JOINT CALL FOR PROPOSALS

Biotechnology for a sustainable bioeconomy

Deadline for submission of full proposals:
July 20, 2017, 13:00 h (CEST)

Data Management

published on May 19, 2017

Data management

Data management is an essential component to the success of projects using systems/synthetic biology approaches and/or bioinformatic tools. Representatives from academia, industry, funding agencies, and scholarly publishers designed and jointly endorsed a concise and measurable set of principles referred to as FAIR Data Principles[1] with the intention to provide a guideline for reusability of data holdings. Four foundational principles – Findability, Accessibility, Interoperability and Reusability– are a necessity of data management. The EC recently published Guidelines on [FAIR Data Management in Horizon 2020](#).

Fulfilling the FAIR principles needs the use of software platforms that enable capturing, cataloguing, and annotating data, associated with well documented SOPs, and supports interlinking data from specialized, as well as local collections. Moreover, in a systems approach, well annotated models (including parameters) must be catalogued and interlinked with relevant data. Each project should have one single starting point starting at which one can find out everything about the project.

Participation in this call requires fulfilling the FAIR principles, including the use of a cataloguing platform as described above. Therefore research data and non-data assets like algorithms, tools and workflows or metadata produced in the projects funded under this call must be: (i) machine-readable (ii) citable and (iii) must be published in a registered repository and (iv) interlinked with other project outcomes in a cataloguing platform. Privacy sensitive data that cannot be published needs to be catalogued, such that the creator of the data can be easily found for possible questions and collaboration. Data that is subject to Intellectual Property for Patents needs to be recorded and published in due course.

The repository must be registered in as a "trustworthy repository" in (i) BioSharing or (ii) re3data (Registry for Research Data Repositories). For this, resources of existing community knowledge and data management platforms in Europe shall be preferably used.

No section for data management is foreseen in the template for pre-proposal.

However, the projects must consider already the cost of data management in their proposals. **The help given to the applicants should enable them to (i) properly estimate what cost will be incurred by performing FAIR data management, and (ii) provide a section in the template for full proposals, in which some guiding questions help the applicants to answer the most relevant and pressing data management questions to the reviewers.**

Building the data management plan should be based on existing check lists, such as the FAIRDOM checklist document: http://fair-dom.org/knowledgehub/data-management-checklist/.

The checklist addresses: (i) responsibilities; (ii) types of study, data and models; (iii) volume and life cycle of the data; (iv) data and model processing and access policies; and finally (v) documentation and metadata.

Answering the questions in the checklist, applicants will identify and resolve key questions about their data, models, SOPs and associated metadata. The data management plan will outline how data flows and the requirements on data, metadata, storage and data transfer throughout the project and beyond the project.

The costs associated to the data management plan (*e.g.* travel expenditures of staff to data management training; salary costs of staff curators; costs of servers, storage, archiving and backup) must be derived from this and clearly presented.

Guiding questions include the following:

Generate and Store

- Who will generate the data/model/SOP?
- Who will receive the data/model/SOP?
- What does the user need from the generator?
- Where will you store data/model/SOP?

---

[1] http://www.nature.com/articles/sdata201618

- How much storage capacity will you need short term?
- How will you transfer it?
- How much will you keep for longer? Who is responsible for this step?
- How will the data be made available for processing?

Curate

- Who will curate the data/model/SOP?
- How will the data/model/SOP be interlinked?
- How is data tracked through processes?
- How are versions of models and SOPs tracked?

Access

- Are you allowed to share data?
- Where, when and how will you make the data/model/SOP available?
- Which public archives will you deposit your data/model/SOPs in?
- How will you make the project's data/model/SOPs available in a unified way through a one-stop single starting point starting at which one can find out everything about the project?

In the pre-proposal phase, these questions can be answered in a preliminary fashion.

While developing answers for the proposal, it will also become clear who is **responsible as a contact person**, for which parts of the data management of the project. This helps in defining whom to choose as data management contacts for the ERACoSysMed data management, and who would make sense for cross-project data management meetings.

Please note that data management is an **ongoing** **activity of iterative improvement and adaptation rather than something that can be finished in one short effort. The data management plans must reflect this.**

All of these points are relevant both in systems biology and systems medicine. In *addition* to the points addressed in the checklist, in systems medicine, the work with privacy sensitive data needs to be addressed. For the purposes of the proposal it makes sense to consider two groups of data: privacy-sensitive and privacy-insensitive data, and consider these independently, as well as considering:

- How can privacy sensitive data become privacy insensitive
- How can privacy insensitive data become privacy sensitive during the project's run time.