



Working Group 4: Annual Report First Year

Dagmar Triebel and Gila Kahila Bar-Gal

The scope of COST MOBILISE action is to foster a cooperative network in Europe to support excellent research activities, and facilitate knowledge and technology transfer around natural science collections. The increase interest in natural science collections research followed with an influx of data from various kings (digitized, genetics, 2D and 3D imaging, isotope etc.). Therefore, data archiving, long-term preservation and availability are important topics for every data producer and data provider in the natural science collections communities.

In the light of the rapid development of IT technology, WG4 will address these problems by assessing current standards and strategies for data archiving and suggesting efficient and practical strategies for different kinds of institutions.

In order to establish guidelines and required standards for data archiving and long-term preservation that the MOBILISE community will adapt, the first-year focus was to establish a group of experts that are interested in the field; learn and become updated with the standards e.g. the Open Archival Information System (OAIS) and procedures and protocols used by the community to archive scientific data.

At the end of year 1 we have 23 colleagues (9 female) that are interested to participate in the activities of WG4 [listed as “contributors” in the OSF platform (see under <https://osf.io/wq7ej/>)]. Among them a group of 14-16 experts from the stakeholder groups, service providers and target group are willing to cooperate and work together as core group. They belong to four major scientific domains: (a) nature science (collection) domain, (b) library domain, (c) informatics and information sciences domain and (d) economics and social sciences domain. Half of the people are experienced in EU-related initiatives. Only 6-7 could be categorized as early-career-researchers.

The WG4 workshop in Sofia brought together 14 group members (see Report of Sofia workshop). The major outcomes of the first workshop were:

- Definition of “data archiving” and related terms and the difference from data backup.
- Establishment of a first draft list of terms to be defined is available under https://costmobilise.biowikifarm.net/wiki/Definitions_of_core_terms_in_the_data_archiving_context

- Defining “long-term archiving” – The need to bridge between the rapid development of computational technology and the availability of the information in the future (100 years). The consensus was 10 years to infinity.
- Access to archived data (present and future time) is required at all time. Therefore, the parameters scalability, reliability and reachability must be in any archived data.
- It has to be distinguished between bit-preservation and functional preservation. Technical standards, e.g. iso-normsd and content standards are crucial.
- Archiving requirements includes migration of data and should result in FAIR data (Findability, Accessibility, Interoperability, and Reusability). Archiving should be done in a FAIR way.
- Discussions on storage concepts and models used. Special attention to the Open Archival Information System reference model.
- 10 presentations held at the WG4 workshop in Sofia are ready to be downloaded under [https://costmobilise.biowikifarm.net/wiki/WG4_Workshop_%22Data_storage_and_archiving_strategies%22_in_Sofia_\(NMNHS\)](https://costmobilise.biowikifarm.net/wiki/WG4_Workshop_%22Data_storage_and_archiving_strategies%22_in_Sofia_(NMNHS))

The communication and discussions within the group occurs through the websites <https://osf.io/j6psx/> and https://costmobilise.biowikifarm.net/wiki/Working_Group_WG4 and are open accessible.

Useful links and material on the topic are available under https://costmobilise.biowikifarm.net/wiki/Useful_links_and_materials