

Butterfly Data Management Plan Meeting Minutes

Action items are highlighted

Date: Aug 5, 2025 (Tuesday)

Start Time: 14:00 CEST / 06:00 (MDT)

Attendees:

Laura Elisabet Drivdal (LD), Nilgun Kulan (NK), Lorraine Balaine (LB), Jeroen van der Sluijs (JvdS), Jorrit Poelen (JP)

Absent: Mirella, Cala, Nicholas.

Link to DMP:

<https://docs.google.com/document/d/1eRsXw-bSWmz9mI1v1HHKdJez11sbjTsBtWsKnAdkDF0/edit?tab=t.0>

Next Meeting: TODO

- TODO

Agenda (Laura D.):

* review current draft DMP

Notes:

Laura is planning to add previous meeting notes to this document.

Laura submits DMP to reviewers (Ivan Meeus, Kjersti Hasle Enerstvedt) on Friday 8 Aug 2025

Jorrit shared a <https://globalbioticinteractions.org/euroappa> page on which the DMP draft and meeting notes are linked and versioned. Jeroen noted that we should check whether the DMP and associated meeting notes are suited to be openly accessible.

The attendees went over the DMP draft and had various discussions updating the draft accordingly.

Reviewers will be asked to provide feedback within 10 days (Monday, August 18).
Nilgun will send a poll to find the next meeting day within that week (Aug. 19-22).

[insert July meeting notes here]

Date: June 27, 2025 (Friday)

Start Time: 14:00 CEST / 06:00 (MDT)

Attendees:

Laura Elisabet Drivdal (LD), Nicola Gallai (NG), Nilgun Kulan (NK), Lorraine Balaine (LB), Jeroen van der Sluijs (JvdS), Jorrit Poelen (JP)

Absent: Mirella, Cala, Nicholas.

Link to DMP:

<https://docs.google.com/document/d/1eRsXw-bSWmz9mI1v1HHKdJez11sbjTsBtWsKnAdkDF0/edit?tab=t.0>

Next Meeting: August 4-8, 2025

- Polling: <https://framadata.org/NgnXI51S1sJMoJ2J>

Agenda (Laura D.):

Action Items

Action items are highlighted

Attendees:

Laura Elisabet Drivdal (LD), Nicola Gallai (NG), Nilgun Kulan (NK), Lorraine Balaine (LB), Jeroen van der Sluijs (JvdS), Jorrit Poelen (JP)

Absent: Mirella, Cala, Nicholas.

Date: June 27, 2025 (Friday)

Start Time: 14:00 CEST / 06:00 (MDT)

Link to DMP:

<https://docs.google.com/document/d/1eRsXw-bSWmz9mI1v1HHKdJez11sbjTsBtWsKnAdkDF0/edit?tab=t.0>

Next Meeting: August 4-8, 2025

- Polling: <https://framadata.org/NgnXI51S1sJMoJ2J>

Agenda (Laura D.):

1. The structure of the DMP. In discussion with Jorrit, we thought the first DMP should have more focus on the data management process and less on the data products. I added some on that in the introduction, but it could be expanded and could be an own chapter, together with the model suggested by Jorrit.
2. Table 1 and 2 will not be included in this DMP, but we could use it for version 2?
3. We need to work more on chapter 3 (on the FAIR principles), eg making data interoperable. The setup of subchapters is a little frustrating, there are many overlaps, so maybe we should merge all the subchapters of 3 into one?
4. **Important:** I will be on holidays week 27 and 28, could anyone of you continue to work on the document in the meantime?

Opening

- Laura noted absences: Mirella and Cala indicated that they could not join today.
- Nicola mentioned technical difficulties with his video, but audio is fine
- Jorrit has arrived a little late as he lives in a time zone 8 hours behind. It is only 6:00 am for him.

Discussion

- LD: This is going to be the 1st version of DMP. Process is more important than what type of data we have and what type of data we will have. This DMP needs to detail how we will do the data management.
- JP: First talk about the structure and where we intend to publish the data. How to review the data (internally and externally).
- NG: To share data with everyone, they need to be structured in a way that is understandable by everyone.
 - Should one person compile a table with all the data (individuals give the data to the compiler)?

- LD: It should not be one person's responsibility. Maybe task leaders. **We need to clarify this point in the DMP, re: who has what responsibility.**
- NG: Then we need rules when we (all of us) create data tables about how to create them.
- JP:
 - This discussion reminds him another project (name?) where data sharing was limited (even for metadata), but data were successfully reviewed. Reviewing is not the same as publishing the data. Share-screen in a meeting would work for reviewing.
 - DMP is a coordination rather than a central repository. Make a table of data where people will store their data.
- NK: Storage issue. We are going to collect/work with lots of different types of data. It is difficult to come up with a standard way to store various types of data. Are there global databases already setup in your field of study? Are there established standard names for variables?
- LB: Not really. "Willingness to pay" is always called that, but no standard name for every variable.
 - More common is to make a list of variable names in the dataset when published. This is easy for quantitative data. Qualitative data, e.g., interviews with agricultural students are more complicated. They can be published, wholly or partially, after removing the bits that may lead to the identification of the interviewee.
 - **Metadata is the common base for all the data we will work with. We can make rules for the metadata and be descriptive there.**
 - Agree with the point on external review. Data should be made understandable / usable by everyone not only by experts in the field.
- NG:
 - Before data is written down, it is a recording. It will be transcribed from an audio recording. Not simple. It can also be a video. **These types of data should be mentioned in the DMP.**
 - Another type of data: National survey by a company. The company will give the results in a giant data table to Nicola's group. 1 country 1,000 people survey. 1 person answers 15 scenarios, etc. Multiple countries...
 - Not easy to share such a huge file! **This must be addressed in the DMP.**
- LD: For biological and ecological data FAIR principles are easier to address in the DMP. But, little information in the human participant data section in the DMP now.
- NG, LB: Cannot be more precise about the data right now, before starting the survey.

