CEPOPA Data Management plan v1.0

19.05.2020

CEPOPA will apply FAIR practices for the management of the data produced in its framework. We will be producing primary (measured) quantitative and qualitative data, from cultivations, photometers, analytical machines (e.g. HPLC, GC, NMR), downstream process, isolation and purification, in silico analysis etc. and our secondary (processed) data will come from evaluation software such as graphpad prism, origin or MS excel. We expect to produce up to 200 GB data over the course of the project, stored and (partly) published in a two-stage data storage solution.

In the first stage, the data will be made accessible from the whole research team via the online SEEK instance, FAIRDOMHub (www.fairdomhub.org) [1]. FAIRDOMHub acts as a public, centrally managed resource with project-controlled spaces, and it offers sustainability of the published data for at least 10 years. More than that, FAIRDOMHub offers integration with modelling tools and support for reproducible publications with leading journals. Metadata will be added to the files with primary data from laboratory instruments, as well as the ones with secondary datasets, and subsequently they will be saved into more share-friendly formats. SEEK software [2] is a content aggregator used as a standards-compliant web-based metadata management and cataloguing Hub, to organize the members of the groups participating in the consortium, as well as their outputs. It uses the ISA structure (Investigation, Study, Assay) to organize the heterogeneous data types/models/SOPs (Standard Operating Procedures), and allows users to upload content directly, or link to content held in other storage drives/repositories. To ensure project wide interoperability, a convention of naming and versioning (revision numbering based on the DDI-Standard) will be established.

In a second stage, part of the raw, primary and secondary data will be made accessible to the public in FAIRDOMHub under an open license (Creative Commons Attribution-Non-Commercial 4.0). The PI of CEPOPA, Assist. Prof. Ioannis Pavlidis, is designated as asset’s housekeeper/gatekeeper, who gives the final permission on the requests to publish the data. Data that are not interesting for commercial/industrial/academic purposes will be released directly. Data that are important for academic purposes will only be released simultaneously or directly after the publication of the peer-review articles that include the secondary data produced. Data that are sensitive for commercialization / industrial purposes will not be made public, as a potential patent application will be compromised. Such data will only be made available after the filing of the patent application, and after the one-year term, when the patent application is open to the public.

To make data findable, they will be described using the ISA-Tab format and get a persistent Identifier (digital object identifier, doi) via FAIRDOMHub. To increase re-useability, all data will be described and stored in common, if possible open and non-proprietary, file formats such as odt, ods/csv, pdf/a, tiff and mp4. If some laboratory equipment produces data in a proprietary format or a specific format that is not widely used, we will deposit the raw data in the proprietary format, and export them and deposit them as primary data in our common depository in a common format, such as txt and/or csv.

There are no privacy sensitive data produced in the framework of CEPOPA.
On the technical level, Department of Chemistry and University of Crete have their own servers and data management routines, as well as their IT services that can support the produced data. Moreover, each academic member of the University of Crete has 1 TB of cloud storage in OneDrive for Business, via the platform “ΔΗΛΟΣ 365”. On the strategic level, the University of Crete recently adopted a guideline for Open Data Management (available only in Greek). We have selected Assist. Prof. Ioannis Pavlidis of our research team as data management responsible, who is willing to more intensely engage with the funding council and the ICT Manager and OpenData Coordinator of Department of Chemistry, Mr. Michael Theodorakis, about planning/implementing the data management throughout the project.

**Cost:** Our data management requirements include the person-months invested from all team members on the proper data management throughout the project for the data preparation for ingest and storage at FAIRDOMHub.

This data management plan will act as a living document that is regularly reviewed and updated over the lifetime of the project, and the consulting from Mr. Theodorakis will be arranged accordingly. The plan will incorporate further specific requirements, such as plans for deposition into public disciplinary depositories, to reach a more specialized audience.

As of 18th of May 2020, the project page for CEPOPA has been prepared in FAIRDOMHub, under the programme "Hellenic Foundation for Research and Innovation - Supporting academic Staff and Researchers". Profiles have been prepared for all team members, and 5 investigations are prepared for the respective workpackages of CEPOPA, all of them being open to the public (Figure 1). Moreover, in WP5, there is already uploaded a manual for the most important actions in FAIRDOMHub (DOI: 10.15490/fairdomhub.1.document.689.1). This Data Management Plan will also be uploaded under the same investigation.

![Figure 1: Overview of the structure of CEPOPA project in FAIRDOMHub.](image)

**References:**
