iCUPE Data Management Principles (ver 1)

Paolo Mazzetti, Stefano Nativi (Consiglio Nazionale delle Ricerche), Alexander Mahura, Tuukka Petäjä, Hanna K. Lappalainen (University of Helsinki); Steffen M. Noe (Estonian University of Life Sciences)

Florence, Italy, 07 March 2018

WP5: Data provision, interoperability and facilitation of data and services

Task 5.3: Compliance of iCUPE to GEOSS and Copernicus data sharing principles and GCI interoperability testbeds (horizontal) / Deliverable 5.3.1: iCUPE data management principles (ver 1)

Version 1

Introduction

The iCUPE WP5 ("*Data provision, interoperability and facilitation of data and services*") facilitates the data provision and services from the iCUPE project (www.atm.helsinki.fi/icupe) to the end-users, decision-makers, and stakeholders. Data obtained, integrated and accessed in WPs 1-4 are distributed in WP5 as data products and assessments. Data (ground-based, time-series, column, etc. observations) will be harmonized and assessed through developed novel methods, proxies and observables. These will be delivered through interoperability tools and services and will be available to iCUPE partners and any other end-users, decision-makers, and stakeholders (in agreement with existing data policies and open source principles).
To achieve the ERA-PLANET (European Network for Observing our Changing Planet; www.eraplanet.eu) overall objectives, and to pursue cross-thematic interoperability and contribute effectively to GEOSS (Global Earth Observation System of Systems; www.earthobservations.org/geoss.php), the iCUPE project will implement the best practices and recommended approaches of ERA-PLANET. This will allow project to contribute to GEOSS via the GCI (GEOSS Common Infrastructure) and to utilize relevant Copernicus data and Core Services and EU capabilities in the EO domain. iCUPE will promote and implement the use of open specifications (i.e. international standards, community specifications) for data sharing and will foster technological development to deliver more timely and high quality data and information, in compliance with the GEOSS Data Management Principles.

To this aim, the iCUPE WP5 releases two dedicated documents in three iterations: D5.3.x "iCUPE data management principles (ver x)" under responsibility of CNR, and D5.1.x "iCUPE data management plan (ver x)" under responibility of the University of Helsinki.

The **iCUPE Data Management Principles** (Task 5.3) describe the requirements for compliance with the obligations due to the participation of ERA-PLANET in the Horizon 2020 Open Data Pilot, and for the required contribution to major global initiatives, namely Copernicus and GEOSS.

The **iCUPE Data Management Plan (DMP)** (Task 5.1) will assure interoperability between the ERAPLANET projects/ strands and with other activities carried out as part of GEO Strategic Plan (2017-2019) (i.e., GEO Initiatives, Flagships, Foundational tasks). As part of making iCUPE research data findable, accessible, interoperable and re-usable (FAIR), iCUPE DMP will include information: on handling of obtained research data (during and after the end of the project); on types of data to be collected, processed, analyzed, etc.; on applied methodological approaches and standards; on whether data will be shared and/or made open access; and on how data will be curated and preserved (including after the end of the project). The most important aspects for data management include: discoverability (data and metadata should be discoverable); accessibility (data should be accessible in online services); usability (encoding, traceability, documentation, quality); preservation (preservation, verification); and curation (review and reprocessing, persistent and resolvable identifiers).
Following the signed iCUPE Consortium Agreement /Section 11 on Data Management/, the appropriate and secure use of material and data of the project will be enabled according to the application of common standards. In the iCUPE project, the DMP will follow the "*Guidelines on FAIR Data management in Horizon 2020*" (version July 2016; [1]) and according to the ERA-PLANET Data Management Plan (Deliverable 4.5). The iCUPE DMP will be updated during the lifetime of the project.

