Before explaining what we do, we first visit a Dystopian world, unfortunately not too far from the truth in some cases. https://www.youtube.com/watch?v=N2zK3sAtr-4

Although this is brought as a bad example, there are surely parts you recognize. I certainly recognize it.

We want to be able to reuse data for a longer period. It needs data management, or data stewardship.

Data Stewardship vs Data Management naming pitfalls: “data management” is sometimes considered to end at project end. “data stewardship” is sometimes considered to start at project end.

Data Stewardship requires good Data Management.
We call it Data Stewardship, and for us this includes Data Management. For me it is any operation that handles digital data. Some others include data on paper and/or samples.

First let's have a closer look at the word reusability.

Our project work should aim to make the data Reusable.
- Reusable for others
- Usable and Reusable for ourselves
  - Now usable: data is complete and well described. You don’t need to search for all the metadata at the end, but properly maintain it.
  - Later reusable: because you won’t remember when the reviewer asks, and because you struggle to reproduce
  - Even later reusable when your postdoc is replaced by a successor

Reusable
To make our data re-usable, we need to make it Findable, Accessible, Interoperable. It is like a four step ladder.

And thinking about this really helps also making decisions in a project.

I would like to exploit common genotype-phenotype relations between Alzheimer’s Disease and Huntington’s Disease… I need to combine AD and HD data… I can help with that!

What does it mean to be re-usable? Interoperability is not something that can be added as an afterthought. It should be planned for, that really saves time.

Data management is the answer to a complicated question, and is not easy! Needs thought.

Data interoperability is more than technical standards: extensive meta data are needed.
DTL DEFINITION OF DATA STEWARDSHIP

Responsible planning and executing of all actions on digital data before, during and after a research project, with the aim of optimizing the usability, reusability and reproducibility of the resulting data.

This is the reason for the need for the word “planning” in the definition of data stewardship.

FAIR Data Stewardship Planning

Let’s first talk about the timing of the DMP activity.

Different funders have different ideas about when a DSP should be made. Some do it before the project starts, but that is hard because that activity is unfunded; it discounts how much work is actually going into good planning. Some ask it as a deliverable, but that makes the budget really hard in some cases. The best way is to start planning before the project proposal, but continue the DMP in the project.

To be able to deal with changes in the project, the data stewardship plan should be updated all along with the project. That way it will also function as part of the documentation of the data after the project is finished.

Big project? Hire a data expert, one of the 500k the EOSC calls for. A data steward should know a lot, and also know where to find the experts in your institute that can help further.
In preparing for battle I have always found that plans are useless, but planning is indispensable

Dwight D. Eisenhower

DMP’s change over time.

What can we learn from warfare? It is normal that Plans change.

Alternative: “no battle plan survives contact with the enemy” by Helmuth von Moltke
Or: When your plan meets the real world, the real world wins

Remember that no matter how carefully you plan your work, reality will always be different from what you expected.

To get the most value out of your data and to run the least data risks, all of these need to be arranged.

All of these are different for each project. All of these benefit from planning. Call upon experts.

I will very quickly pass examples.

* Data needs to come into a repository
* Special = Made for the project. Note Accessibility!
* Different repositories for different data?
* Preferences from funder or institute? Budget affected!
* Go talk to repositories! They may require format, metadata. They can help!
  * Persistent identifier like DOI, Help you get credit
* Especially institutional or special; catalog it!
* Keywords for re-use. What is in there, not why you collected
* Keywords for subsets?
* Librarians/archivists can help! Go look for them!
* Talk to the catalog people early to find out what they need.
ACCESSIBLE

- Assure Longevity
- Check Legal Conditions
- Limit the Embargo
- Ensure proper ICT procedures

Longevity: 20% decay per year on links to web sites. Certified repositories! Transfer if they stop.

Software longevity
Repositories often work with one-time-fee.
Privacy sensitive data: Not coupled to person! Clear rules! Find a data access committee
Non-legal reasons: Limited time! not “once we get the patent” or “after the publication”
ICT procedures: convince reviewer that you won’t lose (track of) the data. Professional?

