EC GEO European Projects Special Biodiversity Workshop
Brussels (Italy), 28th September 2015

Co-funded by
the European Union

ECOPOTENTIAL:
Data portals, community portals, platforms, and GEOSS/GCI

Stefano Nativi, Paolo Mazzetti

(CNR-IIA)

CNR-IIA
ESSI Lab
Florence Division

ECOPOTENTIAL
Virtual Laboratory Platform

- Make **data, scientific results, models and information accessible and available** through a cloud-based open platform implementing virtual laboratories.
- The platform will be a **major contribution to the GEOSS Common Infrastructure**, reinforcing the GEOSS Data-CORE.
- By the end of the project, new prototype products and ecosystem services, based on **improved access (notably via GEOSS) and long-term storage of ecosystem EO data and information** in existing PAs, will be made available.

stefano.nativi@cnr.it

Reference Material

Setting the scene

stefano.nativi@cnr.it

GEOSS Data Sharing Principles

GROUP ON EARTH OBSERVATIONS

DSP1. There will be **full and open exchange** of data, metadata and products shared within GEOSS, recognizing relevant international instruments and national policies and legislation;

DSP2. All shared data, metadata and products will be made available with **minimum time delay** and at **minimum cost**;

DSP3. All shared data, metadata and products being **free of charge or no more than cost of reproduction** will be encouraged for research and education.

stefano.nativi@cnr.it

GEOSS Data Management Principles

GROUP ON EARTH OBSERVATIONS

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.

Usability

DMP-3. Data will 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.

stefano.nativi@cnr.it

GEOSS Data Management Principles

GROUP ON EARTH OBSERVATIONS

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.

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.

stefano.nativi@cnr.it

GEOSS Data Management Principles

GROUP ON EARTH OBSERVATIONS

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

stefano.nativi@cnr.it

GEOSS Architectural Principles

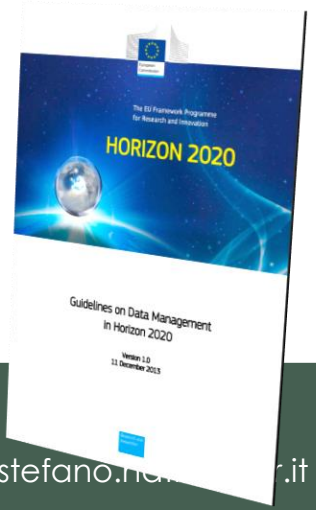
GROUP ON EARTH OBSERVATIONS

- Given the nature of a "system of Systems" it was recognized that the **success would depend on building interoperability among the different and autonomous systems**
- As the basis for evolution and ensure interoperability with relevant research and policy-driven data infrastructures
  - **Openness**
  - **Effectiveness**
  - **Flexibility**
  - **Sustainability**
  - **Reliability**
- Support the implementation of the **Data Sharing and Management principles**

stefano.nativi@cnr.it

EC H2020 Open Data Access (Data Management Plan) Guidelines

- Any **Research Data** and the **associated Software must be**:
  - **Discoverable**
    - readily **located**
    - standard **identification** mechanism
  - **Accessible**
    - **modalities**, **scope**, **licenses**
  - **Assessable and intelligible**
    - scientific **scrutiny**
    - peer **review**
  - **Useable beyond the original purpose for which it was collected**
    - safely stored in **certified repositories**
    - **long term preservation** and **curation**
    - **minimum software**, **metadata** and **documentation** to make it useful
  - **Interoperable to specific quality standards**
    - standards for **data annotation** and **data exchange**
    - **recombinations with different datasets** from different origins



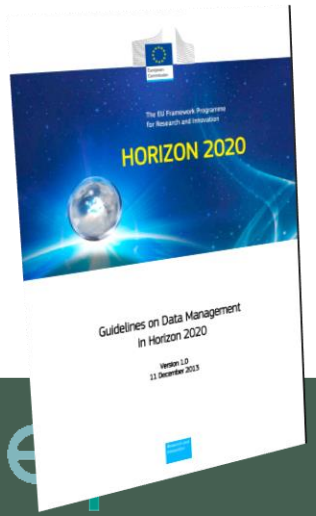
Grant Agreement (model)

