Horizon Europe

Action Number: [101046909]

Action Acronym: REMAP
Action title: Reusable mask patterning

Date: [31-08-2024]

D5.6 Data Management Plan Update 1

Funded by
the European Union

1. Executive Summary

The Horizon Europe Model Grant Agreement requires that a data management plan ('DMP') is established and regularly updated. At the start of the action, the use of a specific template was recommended for Horizon Europe beneficiaries, as utilized for the first version of REMAP's DMP (D5.3).

Recently, an online platform called Argos has been specifically established by OpenAIRE and EUDAT to facilitate Research Data Management (RDM) activities concerning the implementation of DMPs. The platform guides users throughout the process of writing and maintaining DMPs. It also *utilises the OpenAIRE pool of services and inferred sources to make DMPs more dynamic in use and easier to be completed and published*.

As the recommended interface, the REMAP Consortium has decided to migrate its DMP to the Argos platform. D5.3 remains the reference document, with any updates on REMAP's DMP now accessible at the following link:

https://argos.openaire.eu/explore-plans/publicOverview/42eff5b3-e2f6-4e43-ae65-b7212f9790fa

2. Data Summary

Cf. D5.3 and Argos platform.

3. FAIR data

With the aim of fostering a continuous improvement of data management according to the FAIR principles, and in response to issues encountered by the REMAP Consortium (cf. section 8.), we have initiated a protocol for standardized metadata association at the time of data collection.

The protocol consists of an online platform embedded in the relevant Open Data & Software section of the project website, wherein participant researchers can automatically generate a standardized name for a given dataset typically associated to an "experiment", the name itself being a combination of metadata entered by the user through a template. Given that the approach is virtually applicable to any kind of "experiment", from a synthetic lab protocol to a sample characterization, to an ethnographic fieldwork interview, a selection of ad-hoc templates must be designed, which should consider the seamless transfer of metadata to/from specific scientific instrumentation to minimize manual intervention and be amenable to improvements and upgrade.

The current beta versions are dedicated to generic sample naming and PPMS (physical properties measurement system) and can be accessed at the following link: https://www.nm-2.cloud/tx/tools/?c=remap. Currently, we are in the testing phase, i.e. checking their performance and compilers' satisfaction in view of expanding the portfolio of templates to all activities performed within REMAP that are reasonably applicable.

The great added value of such an approach is that metadata encoding is turned from a cumbersome activity merely performed by compilers to abide by the FAIR principles after the data has been acquired, to a functional activity that is instrumental to plan, generate and maintain a well-organized (standardised) research record for the benefit of the Consortium's research participants, as much as external users.

REMAP's data experts are available to discuss the potential implementation of this platform to Argos, which we strongly believe would be transformative for the European research landscape. We believe that freeing researchers from the hassle of manually encoding metadata descriptors as a post-acquisition liability is key to an effective adoption of FAIR practices by the research communities worldwide, and to higher quality research results.

4. Other research outputs

Cf. D5.3 and Argos platform.

5. Allocation of resources

Cf. D5.3 and Argos platform.

6. Data security

Cf. D5.3 and Argos platform.

7. Ethics

Cf. D5.3 and Argos platform.

8. Other issues

While making use of the Argos platform in drafting this deliverable, the Consortium has encountered some difficulties, which are listed below for kind consideration by the Commission.

1. Whereas the REMAP consortium adheres to the FAIR principles since before the start of the action, the scale of commitment requested by the new Argos platform is incompatible with the resources allocated initially.

2. Argos appears to be designed for the description of data at very low hierarchical level, whereas larger datasets and more complex relationships between datasets typically acquired in research projects would require numerous individual entries for a complete description within the platform, aggravating issue No. 1. On the one hand, fully describing and managing each individual data at micro-level is prone to encoding mistakes and is akin to treat stochastic physical systems in a deterministic manner. On the other, by focusing on the micro-level there is a risk to lose connection to the context of the investigated subject and its relationship to other data.
    - Example: a dataset describing a typical set of experiments aimed at identifying a functional relationship between two variables could lead to five samples (physical data), five X-ray diffractograms (digital data), one higher order graph resulting from the combination of the latter data, and so on. It is currently impossible to generate a single entry for the dataset, given the different nature of its individual components, leading to ineffective retrieval of meaningful physical information from users who access the platform, as well as a dispersion of the compilers' resources.

3. For the FAIR principles to be effectively implemented, data duplication is to be avoided, as this inevitably leads to inconsistency and, therefore, to ambiguity of the data, besides entailing an inefficient use of resources[1]. Hence, the current EU requirements to manually populate Sigma, Argos (in substitution of a DMP), a repository platform (e.g. Zenodo), and the project website, combined with each partner Country's request for researchers to compulsorily fill out institutional repositories as part of research quality assessment practices, are against the very principle of interoperability, as there are no cross-talk capabilities, aggravating issues No. 1-2.
    - Example: the matadata pertaining to the dataset example for issue No. 2 and fully described in Argos cannot be transferred machine to machine to the Argos platform or vice versa, which could imply that a researcher invests more resources in the data description than in data acquisition and interpretation.

As a response to such issues, the REMAP Consortium has come up with the strategy described in section 3., which we hope the Commission will consider worthy of implementation at EU funding level upon further development.

[1] Idan Gadbank *et al*., Prevention of data duplication for high throughput sequencing repositories. *Database*, vol. 2018, bay008 (2018). https://doi.org/10.1093/database/bay008