INTEROPERABLE

- Format
- Terminology
- Sufficient metadata

Format
Terms: (controlled) vocabulary. Or an (international) coding system. With relations: Ontology.
Examples after click
Use what others use, or map
All fields: even if you do genetics, store sample locations like climate scientists.
There will be lots of development, demands on interoperability and benefits will grow!

Example
1854 London Cholera outbreak, John Snow made a map and identified a water pump that was the origin of the infection
Arabidopsis study where correlation was made with height through google maps API.

You need sufficient metadata to be able to interpret the data

REUSABLE

- Document provenance
- Use minimal metadata standards
- Choose a liberal license
- Make sure results are not only narrated

Annotate to help self and others, avoid clarification requests
How was data obtained, provenance. Instrument settings. Automate collection! Helps yourself!
Minimal metadata. Quite extensive, and volunteer to do optional too. Field specific+DC.
License. Are you obliging people to Cite? Forbid commercial use? Such clauses have Consequences. Unnecessary restrictions!
See the example of the Gauss curve. Which nobody cites any more!
Ewan Birney insight: other scientists will cite, they need to prove they got data from a reliable source!

Do not encode knowledge only in narrative. Please help to make text mining obsolete.

Use experts, be experts.
Making the actual DSP requires that you think about it. You can’t simply copy a DMP from somewhere. How do you manage to make your data FAIR?
Some aspects can be more difficult than you imagined. Especially where it is far from your own specialty.

You may have a well maintained photo library at home. One that you are proud of. But: Maintaining data in the lab is more complicated than a photo library. (1) It is larger (2) It is varied (3) It is operated on by different people
And even in a photo library you are sometimes looking forever.

Regarding this photo library: everybody makes the same mistakes.
You should really make use of the expertise of others. Look for the experts!

Broad expertise is needed: from (library, IT) and from field experts
Not all the data related expertise may be in your existing network

Service providers make a list. Does that work? No, Lists only solve the easiest problem: finding back something you already know that exists.

It is irritant if you don’t know where to find the expertise. Lists solve that.
It is painful if you don’t know it exists. Lists don’t solve that.
If is dangerous if you don’t know you need it. Lists don’t solve that.

Case in point: How do you get your apps on your phone? You ask a friend, an expert.
I think I can help in another way.
With a tool that we are developing in ELIXIR. A tool that behaves like an expert.

In ELIXIR-NL we work together with ELIXIR-CZ on a tool that can guide researchers to good data management: the wizard. [https://dsw.fairdata.solutions/](https://dsw.fairdata.solutions/)
The wizard opens questions that are relevant. It attempts to ask the first few questions a broad set of experts would ask and point to more information.
It opens relevant questions [only]
It contains more detailed guidance on some topics.
It contains links to Barend’s book pages.
It measures the FAIRness of the result per chapter.
The questions in the questionnaire are based on a mind map I have been collecting over the last 5 years in interviews with experts.

For each answer in each question, the system also encodes six objectives.

**COMMUNITY AND TRAINING ACTIVITIES**

- DTL/ELIXIR-NL:
  - FAIR trainings (for researchers and data stewards)
  - "Bring Your Own Data" meetings (F/A BYOD)
  - FAIR Hackathons (F/A BYOD)
  - Data Stewards Interest Group (increasingly international)
- ELIXIR DMP Implementation study
  - Inventory of available materials
  - Trainings
  - Establishing the Data Stewards job description

See also [https://www.dtls.nl/](https://www.dtls.nl/)
Proper Data Stewardship makes and keeps data FAIR, during and after the project.

- (Researchers do not want to make DSP? Do they want to make Papers? Both belong to science!)
- (Requirements on FAIR will grow, FAIR is a good start)
- FAIR approach to DMP will not only benefit others, but also yourself.
- At the end of the project, the DSP has become a map of the road you walked.