iCUPE is one of the four Transnational Projects selected through the ERA-PLANET call for proposals (the others are SMURBS, GEOEssential and IGOSP) launched on 2015. Since the four projects are all funded by the European Union through ERA-PLANET, they share the obligations from the participation of the ERA-PLANET action in the Horizon 2020 Open Data Pilot. Moreover, they have similar requirements of contribution to global and European initiatives, namely Copernicus and GEOSS. Therefore, the ERA-PLANET Consortium decided to have a harmonized approach to the definition of Transnational Projects DMPs. This harmonization activity is part of the ERA-PLANET WP4 on “Follow-up and monitoring of projects” (“*WP-4 will monitor the communication and dissemination plans and the data management plans drafts that will report on how the projects' partners will participate in the EC Open Data Pilot*”). As part of this activity, the release of an ERA-PLANET Data Management Plan (ERA-PLANET D4.5) was initially planned. Then, since ERA-PLANET does not share data itself, but only through the Transnational Projects, it was decided to focus the ERA-PLANET D4.5 on the definition of the ERA-PLANET Data Management Principles and the methodology for a harmonized approach to the Transnational Projects DMPs.

[1] http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

The present document bases on the ERA-PLANET D4.5 with specific reference to the implementation of the methodology in iCUPE.

1. Background and References

1.1. Horizon 2020 Open Research Data Pilot

Over the last years, the European Commission run a flexible pilot under Horizon 2020 called the Open Research Data Pilot. The Open Research Data pilot aims to improve and maximise access to and re-use of research data generated by Horizon 2020 projects and takes into account the need to balance openness and protection of scientific information, commercialisation and Intellectual Property Rights (IPR), privacy concerns, security as well as data management and preservation questions. In general terms, it establishes that research data should be 'FAIR', that is findable, accessible, interoperable and re-usable. These principles precede implementation choices and do not necessarily suggest any specific technology, standard, or implementation-solution.

In the 2014-16 work programmes, the Open Research Data pilot included only selected areas of Horizon 2020. Under the revised version of the 2017 work programme, the Open Research Data pilot has been extended to cover all the thematic areas of Horizon 2020 (“Open Research Data (ORD) - the uptake in Horizon 2020,” 2016).

The ERA-PLANET project has been funded under the Topic SC5-15-2015 “Strengthening the European Research Area in the domain of Earth Observation” in the call H2020-SC5-2014-2015 on “Growing a Low Carbon, Resource Efficient Economy with a Sustainable Supply of Raw Materials” with deadline on September 2015. As a project funded in the Horizon 2020 Societal Challenge 5 on Climate Action, Environment, Resource Efficiency and Raw Materials, ERA-PLANET participates in the Open Research Data Pilot.

In order to support the elaboration of DMPs, the European Commission released several documents including: “Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020” (European Commission Directorate-General for Research & Innovation, 2017); “Guidelines on FAIR Data Management in Horizon 2020” (European Commission Directorate-General for Research & Innovation, 2016); “Template for the Data Management Plan” (European Commission Directorate-General for Research & Innovation, n.d.).

1.2. ERA-PLANET Grant Agreement

Article 29.3 of the ERA-PLANET Grant Agreement establishes the minimum requirements for compliance with in the participation in the Horizon 2020 Open Research Data Pilot (European Commission, 2015):

Regarding the digital research data generated in the action[2] *('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 1);*

(b) ***provide information — via the repository — about tools and instruments at the disposal of the beneficiaries and necessary for validating the results*** *(and — where possible — provide the tools and instruments themselves).*

This does not change the obligation to protect results in Article 27, the confidentiality obligations in Article 36, the security obligations in Article 37 or the obligations to protect personal data in Article 39, all of which still apply.

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 1, 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.

1.3. The Group on Earth Observation (GEO) and the Global Earth Observation System of Systems (GEOSS)

The SC5-15-2015 call explicitly asked to launch a joint call for transnational projects that "address the issue of the coherence of European participation within the Group on Global Earth Observation (GEO) and should also provide a research and innovation component to the Copernicus programme" (European Commission, 2014). This means that, also in terms of data sharing and data management, the potential contribution by ERA-PLANET transnational projects must be compliant with GEO and Copernicus requirements.

In this direction, one of the objectives of the ERA-PLANET project is to "provide more accurate, comprehensive and authoritative information to policy and decision-makers in key societal benefit areas, such as Smart cities and Resilient societies; Resource efficiency and Environmental management; Global changes and Environmental treaties; Polar areas and Natural resources".

