First version of Data Management Plan

June 2023

Action Number: 101079310
Action Acronym: EarthBridge
Action title: Building Bridges between Earth Observation and Environmental Sciences
Date: 30.6.2023
DMP version: 1.3.

Authors:

- Vitezslav Moudry
  moudry@fzp.czu.cz, 0000-0002-3194-451X
  Roles: Project Leader
  Affiliation: Czech University of Life Sciences Prague
- David Moravec
  dmoravec@fzp.czu.cz, 0000-0003-4348-2815
  Roles: Contact Person, Data Steward, Project Member
  Affiliation: Czech University of Life Sciences Prague
- Emilie Trakalova
  trakalova@fzp.czu.cz
  Roles: Project Manager
  Affiliation: Czech University of Life Sciences Prague
- Jan Komarek
  komarekjak@fzp.czu.cz, 0000-0002-3505-6755
  Roles: Data Curator, Project Member, Researcher
  Affiliation: Czech University of Life Sciences Prague
- Christopher Marrs
  christopher.marrs@tu-dresden.de
  Affiliation: TU Dresden
- Ruben Remelgado
  ruben.remelgado@tu-dresden.de, 0000-0002-9871-5703
  Affiliation: TU Dresden

Editors: Management committee members
Contact: Czech University of Life Sciences Prague
        Vitezslav Moudry
        moudry@fzp.czu.cz
        Kamycka 129
        165 00 Prague
        www.fzp.czu.cz
Website: www.EarthBridge.eu

Earth
Bridge

Content

Executive summary

This document informs on the Data Management Plan (DMP) of the EarthBridge project. As a DPM, this is a "live" document we expect it to be updated in response to the progress of the project and its emergent data and storage requirements. The DMP will focus on the maintenance of research data, whereas data derived in the scope of project management activities will be addressed in a dedicated Project management plan (PMP) and Communication and Dissemination Plan (CDP).

Project Abstract

Unprecedented human-induced changes in climate and biodiversity are increasingly threatening nature, human life, and well-being around the world. Earth Observation (EO) has become one of the most important methods for regularly obtaining standardised information about the Earth's environment and is therefore essential for addressing climate change, biodiversity loss, and other societal challenges. However, the full utilisation of EO is still limited. This is especially due to (1) a lack of experts in the field of EO with background in environmental sciences as well as (2) a lack of general awareness of the potential of EO in the field of environmental sciences or policy support and implementation. Universities play an essential role in bridging such knowledge and communication gaps by conducting research, developing new methods, and training new professionals. The EarthBridge project will:

(i)   substantially increase the interdisciplinary research capacity of the Czech University of Life Sciences Prague (CZU) in the field of environmental EO.

(ii)   and further improve the use of environmental EO tools by a wider community through twinning with two internationally leading universities: Technische Universität Dresden (TUD) in Germany, and Alma Mater Studiorum - Universitá di Bologna (UNIBO) in Italy. EarthBridge focuses on the development and application of environmental assessment methods for monitoring, restoration, and conservation of biodiversity in agricultural landscapes, a research topic of high relevance to Europe.

The project will benefit from ongoing research activities that will feed into existing EO data. Ways to achieve the bridging will include staff exchanges; scientific training and training in preparing research proposals; organisation of a summer school and scoping for a joint study programme; a synthesis project.

for early-career researchers; and involvement of regional/national institutions according to their needs or requirements.

1. Data Summary

Within the EarthBridge project, we expect to use several datasets, such as:

*Instrument datasets:*The following instrument datasets will be acquired in the project:

- **UAV campaigns; temperature, moisture, chlorophyll content and leaf area index field measurements**

This dataset will be collected by experts in the project, with our own equipment, which is well-described and known. The metadata describing the Data will be provided so that other researchers working in the same field of research could use these data.

Re-used datasets
We have found the following reference datasets that we have considered for re-use:

- **Copernicus Open Access Hub (https://scihub.copernicus.eu/)**

Owner of this dataset: The Copernicus Open Access Hub (previously known as Sentinels Scientific Data Hub) provides complete, free and open access to Sentinel-1, Sentinel-2, Sentinel-3 and Sentinel-5P user products, starting from the In-Orbit Commissioning Review (IOCR). The Copernicus programme is a cooperation between the European Union and the European Space Agency (ESA) for Earth observation. ESA coordinates the acquisition and delivery of Earth observation data from space. Starting with the successful launch of Sentinel-1A in 2014, the series of Sentinel satellites will deliver a wide range of data for environmental and civil security purposes. Under the ESA – European Community GMES Agreement of 28 February 2008, as amended subsequently, the agency is charged with managing the GMES Space Component (GSC) Programme. Under the Copernicus Agreement signed by ESA and the European Union (EU) on 28 October 2014, consistent with the EU Copernicus Regulation, ESA continues the activities related to the Copernicus Space Component. The Copernicus Data Hubs (Open Access Hub, Collaborative Data Hub, International Access Hub, Copernicus Services Data Hub) are ESA online portals, which make Sentinel data available to individuals and entities worldwide. The portals allow the searching for Sentinel data and the ability to download that data, as selected by the user.

