

## Exercise 4. Data Flow Map 2

**Note:** *This assignment is only a reflective practice, you do not need to submit it nor will you receive feedback on it.*

In Module 3 you have learned about the main elements of the FAIR principles (i.e. Documentation & Metadata, Interoperability, Access to Data, Persistent Identifiers and Licences). In Module 4 we have explored some best practices for data organisation. Also we explored some documentation tools and we discussed data publication in data repositories.

We provide you with the following [template](#). You can either use the template, or paper and post-its - use your own creativity to make this assignment. You'll find the step-by-step instructions below.

### Step 1 - Use the datasets and description from part 1

1. Go back to the results of Exercise 1 where you listed the datasets you will be collecting/creating in your project and you also briefly described them.
2. Reflect on how you can apply what you have learned in Module 3 and 4 to the different datasets in your list. We are providing you with new themes and some guiding text that can help you in your reflection and will allow you to structure the information. Remember to take into consideration the flags (relevant attributes) you have added to the different data types, which can determine the way you handle each of the themes. Also, remember that you can duplicate slides, if you have more than three datasets.
3. Copy the datasets to the new template

### Step 2 - Reflect on the theme Data organisation

**Data organisation** - good documentation starts with good data organisation. Think about a good folder structure for your project. In which folder will you place the datasets you listed? Think also about some naming conventions for your files. You can select one convention to use across all datasets or you can define different ones for each dataset. But, remember that once you choose a convention you need to be consistent.

Reflect and answer this as one theme for all the datasets:

1. What folder structure will you use for your project? (and how does the data fit in there)

2. What naming convention are you going to use? (show us an example based on your expected dataset)
3. Additional remarks

## Step 3 - Reflect on the theme Documentation

**Documentation** - think about what type of documentation is relevant for the different datasets you will be collecting/creating. For some datasets, you might need to record only metadata, for others you might need to create a data dictionary or in other cases you might need more extensive information describing the process of data collection. Additionally, indicate if you will be using any of the documentation tools we explored in Module 4.

Reflect and answer these questions for each dataset:

1. What type of documentation do you need to generate/write? (e.g., metadata, data collection process, data dictionary, code versioning, etc.)
2. What tools are helpful to generate/write the documentation needed?
3. Additional remarks

## Step 4 - Reflect on the theme Metadata

**Metadata** - Metadata is a relevant type of documentation. Think about what type of metadata is necessary to add to your datasets. Do you need to collect just generic metadata e.g. Date of collection/creation, author? Or, do you need more specific metadata e.g. location of collection, instrument model, name of the software to be able to read the data? Remember that you can always look in [FAIRsharing.org](https://fairsharing.org) if a metadata standard exists for your discipline or type of data. If you find one that makes sense to use for your datasets, provide the name and the link to it.

1. List some of the relevant metadata you need to record for each dataset and/or indicate (and provide the link to it) if you will use a metadata standard existing in your discipline.

## Step 5 - Reflect on the theme File formats

**File formats** - Remember that file formats can enable or not interoperability and reusability. You already provided the data format in which you will collect/create the different datasets in your project (assignment 1). Go back to that information and think if you can convert those formats to an open file format that would increase interoperability if somebody would like to reuse the data.

Reflect and answer these questions for each dataset:

1. Are your files in an open file format? Or, is the code in an open programming language?
2. Can you convert the proprietary file formats to an open file format? If yes, to which open file format?
3. If the data/code is in a proprietary format, what information/software would others need to re-use the data/code?

## Step 6 - Reflect on the theme Access

**Access** - It is important to reflect early in your project about who will be accessing the datasets you collect/create, when they will access the data in the different steps of the project and how. Also, before preparing your data for publication, you need to reflect if it is possible/allowed for you to publish the data. You might need to process some data before you can make them available, especially if you are working with confidential data.

Reflect and answer these questions for each dataset:

1. Who will have access to this dataset during the project?
2. If others than you will have access to the dataset during the project, how will you share the data?
3. At the end of the project this dataset can be:
  1. a. 'open' - no confidential information in it , it can be published
  - b. 'restricted access' - there is confidential information in it, can't be anonymize, can't be published
  - c. 'restricted access with public metadata' - there is confidential information in it, but the description of the dataset can be published
4. Go to the toolbox and use the action arrows to indicate if you need to process the data before publishing them

## Step 7 - Reflect on the theme Data Publication

**Data Publication** - this theme needs to be worked out based on the previous theme 'Access'. Depending on your reflections on accessibility answer the following for each dataset:

Depending on your reflections on accessibility answer the following for each dataset:

1. If the dataset is marked as 'open', in which repository would you publish it?

2. If the dataset is marked as 'restricted access' how can somebody request the data after you finish your project?
3. If the dataset is marked 'restricted access with public metadata' where will you publish the metadata and where will the data be stored?
4. Do the repositories you plan to use provide a DOI for the dataset? Does it allow you to provide a licence? Which licence would you use?