[2] In the Grant Agreement, the term 'action' refers to the funded project as stated in Article 2: "*The grant is awarded for the action entitled 'The European network for observing our changing planet — ERA-PLANET' (`action'), as described in Annex 1.*"

In its first decade of activities, GEO published a set of Data Sharing Principles (GEOSS-DSP) and a set of Data Management Principles (GEOSS-DMP) (GEO, n.d.). The current 2017-2019 GEO Work Programme includes activities aiming at the advancement and implementation of the DSP and GEOSS-DMP. In particular, it includes a dedicated Advancing GEOSS Data Sharing Principles Foundational task, while the GEOSS Common Infrastructure (GCI) Operations Foundational Task includes an activity on GCI Resource Quality that also aims at implementing the GEOSS Data Management Principles.

1.3.1. GEOSS Data Sharing Principles

The GEOSS 10-Year Implementation Plan as adopted on 16 February 2005, defined the following three GEOSS Data Sharing Principles:

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 **free of charge or no more than cost of reproduction will be encouraged for research and education**.

As it embarks on its second decade, GEO now aims to implement the following GEOSS Data Sharing Principles:

NDSP1. **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;

NDSP2. 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; and

NDSP3. all shared data, products and metadata will be made available with minimum time delay.

1.3.2. The GEOSS Data Management Principles

Since its inception in early 2000s, GEO has been working to facilitate 'data management approaches that encompass a broad perspective of the observation data life cycle'. GEO therefore strives to promote and encourage the implementation of Data Management Principles laid out below under five headings: discoverability, accessibility, usability, preservation, and curation.

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.

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.

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.

In the recent years, some initiatives have been launched on the practical implementation of GEOSS-DMPs through the realization of trustworthy repositories. For example, ICSU World Data System (WDS) and the Data Seal of Approval (DSA) recently launched a new certification organization: CoreTrustSeal. CoreTrustSeal offers to any interested data repository a core level certification based on the DSA-WDS Core Trustworthy Data Repositories Requirements catalogue and procedures. This universal catalogue of requirements reflects the core characteristics of trustworthy data repositories and is the culmination

of a cooperative effort between DSA and WDS under the umbrella of the Research Data Alliance to merge their data repositories certifications (Nativi et al., 2017) .

1.4. The Copernicus Programme

The Regulation n. 377/2014 of the European Parliament and of the Council of 3 April 2014 establishing the Copernicus Programme states that "Copernicus data and Copernicus information should be available freely and openly to support the Digital Agenda for Europe". More explicitly art. 36 says that "**the data and information produced in the framework of Copernicus should be made available on a full, open and free-of-charge basis subject to appropriate conditions and limitations**, in order to promote their use and sharing, and to strengthen European Earth observation markets, in particular the downstream sector, thereby enabling growth and job creation." (European Parliament and Council of the European Union, 2014)

The Regulation is implemented providing policies defined per user category and per data product, as for example in the "Copernicus Space Component Data Access Portfolio: Data Warehouse 2014 – 2020" (Hoersch and Amans, 2017).

2. The ERA-PLANET Data Management Principles

The ERA-PLANET Data Management Principles (EDMP) stems from the previously described requirements and constraints:

EDMP-1 All data generated in the action must be deposited in a research data repository and made accessible free of charge and at the FAIR conditions described in the DMP;

EDMP-2 All the scientific results generated in the action (e.g. presented in a publication) must be reproducible providing the required data and information about tools and instruments necessary for validation;

EDMP-3 All data generated in the action, which are relevant, directly or indirectly, for information to policy and decision-makers in key societal benefit areas must be accessible through GEOSS and Copernicus at the conditions described in the DMP and in compliance with GEOSS-DSP and GEOSS-DMP;

EDMP-1 and EDMP-2 assure compliance with the H2020 Open Research Data Pilot and ERA-PLANET Grant Agreement (see section 2.2). EDMP-3 assures compliance with GEOSS (see section 2.3). EDMP-1 also assures compliance with the general Copernicus sharing and management principles (see section 2.4)

3. Towards the iCUPE Data Management Plan

Figure 1 summarizes the planned process for the definition of the iCUPE Data Management Plan. The process involves four Actors with different roles and responsibilities:

- The ERA-PLANET Data Management Plan Responsible Party (CNR), which is in charge of defining the general Data Management Principles that Transnational Projects must comply with (ERA-PLANET T4.5).
- The ERA-PLANET Monitoring Responsible Party (CREAF), which responsible of the follow-up and monitoring of transnational projects (ERA-PLANET WP4)
- The iCUPE Data Management Plan Responsible Party (UHEL), which is in charge of defining the project DMP based on inputs from project data providers.
- The iCUPE Data Providers, which are in charge of providing the necessary information about the management of their data and how they will be made accessible. The iCUPE Data Providers are all the iCUPE partners which generate data and products as an outcome of the project.