- Article 29.3 of the GA model
  - *Regarding the **digital research data generated** in the action ('data'), the **beneficiaries must**:*

    *(a) **deposit in a research data repository** and take measures to **make it possible** for third parties to **access**, **mine**, **exploit**, **reproduce** and **disseminate** — **free of charge** for any user — the following:*

    *(i) the **data**, including **associated metadata**, needed **to validate** the results presented in **scientific publications as soon as possible**;*

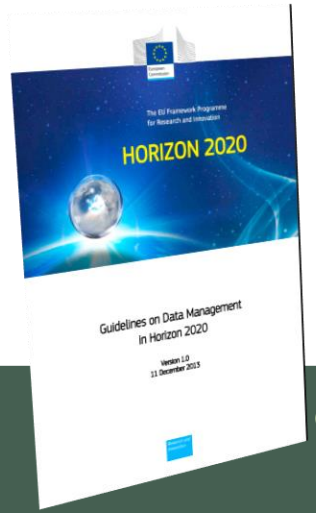
    *(ii) **other data**, including **associated metadata**, as specified and **within the deadlines laid down in the data management plan** (see Annex I);*



stefano.nativi@cnr.it

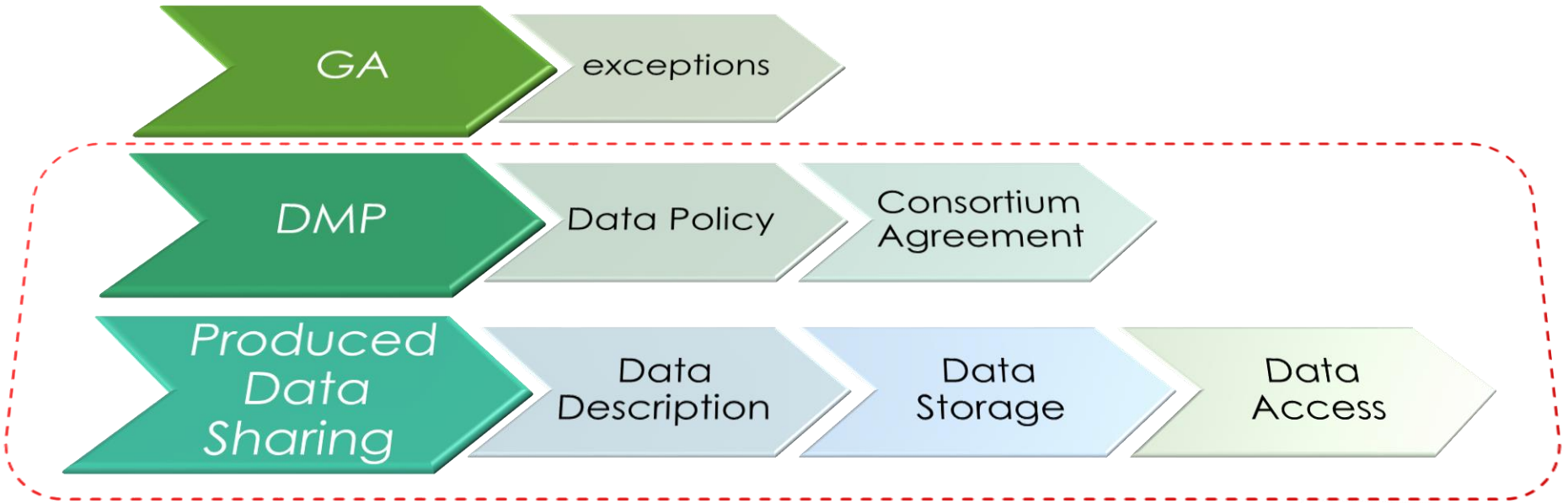
GA: exceptions….

