

Project Based Learning: use cases.

Project 4: The evolutionary origin Altai Falcon

This document serves as the basis of an exercise that spans several sessions of the training course. Based on the use case that is described below, you will dive into some details of research data management; we aim to tackle basic FAIRification, documentation, archiving as well as some research integrity issues concerning data.

The FAIR principles are guidelines that together represent an 'ideal', and researchers can make small steps towards 'FAIRness': the aim is to explore which steps for further FAIRification you could take with these data, not to judge previously performed work.

Project Description:

You are writing a grant proposal to perform a study on the evolutionary origin of the Altai falcons from Central Asia. This large-bodied and mighty bird of prey with a rare and unique look has always attracted the attention of humans. After decades of illegal harvesting for falconry, the Altai falcon is on the brink of extinction and the aim of your project is to contribute to effective conservation priorities and strategies based on a genomic approach to investigate the genetic status and phylogenetic position of this species.

In this study you propose a genome-wide approach, Restriction-site Associated DNA sequencing, using sympatric eastern sakers falcons, allopatric western saker falcons and gyrfalcons as outgroup to explore the evolutionary relationships between the the Altai falcon and other populations.

A data management section is a mandatory part of your research proposal; in this section, you will

consider in advance how you will manage the data that this project will generate and to plan which data will be preserved and made publicly available. Whilst doing this, you should also consider costs for your data management. These costs are eligible for funding, if you include them in the project budget under the budget module 'material budget'.

Assignment 1: FAIRness of data

You can find a dataset related to this project at:

<https://doi.org/10.48428/ADATTAR/UUXM4D>

Take some time to have a look at the dataset. Can you understand the structure of the files and the documentation?

In the session, we showed you the FAIR Aware Tool; it provides a list of questions that guide you and help you decide on what you need to do to make data FAIR,

<https://fairaware.dans.knaw.nl>.

For this assignment we made some adjustments to the questions of the FAIR Aware tool, so that we can use them to assess the FAIRness of data (instead of your level of awareness).

Can you apply the questions taken from the FAIR Aware tool to the above dataset and answer them in the table below in order to assess the FAIRness of the dataset? If you need additional information, please check the online tool and click on the 'i' icon:



FAIR principle	Question	Yes/No/Not sure
Findable	Has this data(set) been assigned a globally unique persistent and resolvable identifier when it was deposited in the data repository?	
	Does the data(set) have discovery metadata that make the data(set) findable, understandable and reusable to others?	
	Are the metadata describing your data(set) available in a format readable by machines as well as humans?	
Accessible	Is access to this data(set) controlled and do the metadata include licence information under which the data(set) can be reused?	
	Do you think that metadata will remain available over time, even if the data(set) is no longer accessible?	
Interoperable	Do the metadata describing this data(set) use controlled vocabularies?	
Reusable	Is provenance information about the collection and/or generation of data included in the metadata?	
	Do the metadata describing this data(set) follow the specifications of a community-endorsed standard?	
	Has this data(set) been deposited in a file format that is open and supported by the data repository for long-term preservation?	
	Do you think that keeping your data(set) FAIR over time requires professional data curation and digital preservation?	

Please note: It's very likely that you answered 'not sure' for some (or many) of the questions above. Don't worry! These topics will be covered in depth in the later sessions of this learning program

Assignment 2: Project DMP

Writing a data management plan can help you to create a structure for all the activities that are needed to handle the research data in a project.

In the table below you will find some questions that you could expect to find in a DMP. This is, of course, a simplified structure. The actual amount or order of the questions varies per DMP.

Please provide answers to the questions below:

1. Data collection	How do you collect or generate your data? Do you (also) use data that have been produced by somebody else? What method do you use to create your data? What tools, software or other equipment do you use?
<i>[please provide your answer here]</i>	
2. Data storage and security	Where do you store your data? What do you do to make a backup of your data? Who has access to your data? Do your data need extra protection?
<i>[please provide your answer here]</i>	
3. Data documentation	What documentation do you provide to make your peers understand your data? Are there any documentation standards in your field of research? What documentation do you need to provide to the repository to make your data Findable?
<i>[please provide your answer here]</i>	

4. Data sharing	With whom can you share (which part) of your data DURING the project? With whom can you share (which part) of your data AFTER the project? Do you have any restrictions on sharing? If so, what are the reasons for those restrictions?
<i>[please provide your answer here]</i>	
5. Data preservation	Which data do you need to archive? Do you have a repository that is prescribed by your institution? Do you yourself have a specific repository in mind? Do you need to spend a lot of time to prepare your data for archiving? Who is responsible for the data after the project?
<i>[please provide your answer here]</i>	

Assignment 3: Project documentation

You can find the metadata for the dataset at:

Datacite metadata are available at:

<https://adattar.unideb.hu/api/datasets/export?exporter=Datacite&persistentId=doi%3A10.48428/ADATTAR/UUXM4D>

Other metadata are in the attached file.