The dataset can be used in the provided format without any conversion needed.

All the data have different versions according to the level of pre-processing (nowadays Level-1B, Level-1C, Level-2A) and we will use all versions flexibly to meet project requirements. If new versions become available during the project, new analyses will be done with those new versions.

The provider keeps old versions archived so the same reference data will be available to reproduce our results.

We will use them to answer the scientific questions raised in the Earth Bridge project - mainly to observe conditions in our areas of interest. We will use the dataset which are the primary source of Remote Sensing satellite observation for the EU.

We have found the following <u>non-reference datasets</u> that we have considered for re-use:

- **Airborne laser scanning point clouds**

Owner of this dataset: Individual National Parks. It is necessary to request access to the dataset from its owners. The dataset can be used in the provided format without any conversion needed. We already have a copy of some of these datasets. It is a fixed dataset; changes will not influence the reproducibility of our results. Only part of the dataset will be used; any filtering or selection will be well-documented. We will use the dataset as followsto calculate vegetation metrics.

- **Global Ecosystem Dynamics Investigation** (GEDI) (https://search.earthdata.nasa.gov/search)

Owner of this dataset: National Aeronautics and Space Administration (NASA). We will first need to convert the format before using it. The dataset may be updated in the future; therefore, we need to make a snapshot. Only part of the dataset will be used; any filtering or selection will be well-documented. We will use the dataset to calculate vegetation structure and to evaluate the data's usability.

- **UAV mosaics**

Owner of this dataset: CZU are owner of the dataset. The dataset can be used in the provided format without any conversion needed. We already have a copy of this dataset. It is a fixed dataset; changes will not influence the reproducibility of our results. Only part of the dataset will be used; any filtering

or selection will be well-documented. We will use the dataset as follows: To get the necessary information (spectral properties, spatial properties, temporal properties) of study localities (such as CZU experimental site Amalie) within the project. Those datasets will be used for prediction of biodiversity and other ecological variables defined by the project. We will need to harmonise different sources of existing data.

Data formats and types

We will be using the following data formats and types:

- **NetCDF4**

It is a standardised format. This is a suitable format for long-term archiving.

- **Geographic Tagged Image File Format**

It is a standardised format. This is a suitable format for long-term archiving.

- **Environmental Systems Research Institute Shapefile**

It is a standardised format. This is a suitable format for long-term archiving.

- **Comma-separated Values**

It is a standardised format. This is a suitable format for long-term archiving. We will have only a small amount of data stored in this format.

- **GeoJSON**

It is a standardised format. This is a suitable format for long-term archiving. We will have only a small amount of data stored in this format.

- **LAZ**

It is a standardised format. This is a suitable format for long-term archiving.

- **Hierarchical Data Format**

It is a standardised format. This is a suitable format for long-term archiving.

2. FAIR Data

Within the EarthBridge project we will fully comply with and follow the FAIR principles. EarthBridge will ensure public access to data hosted in coherence with GDPR regulations, Open Geospatial Consortium (OGC) standards and compatibility with the Global Earth Observation System of Systems (GEOSS). The collection of existing data and new data generated by the collaborative work of EarthBridge researchers will be stored in local secure servers at CZU and will be accessible via VPN or secure FTP.

2.1. Making data findable, including provisions for metadata

In the EarthBridge project methods and data will be shared for example through open repositories such as GitHub, DRYAD, and Zenodo or open journals specialised in data publishing that allow rigorous data scrutiny (e. g. Earth System Science Data). The data will be associated with the citable DOI (Persistent digital object identifier), which enable data to be findable. All derived data products will be described with corresponding metadata following the INSPIRE directive based on ISO standards 19115-3 (19115-2) and made accessible via the OGC CSW service and the emerging STAC (spatial-temporal asset catalogue).

We will use an electronic lab notebook to make sure that there is good provenance of the data analysis. We made a SOP (Standard Operating Procedure) for file naming. Naming files is irrelevant as most of the datasets consists of multiple (thousands of) files. Folders will be named according to following structure:

projectname/workpackage_number/task_number_and_name: -input_data/dataset_name/ - output_data/dataset_name/. We will be keeping the relationships between data clear in the file names. All the metadata in the file names also will be available in the proper metadata. The metadata will versioned according to the data update.

At this project stage (M6), we are considering the use of the CZU Data Platform (DP), which has been currently launched at CZU. DP natively solves the complete system of working with data (all standard types and formats), including their storage and advanced cold storage in compliance with security standards. Data are stored in accordance with the rules and needs in the CZU DP. The length of storage is practically unlimited unless the owner or economic restrictions provide otherwise; here is always the possibility of data

shifting (migration). Within the DP, data are stored and made available, among other things, in accordance with FAIR principles, enabling their machine traceability, accessibility, interoperability, and reuse.

