciesin.columbia.edu)  
**Gregory Giuliani** [gregory.giuliani@unige.ch](mailto:gregory.giuliani@unige.ch)  
**Joan Maso** [joan.maso@uab.cat](mailto:joan.maso@uab.cat)



**#GEODATA18**  
[bit.ly/GEodataworkshop](http://bit.ly/GEodataworkshop)



3rd GEO  
**DATA**  
PROVIDERS  
WORKSHOP

DATA PROVIDERS  
MEET USERS

FRASCATI, ITALY  
2-4 MAY 2018

making data work for our planet

**GEOSS / HACK** 2018

powered by **GEOSS**