## Data Archiving and Networked Services : CESSDA list of Data Management Questions

### Overview

Title of the project/study

Date of this plan

**Description of the project**
What is the nature of the project?

**Origin of Data**
What kind of data will be used during the project?
If you are reusing existing data: What is the scope, volume and format?
How are different data sources integrated?
If you are collecting new data can you clarify why this is necessary?

**Principal researchers**
Who are the main researchers involved?
What are their contact details?

**Collaborating researchers** (if applicable)

**Funder** (if applicable)

**Data producer**
Which organisation has the administrative responsibility for the data?

**Project data contact**
Who can be contacted about the project during and after it has finished?

**Data owner(s)**
Which organisation(s) own(s) the data?

**Roles**
Who is responsible for updating the DMP and making sure that it’s followed?
Do project participants have any specific roles?
What is the project time line?

**Costs and Resources**
Are there costs you need to consider to buy specific software or hardware?
Are there costs you need to consider for storage and backup?
Are potential expenses and resources for (preparing the data for) archiving covered?
What resources will be dedicated to data management ensuring that data will be FAIR?

### Organising and documenting your data

**Data collection**
How will the data be collected?
Is specific software or hardware or staff required?
Who will be responsible for the data collection?
During which period will the data be collected?
Where will the data be collected?

**Data organisation**
How will you organise your data?
Will the data be organised in simple files or more complex databases?
How will the data quality during the project be ensured?
If data consists of many different file types (e.g. videos, text, photos), is it possible to structure the data in a logical way?

**Data type and size**
What type(s) of data will be collected?
What is the scope, quantity and format of the material?
After the project: What is the total amount of data collected (in MB/GB)?

**File format**
In what format will your data be?
Does the format change from the original to the processed/final data?
Will your (final) data be available in an open format?

**Folder structure and names**
How will you structure and name your folders?

**File structure and names**
How will you structure and name your files?

**Documentation**
What documentation will be created during the different phases of the project?
How will the documentation be structured?

**Metadata**
What metadata will be provided with the collected/ generated/ reused data?
How will metadata for each object be created?
Is there any program that can be used to document the data?
Can metadata be added directly into the files or will the metadata be produced in another program or document?

**Metadata standard** (if applicable)

### Processing your data

**Versioning**
What is your strategy concerning versioning your data files (and scripts) during the project?
Will you create and/or follow a convention for versioning your data?
Who will be responsible for securing that a “Masterfile” will be maintained, documented and versioned according to the project guidelines?
How can different versions of a data file be distinguished?

**Interoperability**
Will you make use of established software and hardware? If not, how does the software and hardware you use relate to other research?

If applicable:

Will you make use of established terminologies/ontologies (i.e. structured controlled vocabularies) in the project? If not, how do your terminologies relate to established ones?
Which coding is used (if any)? Will you build on established coding schemes? If not, how does your coding relate to other research?

**Data Quality**
How will data quality be evaluated?
What data quality control measures will be used?

### Storing your data and metadata

**Storage**
How and where will the (meta)data be stored during the project?
For how long will the (meta)data be stored?

**Backup**
How, where and at what intervals will the (meta)data be backed-up?
How will data be recovered in the case of a (meta)data loss incident?

**Security**
How will sensitive (meta)data be protected? (if applicable)
How will (meta)data access be managed?

### Protecting your data

**Ethical review** (if applicable)

**Informed consent** (if applicable)

**(sensitive) Personal data /confidential information** (if applicable)

**Intellectual property rights (IPR)/Copyrights**
Are there IPR or copyright issues to consider?
Will permission be needed to collect/reuse the data?
Will these rights be transferred to another organisation for data distribution and archiving?

**Agreements** (if applicable)

**Restrictions** (if applicable)

### Archiving and publishing your data

**Archiving**
How and where will the data be stored after the project’s completion?
Will you archive your data in a trusted data repository?
Will the application of a persistent identifier to your data be ensured?

**Data formats**
What formats will you provide your data in for archiving (and sharing)?
Will specific software be required to process your data? Can this software be deposited with the data?

**Access** (if applicable)

### Discovering data

**Identification of needs**
Do you plan to use existing data for your research?
What is the purpose for which you need the data?
What do you want to learn from the data?
What type of data do you need?

**Search for data**
Do you know where the data may be located?
How do you plan to search for the data?

**Evaluation of data quality**
What is the minimal required quality of the data (in terms of origin, contents, scope, size, methods, etc.)?
How do you plan to evaluate data quality (evaluation of metadata, tests, analysis, comparisons)?

**Gaining access to data**
What are the (expected) terms and conditions for data access and use?
What is the (expected) process for gaining access to the data?
What is the (expected) time-span of the process for gaining access to the data?
What are the (expected) costs for data access and use?