- JP:
 - This (meeting) is high level of sharing and getting to know who is working with what type of data.
 - Make a list of all the various data we are interested in with a very short description and names of people who will be dealing with that data.
How do we do this?
 - Then we can start finding tasks associated with it and reviewers.
- LD: Butterfly is a transdisciplinary data, we need to understand each other's data. **Internal reviewing of data would facilitate this.**
- JP: Tables and files are common storage media. Start a conversation about what data go in them, rather paying attention to the format.
- NG:
 - WP3 will work on macroeconomics model by using both ecological data + social data. For this alone we need to create a table about types of data.
 - Ali and Georgios also have incentive in creating a table with all the data they need for their modelling.
 - **NG is also part of WP3. They should discuss how to create this data table.**
- JvdS:
 - **Transdisciplinary is an important aspect of Butterfly project. It can be brought in by internal review.**
 - So, if we have ecological data, there should be at least one social or humanities scholar in the internal review. To make sure data is understandable to others outside the experts' field.
 - Need to agree on a procedure for both the internal and external review.
 - Need also some requirements about how the data sets are documented and versioned, etc.
 - We should also store our data in multiple places / repositories.
 - Who is responsible for what: collecting, storing, reviewing, uploading to repositories, etc.
 - **How much space we need? 10 terrabyte???** Probably too much... Depends on the type of data, e.g., videos.
- LB: We are not allowed to share videos anyway. Just the metadata.
- NK:
 - We have 1 TByte space in Butterfly Next Cloud.
 - Maybe what we need first is a table like Jorrit mentioned. A list of types of data and who will be working with that data.
- LD:
 - Tried to categorize data in the DMP. Biological and ecological data are different from each other according to Jeff. Now separated.

- Human data with econ data – is it okay?
- WP6 data is very confusing.
- WP7 and WP2 interconnected. Econ, landscape use data. How to deal with them?
- Made another table for REUSED data.
- NK: Should we get advice from people in these tasks first?
- LD: Some gave advice already.
- LB:
 - Some of the work we do in WP2 can fit in multiple categories. Are we making it too complicated by dividing, subdividing, etc.
 - Maybe we just need to agree what should be in the DMP. If this much detail is needed, then yes, but putting so much detail in the deliverable at such an early stage in the project maybe limiting.
 - Data collectors who need ethical approval and informed consent. Need for anonymization. Nicola made a file about that which can be linked here. But it is necessary?
- JP:
 - Notice that we are eager to go into detail. But let's start with a discussion on what to put in the DMP.
 - Different datasets will have different approaches and ways of collecting and storing data.
 - Rather than trying to understand all the different data types, let's come up with a framework. For example, we will meet every month (?) and review the data, update our data registry.
 - Learn from each other. More active knowledge sharing instead of a fixed DMP.
 - How do other projects do it?
- LD: Many different ways, as far as we know from other projects. They all say the standard things like, "we will make it interoperable", "we will put our data on Zenodo", etc.
- JP:
 - This is a group skill to develop during the course of the project.
 - What is a good dataset may differ from person to person.
- LD: **We should highlight this co-creation / review process in this first version of the DMP.** See notes (comments in the DMP document).
 - Moving on to the other agenda points.
 - End of the DMP document is repetitive and does not make sense. Hard to write.
- JP: It is like a checklist of "did you think about this, did you think about that, etc." Useful if we review things regularly and update. Hard to do upfront.
- LD: It is just a checklist to make at the beginning and forget about later.

- NK:
 - Maybe we should resolve to create a brief DMP at this point and say the things we are expected to do to make our data FAIR. Then add that we will have monthly or 3-monthly regular data meetings to review and update our DMP.
 - Next DMP version, delivery date Month 24. But we do not have to wait until then to internally update our DMP.
- LB:
 - Good to **make a point of monthly or quarterly meetings in the DMP.** Say that all these processes will be decided as data is collected.
 - Lorraine was asked (in an ethics approval application she made) to **hold workshops with partners to outline their responsibilities.**
 - In terms of reusability of the data, we will ensure **human participants will be asked to give informed consent about the use of their data in the future / future projects.**
 - **Concent form: LB to add directly or send it to us by email.**
 - External data review: Publishing in a peer reviewed journal would count as external peer review? Otherwise it might be difficult to find external reviewers for our data.
- JP:
 - In the diagram we put in the DMP, we thought external review can be as informal or formal as possible.
 - Could be as simple as showing it to a friend.
 - Do journal reviews and see lots of shortcomings in the datasets.
- LD: Will be away for two weeks.
- NK:
 - The deadline for the first version of DMP is Month 6: August 31, 2025.
 - We can treat this 1st version as a clean, presentable draft DMP and in it say that we will update it every 3-months internally.
- LD:
 - **Jeroen (Nilgun) to work on the DMP while LD is away.**
 - **Nicolo to complete the data registry table** (two example datasets, one highly regulated, one easier. See JP's comment below).
 - **Anyone can add to the document** (link above).
- JP:
 - For the August 31 deadline, we can not complete but add two examples: *"... one example with data that is highly regulated... and one data set that might be a little bit less so."*

- Curious to see how a highly regulated (personal data) data set looks like as a knowledge sharing activity.

Next Meeting

- Fill in the poll to find the best date in the **first week of August** (please answer before **July 11, Friday**):

<https://framadate.org/NgnXI51S1sJMoJ2J>