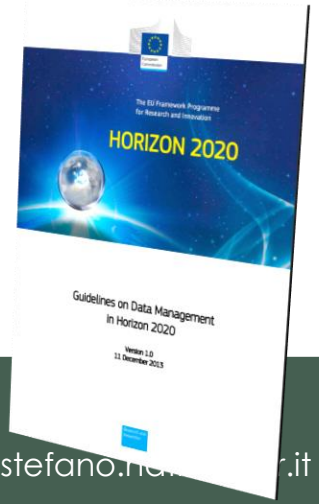
- *As an exception, the beneficiaries **do not have to ensure open access to specific parts of their research data** if the **achievement of the action's main objective**, as described in Annex I, would be **jeopardised by making those specific parts of the research data openly accessible**.*
- *In this case, the **data management plan must contain the reasons for not giving access**.*



stefano.nativi@cnr.it

Approach and actions





ECOPOTENTIAL PRINCIPLES

stefano.nativi@cnr.it

ECOPOTENTIAL Interoperability (e-infrastructure) Principles

1. To build on ECOPOTENTIAL existing and under development digital systems
   - Noticeably, digital systems managed in the project (EO data, EO products, in-situ data,...)
2. Not to impose any "common solution/specification" but advocate the use of open standards and interoperability APIs
3. To provide a common, consistent, and "high-level" entry point ECOPOTENTIAL platform for discovering, accessing, and using ECOPOTENTIAL ecosystem services
   - interoperability to GEOSS, Copernicus, and other EC-funded programmes
4. To adopt/implement the GEOSS Architecture Principles
5. To adopt/implement the GEOSS "resource sharing" and "resource management" principles –including quality and preservation
6. To adopt/implement the EC Open Data Access principles

stefano.nativi@cnr.it

Services protocol and APIs

To support interoperability and contribute to the ECOPOTENTIAL Virtual Research Laboratory (Eco-VRL)

stefano.nativi@cnr.it

App (experts)

App (activist)

. . .

App (Decision Makers)

. . .

App (teachers)

APIs

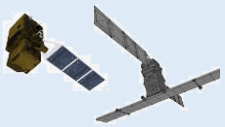
Interoperability

Knowledge management (including Ecosystem Modeling)

Mediation, Harmonization (including uncertainty), workflow (brokering)

Data

EO Data



Monitoring Data

Other Resources

Scientific Processes

Knowledge

Models

ECOPOTENTIAL Virtual Laboratory : ECOPOTENTIAL Platform

ECOPOTENTIAL Platform Service Layers

1. Discovery, access, and harmonized use of:
   - open **data** –both remote and *in-situ*
   - **semantic** services
   - **analytics** capacities
   - scientific **models** and results –including ecosystems m… scenarios
2. Provision of the VRE for generating and … knowledge for the Ecosystem services
   - analyze **ecosy…**
   - … and land-use **change scenarios**
   - **u…scaling** to larger areas
   - definition of the requirements of **future protected areas**
3. Provision of specific applicati… tals for different users
   - experts, activists, decisions m…kers
   - They will be defined in the project addressing their specifi… -cases
4. Provision of specific platform and (open) APIs to int… with GEOSS, Copernicus, INSPIRE, ESFRI, etc.

Data, Services

Knowledge bodies, workflows

APPs

APIs

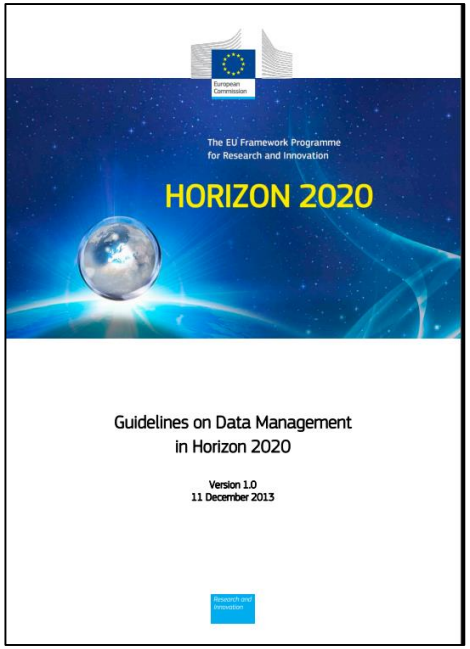
stefano.nativi@cnr.it

ECOPOTENTIAL Data Management Plan (DMP)