To assess the completeness of the metadata, please review the attached file and the metadata that you find in the repository.

Question	Your answer
Do you think that the information in this record and in this file is sufficient to understand the data(set)?	
Is the documentation understandable?	
Can you add at least 4 elements to the metadata to improve the usefulness of this file for others?	

Assignment 4: Data Repository

This dataset was deposited in

Adattár, the University Institutional Dataverse Repository, <https://adattar.unideb.hu>

Let us try to have a closer look at this repository and its FAIR enabling features. If you do not know the answer to a question, we can address this in the 5th session. Use Google Translate (or comparable AI tools) to help you go through the Hungarian instruction in the repository.

This page may give you an impression of the underlying system, Dataverse,

<https://guides.dataverse.org/en/5.14/user/> and we also recommend you look at the record for this repository in Re3data, the Registry of Research Data Repositories:

<https://www.re3data.org/repository/r3d100013804>

Question	Your answer
- Is the repository certified? Can you find any information on this? - Does this repository accept the file (formats) that you use? - Does the repository assign persistent identifiers to datasets - Does the repository adhere to one or more metadata standards? Which ones? - What licenses does this repository apply?	
If you consider the subject and scope of the study, could you name three more repositories and provide an argumentation why those are suitable alternatives.	

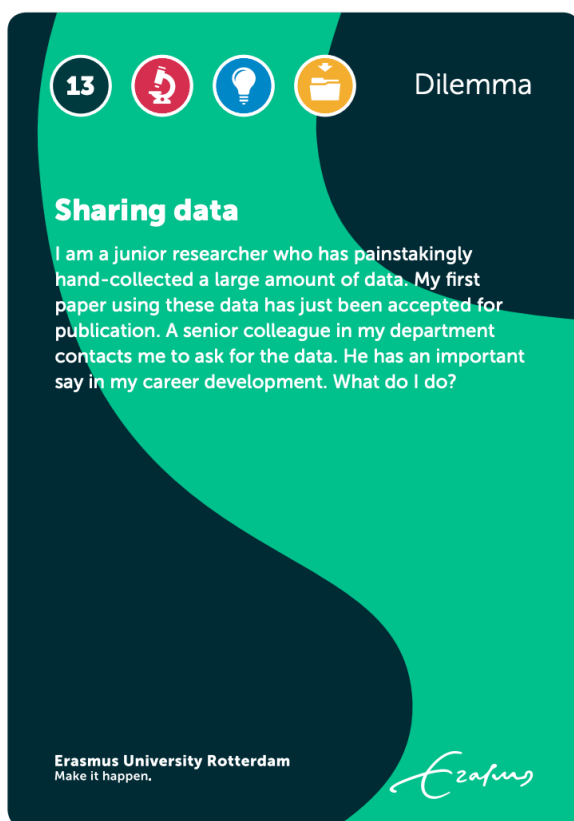
For inspiration, take some time to browse through the Registry of Research Data Repositories <https://www.re3data.org>

Assignment 5: Policy and practice

Independent from your subject background or the project that you have chosen in this course, we would like to draw your attention to some issues that could hamper your research and FAIR data sharing practices.

The Dilemma Game is a fun tool, developed at Erasmus University, to lower the threshold to discuss serious real life research integrity issues. The game was originally a card game, but you can also download and use an app to play it in an online setting.

In this exercise, we ask you to reflect on 5 dilemmas: 13, 16, 35, 43, and 59 and to discuss them in breakout groups.



The card features a dark teal background with a large, light teal abstract shape on the left. At the top, there are four circular icons: a green circle with the number 13, a red circle with a microscope, a blue circle with a lightbulb, and an orange circle with a folder. The word "Dilemma" is written in white at the top right. The title "Sharing data" is in bold white text. The text of the dilemma is in white. At the bottom left, the Erasmus University Rotterdam logo and tagline are present. At the bottom right, the Erasmus University script logo is visible.