The process includes four main steps (see Figure 1):

1) The ERA-PLANET Data Management Plan Responsible Party releases:
   a) the Data Management Principles document summarizing the general principles for the Transnational Project DMPs;

b) the Data Management Plan template;
c) the template for collecting information by data providers.

2) The iCUPE DMP Responsible Party (UHEL):
   a) revise and adapt the survey template;
   b) circulate the survey to the project potential data providers, collecting information about data that will be provided in the project.
3) The iCUPE DMP Responsible Party (UHEL):
   a) elaborates the draft of the project DMP based on survey responses and assessing compliance with the ERA-PLANET Data Management Principles according to the Data Management Plan template
   b) Submits the draft DMP to the ERA-PLANET Monitoring Responsible Party (CREAF), which, according to ERA-PLANET T4.2 “Monitor the quality of the deliverables, review them and request modifications if needed”
   c) Produces the final version of the DMP
4) The iCUPE DMP Responsible Party submits the DMP as an official iCUPE deliverable

Steps 2, 3 and 4 are repeated for the release of new versions of the DMP. Since the release of the first version of the iCUPE DMP was due before the release of ERA-PLANET D4.5, the previous procedure starts with the second iteration to release the second version of the iCUPE DMP.

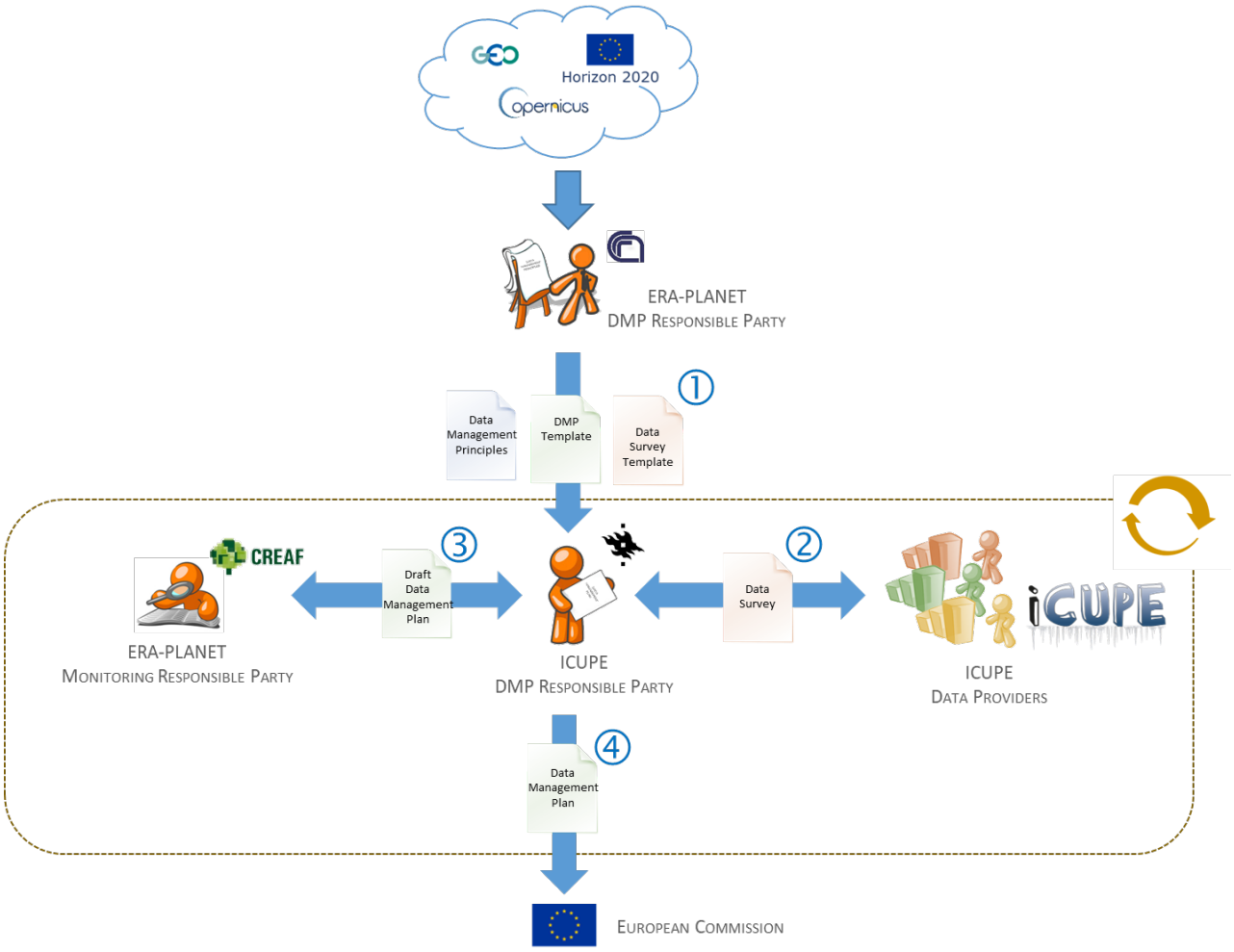

Figure 1 Process for the definition of the iCUPE Data Management Plan

3.1. The ERA-Planet Data Management Principles Document

The ERA-PLANET D4.5 is the ERA-PLANET Data Management Principles document.

3.2. The Template for the Transnational Project Data Management Plans (Annex I)

The ERA-PLANET D4.5 provides the Template for the Transnational Project Data Management Plan as annex.The Template for the Transnational Project Data Management Plan is based on the "Template for the Data Management Plan" (European Commission Directorate-General for Research & Innovation, n.d.). It includes a general section directly based on the "Template for the Data Management Plan" which should summarize the result of the data providers' survey. It also includes a specific section to assess and monitor the compliance with the ERA-PLANET Data Management Principles.

