

## Questionnaire concerning the data management requirements

The purpose of this questionnaire is to provide basic information in order to identify the research data and datasets that are used and created during the project. By identifying data used during the course of the project, we can better estimate the requirements for research data infrastructure needed to run your fellowship.

The questionnaire is split into two parts: the requirements for the collaborative cloud system (*Mecila Cloud*) and the requirements for the long-term archive.

- The collaborative cloud system will facilitate the day-to-day operations on “living” data in a Nextcloud - a Dropbox-like functionality, paired with a collaborative Office-style online solution, video chat, project management tools and such like. All data will be backed up in a tiered storage solution at the University of Cologne. This is not a backup service for your whole computer, but for all your research data.
- The archival system will provide the long-term citable data store for all research data which is deemed “dead”, i.e. data which will in itself have no more changes but which has to be kept for citations, re-use and accountability.

Please fill out the questionnaire and send by mail to [milagros.pacco@uni-koeln.de](mailto:milagros.pacco@uni-koeln.de)

---

Surname(s), First name(s)

### A) Cloud system

1. How much work-in-progress research data do you have at the moment? Please state amount in GB:

2. How much work-in-progress research data do you estimate you will have in seven months' time? Please state amount in GB:

3. What kind of data do you have? Which formats do you use to produce your data?

*Examples: text documents (i.e. DOC, ODF, PDF, TXT etc.), structured text (i.e. HTML, JSON, TEX, XML etc.), tables (i.e. CSV, ODS, XLS, SAS, Stata, SPSS etc.), databases (i.e. MS Access, MySQL, Oracle etc.), images (i.e. JPEG, SVG, PNG, GIF, TIFF etc.), audio (i.e. MP3, WAV, AIFF, OGG etc.):*

4. How big are the largest individual files (provide amount in KB, MB, or GB)?

5. Does your research include working with sensitive personal data? Do you anonymise it and if so, how?

**B) Archive**

1. How much archival ("dead") data do you have at the moment? Please state amount in GB:

2. How much archival ("dead") data do you estimate you will have in seven months' time? Please state amount in GB:

3. Do your long-term storage formats differ from your work-in-progress file formats? If so, which and why?:

4. How many people do you expect to re-use your data in the next years?

5. How would you like your finished research data to be presented? How would you like it to be found online?

6. Does your field of work have a metadata and digital object identifier standard? If so, which?

7. Do you want your archival data to be harvested by a specific archival institution for your field of study? If so, by which?

8. Under which license do you want to publish your research data?

**Training exercises**

Do you feel the need to have training exercises concerning the handling of living or dead data, metadata and the like? We will provide extensive documentation anyway, but if you would like specialized training, don't hesitate to ask for it.