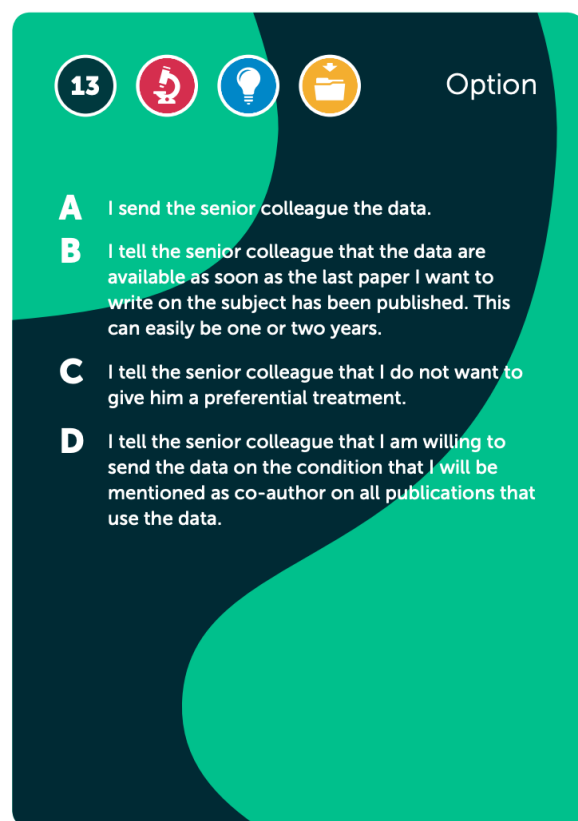
13 **Dilemma**

Sharing data

I am a junior researcher who has painstakingly hand-collected a large amount of data. My first paper using these data has just been accepted for publication. A senior colleague in my department contacts me to ask for the data. He has an important say in my career development. What do I do?

Erasmus University Rotterdam
Make it happen.

Erasmus



The card features a dark teal background with a large, light teal abstract shape on the left. At the top, there are four circular icons: a green circle with the number 13, a red circle with a microscope, a blue circle with a lightbulb, and an orange circle with a folder. The word "Option" is written in white at the top right. The title "Option" is in bold white text. The text of the options is in white. At the bottom left, the Erasmus University Rotterdam logo and tagline are present. At the bottom right, the Erasmus University script logo is visible.

13 **Option**

A I send the senior colleague the data.

B I tell the senior colleague that the data are available as soon as the last paper I want to write on the subject has been published. This can easily be one or two years.

C I tell the senior colleague that I do not want to give him a preferential treatment.

D I tell the senior colleague that I am willing to send the data on the condition that I will be mentioned as co-author on all publications that use the data.

Erasmus University Rotterdam
Make it happen.

Erasmus

16

Dilemma

Data check

My data is archived in a database, stored in the central data archives of the university. As part of a routine integrity check by the university, I am requested to provide information on my respondents to ensure that these are actual existing people, and that there is no fabricated data. How do I respond to the request?

Erasmus University Rotterdam
Make it happen.

16

Option

A The confidentiality of my respondents is paramount. I do not comply, and explain this in my response to the request.

B I provide some access to the identity of my respondents, in such a way that they cannot be traced back to the observations in the data archive.

C I provide full access to the identity of my respondents, but I request a signed confidentiality agreement of the person making the request.

D I provide full access to the identity of my respondents. Showing data collection integrity supersedes respondents' confidentiality.

35

Dilemma

Anonymous data

As a researcher I participate a lot in research projects funded by business. One of the explicitly stated wishes of the directors of a company is that anonymity of the persons interviewed is ensured as the research is on a sensitive issue. I have stated that the anonymity is ensured, but we have no written agreement. While writing the report and discussing the results it becomes clear that maintaining anonymity will make it impossible to verify the results. This can jeopardize the chances of the article being published. What do I do?

Erasmus University Rotterdam
Make it happen.

35

Option

A I respect the requested anonymity and leave it like that.

B I allow reviewers and other researchers access to my data.

C I ask the director to agree to limited anonymity. If he does not agree I will maintain full anonymity.

D I notify all participants about the use of their personal data.

43



Dilemma

Smart use of data

One of my PhD students came to me yesterday to tell me that she had been accused of stealing data by writing a paper based on data she and two other PhD students had collected. After questioning her about the situation I find out that the data she used was collected by all three together in the context of another research paper that they co-authored. What do I do?

Erasmus University Rotterdam
Make it happen.



43



Option

- A** I tell her that it is a smart move. She was the first to write the article so she can use the data.
- B** I tell her that that she should not write the article on her own and should apologize to her colleagues.
- C** I tell her that the situation is one they have to resolve themselves.
- D** I point her to the applicable code of conduct and I let her decide whether she can continue writing the article or not.

59



Dilemma

Spoilsport

I am using data from a widely used data source within my institute. While processing the data, I come across some systematic problems (missing values, outliers) that apparently nobody has ever bothered about before. Remedying the error accurately would take me half a year. My supervisor suggests following "common practice", without specifying. Common practice is not to report the problem. Alternative sources are not readily available. What do I do?

Erasmus University Rotterdam
Make it happen.



59



Option

- A** I take extensive time to analyse the problems, even if that implies that my PhD will be delayed.
- B** I go to the head of the institute and ask for an investigation into past and current research based on the data set. The results might be problematic.
- C** I change the scope of my research project so that I no longer have to use the data.
- D** I contact researchers who published earlier on the database. If they agree with the supervisor I follow common practice.