2.2. Making data accessible

We will be working with the philosophy as open as possible for our data. It should be noted, however, that some data may not be published until they have been fully exploited scientifically. If some of the data could not become completely open immediately because of legal reasons (for example: privacy rights of single farmers) a data-sharing agreement will be required. Data will be released only as soon as restrictions are lifted. Intellectual property rights are given by the consortium agreement.

There are three levels of access employed in this project as follows.

Open Access – No restrictions and without ethical or legal restrictions

Partial Access – Refers to the part of the dataset that cannot be shared due to legal or ethical reasons.

Closed Access – Data is only available to project partners and not shared beyond the partnership.

Metadata will be openly available including instructions on how to get access to the data. Metadata will not be available in a form that can be harvested and indexed. For the reference and non-reference data sets that we reuse, the conditions are as follows:

Data sets	Licence/Conditions
Copernicus Open Access Hub	Freely available with obligation to quote the source (e. g. CC-BY).
Airborne laser scanning point clouds	Available under specific restrictions, which we will follow in our project: Available for research purposes.
Global Ecosystem Dynamics Investigation (GEDI)	Available under specific restrictions, which we will follow in our project: Available for research purposes.
AV mosaics	Available under specific restrictions, which we will follow in our project: Available for research purposes.

2.3. Making data interoperable

We will be using the following standardised data formats and types:

- NetCDF4
- Geographic Tagged Image File Format
- Environmental Systems Research Institute Shapefile
- Comma-separated Values
- GeoJSON
- LAZ
- Hierarchical Data Format

2.4. Increase data re-use

As explained above in Section 2.2, some of our data may not become completely open immediately. There could be some Intellectual Property reasons why our data cannot be open. A nominated person from the project consortium will be given the decision power to move documents or data to a new location after the project has finished. We do not plan on archiving data (using so-called cold storage) for long term preservation during the project.

To validate the integrity of the results, the following will be done:

- We will be instrumenting the tools into pipelines and workflows using automated tools.

- We will use independently developed duplicate tools or workflows for critical steps to reduce or eliminate human errors.

- We will run part of the datasets repeatedly to catch unexpected changes in results.

3.   Other research outputs

We use Data Stewardship Wizard for planning our research data management and creating this DMP. The management and planning of other research outputs are done separately. Still, we benefit from data stewardship guidance (e.g., FAIR principles, openness, or security) and it is reflected in our plans with respect to other research outputs.

4.   Allocation of resources

FAIR is a central part of our data management; it is considered at every decision in our Data Management Plan. We use the FAIR data process ourselves to make our use of the data as efficient as possible. Making our data FAIR is therefore not a cost that can be separated from the rest of the project.

At this project stage (M6), we are considering the use of the CZU Data Platform (DP), which has been currently launched at CZU, project coordinator. The costs associated with the use of such Platform are not known at this stage.

None of the other used repository's charge for their services.

Vitezslav Moudry is responsible for implementing the DMP, and ensuring it is reviewed and revised.

Jan Komarek is responsible for reviewing, enhancing, cleaning, or standardising metadata and the associated data submitted for storage, use and maintenance within a data centre or repository.

David Moravec is responsible for the management and proficiency of data including data processing, data policies, data guidelines, and data availability.

To execute the DMP, we have trained support staff available within the project consortium.

We do not require any hardware or software in addition to that available in the project partners' institutes.

5. Data security

Project members can carry data with them on password-protected laptops. All data centres where project data is stored carry sufficient certifications. All project web services are addressed via secure HTTP (https://...). Project members have been instructed about both generic and specific risks to the project. The collection of existing data and new data generated by the collaborative work of EarthBridge researchers will be stored in individual and join storage space e.g., local secure servers at CZU and will be accessible via VPN or secure FTP.

The risk of information loss in the project or organisation is acceptably low. The risk of information leak in the project or organisation is acceptably low. The risk of information vandalism in the project or organisation is acceptably low.

6. Ethics

We will not collect any research data connected to a person, i.e. "personal data". Only personal data within the projects are those related to the project participants and members of Stakeholder Advisory Board, such as Name, Surname, Institution, and email addresses. While collecting the personal data for the purposes of the events organized under the EarthBridge umbrella, we follow GDPR rules via participant consent.

For reused non-reference datasets, the consent for privacy sensitive data will be solved as follows:

- Airborne laser scanning point clouds

   None of the data from this dataset is personal data.

- Global Ecosystem Dynamics Investigation (GEDI)

   None of the data from this dataset is personal data.

- UAV mosaics

   None of the data from this dataset is personal data.

7. Other issues

We used the Data Stewardship Wizard with its Common DSW Knowledge Model (ID: dsw:root:2.4.4) to make our DMP. More specifically, we used https://czu.ds-wizard.org. DSW instance where the project has a direct URL: https://czu.ds-wizard.org/projects/fbeac497-5350-4ea0-8911-ea7a553931ad. Data management in the project will be carried out in accordance with the regulations in the field of data management and handling at CZU, TU Dresden, and UNIBO.