- Mandatory
- First Draft by Month 6
- To be kept Updated for the entire duration of the project
- Part of the annual and final Review

stefano.nativi@cnr.it

DMP Guidelines (Reference Material)



Data Documentation

- **Data set reference and name**
  - **Identifier** for the data set to be produced.
- **Data set description** (Description of the data that will be generated or collected)
  - its **origin** (in case it is collected)
  - **nature** and **scale**
  - to **whom** it could be useful
  - whether it underpins a **scientific publication**
  - Information on the existence (or not) of **similar data**
  - possibilities for **integration** and **reuse**
- **Standards and metadata**
  - Reference to existing **suitable standards** of the discipline.
  - If these do not exist, an outline on **how** and **what metadata** will be **created**.

Identification metadata

Provenance metadata

Use metadata

stefano.nativi@cnr.it

Data Documentation

- **Data sharing** (Description of how data will be shared)
  - **access procedures**
  - **embargo** periods (if any)
  - outlines of **technical mechanisms** for **dissemination** and necessary software
  - other **tools** for enabling **re-use**
  - definition of whether **access** will be widely open or restricted to **specific groups**
  - Identification of the **repository** where data will be stored, if already existing and identified, indicating in particular the type of repository (institutional, standard repository for the discipline, etc.).
  - In case the dataset **cannot be shared:** the **reasons** for this should be mentioned (e.g. ethical, rules of personal data, intellectual property, commercial, privacy-related, security-related).

Access metadata

Usability metadata

stefano.nativi@cnr.it

Data Documentation

- **Archiving and preservation** (including storage and backup)
  - Description of the **procedures** that will be put in place for **long-term preservation**
  - Indication of **how long** the data should be **preserved**,
  - its approximated **end volume**
  - the **associated costs** and how these are planned to be covered.

Preservation metadata

stefano.nativi@cnr.it

EDSI Coordination Task force

stefano.nativi@cnr.it

Proposal

- **Objective**
  - To coordinate the WP actions to address the Data and Services Interoperability challenges
  - To coordinate the WP actions to integrate the ECOPOTENTIAL digital systems and realize the ECOPOTENTIAL VRE/Platform
  - To resolve conflicts for interoperability between two or more ECOPOTENTIAL digital systems
- **Coordination Team**
  - Members from each WP
  - "ECOPOTENTIAL Virtual Laboratory Platform" WP coordinates the team, considering its scope and mandate
- **Activities**
  - Regular meetings (mostly virtual)
  - Produce good practice and guidelines on interoperability
  - Produce and maintain the ECOPOTENTIAL DMP

stefano.nativi@cnr.it

Conclusions

- What is the overall vision for the portal and/or platform?
  - ECOPOTENTIAL Virtual Laboratory Platform with resource sharing services accessible through apps/APIs
- Who is it for (who are the users)?
  - Scientists, decision-makers, citizens (activists),…
- What problem does it solve (what user need will it fulfill)? Why is it the right solution to the problem?
  - Open and interoperable access to data and knowledge on identified PAs
- How is it different from what is already available or planned?
  - Unified framework for ecosystem studies and management of protected areas

stefano.nativi@cnr.it

Conclusions

- Why should a user go to your portal?
  - Tailored apps, open APIs for developing new ecosystem information services
- How engaged are the users? Are they part of the planning and development process?
  - A permanent stakeholder consultancy group (GEO Ecosystem Community of Practice) will be created
- Will it use the GCI to obtain some of its data?
  - Full interoperability with the GCI is a requirement
- What is the plan for the portal once the project ends (post-project sustainability)?
  - Data Management Plan includes sustainability aspects

stefano.nativi@cnr.it

Thank you !

stefano.nativi@cnr.it