3.3. The Data Survey Template (Annex II)

The ERA-PLANET D4.5 provides the Data Survey Template as annex. The Data Survey Template is based on the "Template for the Data Management Plan" (European Commission Directorate-General for Research & Innovation, n.d.), with some questions eliminated and/or modified for facilitating the work of data providers who are not expert in interoperability. Some of those questions could be added in future surveys for the next releases of the DMPs.

References

European Commission. 2014. "TOPIC : Strengthening the European Research Area in the Domain of Earth Observation." http://ec.europa.eu/research/participants/portal/desktop/en/opportunities/h2020/topics/sc5-15-2015.html.

———. 2015. "Grant Agreement Number: 689443 - ERA-PLANET."

European Commission Directorate-General for Research & Innovation. 2016. "Guidelines on FAIR Data Management in Horizon 2020." http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf.

———. 2017. "Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020." http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf.

———. n.d. "Template Horizon 2020 Data Management Plan (DMP)." https://view.officeapps.live.com/op/view.aspx?src=http://ec.europa.eu/research/participants/data/ref/h2020/gm/reporting/h2020-tpl-oa-data-mgt-plan_en.docx.

European Parliament, and Council of the European Union. 2014. "Regulation (EU) No 377/2014 of the European Parliament and of the Council of 3 April 2014 Establishing the Copernicus Programme and Repealing Regulation (EU) No 911/2010." http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32014R0377&from=EN.

GEO. n.d. “Data Sharing & Data Management.” http://www.earthobservations.org/dswg.php.

Hoersch, B., and V. Amans, eds. 2017. “Copernicus Space Component Data Access Portfolio: Data Warehouse 2014 - 2020.” ESA. https://spacedata.copernicus.eu/documents/12833/14545/DAP_Release_v2.2.

Nativi, Stefano, Paolo Mazzetti, Paola De Salvo, and Steven Ramage, eds. 2017. “Report on the 2nd GEO Data Providers Workshop.” GEO. https://www.earthobservations.org/documents/meetings/201704_2nd_dpw/201704_2nd_dpw_final_report.pdf.

“Open Research Data (ORD) - the Uptake in Horizon 2020.” 2016. European Commission Directorate-General for Research and Innovation.