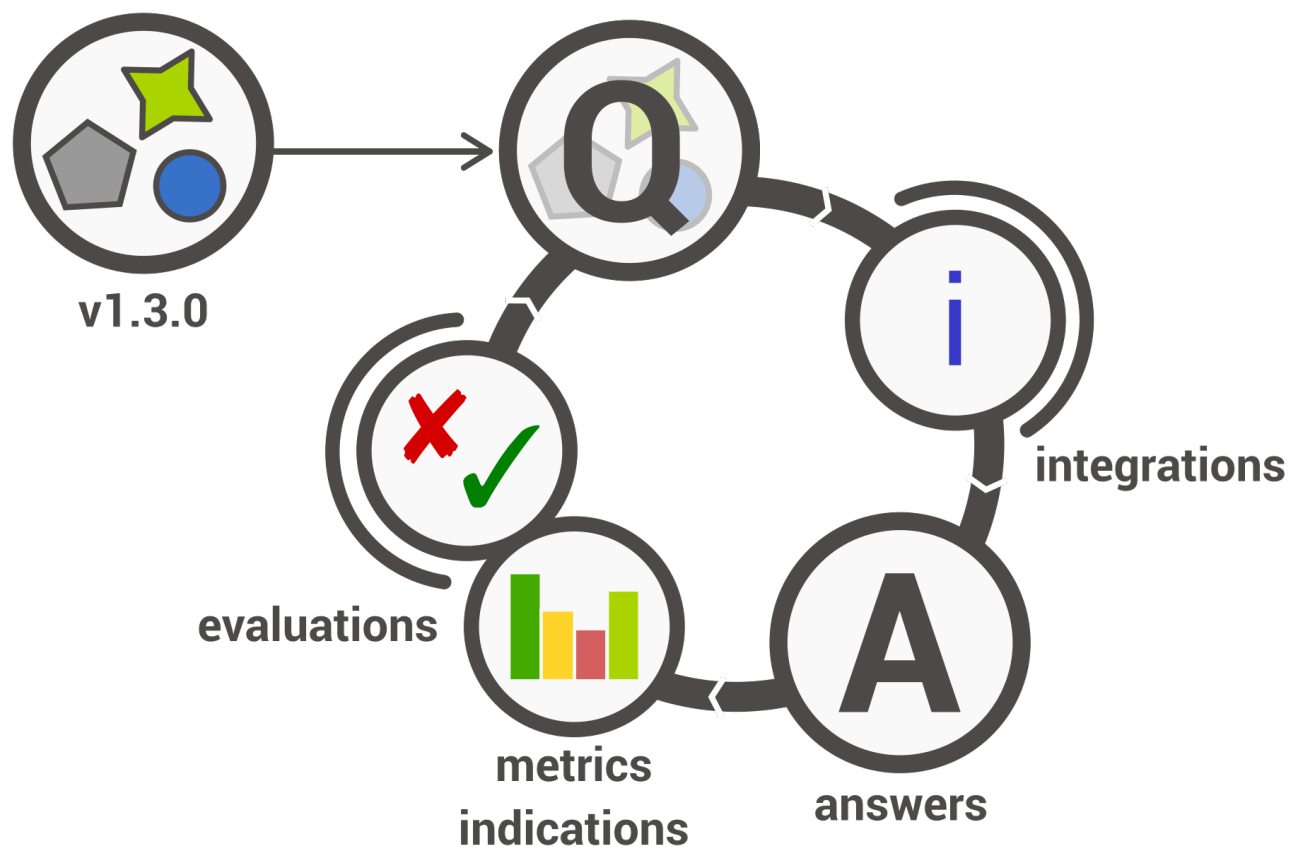




Data Stewardship Wizard



Korbinian Bösl
Data management coordinator
ELIXIR Norway/Digital Life Norway
10th of March 2021



Developed by



elixir-no.ds-wizard.org

Filled questionnaire → **Template** → **DMP in various formats**
once per funding body .docx, .tex, .html, .json

elixir-no.ds-wizard.org

Filled questionnaire → **Template** → **DMP in various formats**
once per funding body .docx, .tex, .html, .json

Full compliance:



elixir-no.ds-wizard.org

<https://ds-wizard.org>
<https://github.com/ds-wizard>

Filled questionnaire → **Template** → **DMP in various formats**
once per funding body .docx, .tex, .html, .json

Adapted for Norwegian users



elixir-no.ds-wizard.org

Full compliance:



Filled questionnaire → **Template** → **DMP in various formats**

once per funding body .docx, .tex, .html, .json

Adapted for Norwegian users



Login:
 **Feide**



Full compliance:

SCIENCE
EUROPE 



elixir-no.ds-wizard.org

Build your DMP by following a decision tree with defined questions

elixir-no.ds-wizard.org

1 Is there any pre-existing data?

Horizon 2020 DMP Science Europe DMP maDMP

Are there any data sets available in the world that are relevant to your planned research?

☒ Desirable: Before Submitting the Proposal
☒ Data Stewardship for Open Science: [atg](#)
☒ External links: [Google dataset search](#), [Datacite Search](#)

☐ a. No

☒ b. Yes

Clear answer

1.b.1 Will you be using any pre-existing data (including other people's data)?

Horizon 2020 DMP Science Europe DMP maDMP

Will you be referring to any earlier measured data, reference data, or data that should be mined from existing literature? Your own data as well as data from others?

☒ Desirable: Before Submitting the Proposal
☒ Data Stewardship for Open Science: [ezi](#)

☐ a. No

☒ b. Yes

Clear answer

1.b.1.b.1 What reference data will you use?

Horizon 2020 DMP Science Europe DMP


Much of today's data is used in comparison with reference data. You may be comparing your own data with a "standard set" which is maintained as a collection by someone else. Or you could be determining differences to a standard (for example in bioinformatics, a genome is often compared with a reference genome to identify genomic variants). If you use reference data, there are several specific issues that you should consider. What are the reference data sets that you will use?

☒ Desirable: Before Submitting the DMP
☒ Data Stewardship for Open Science: [dgg](#)

1.b.1.b.1.a.1 Reference database or dataset

Horizon 2020 DMP Science Europe DMP

Ensembl

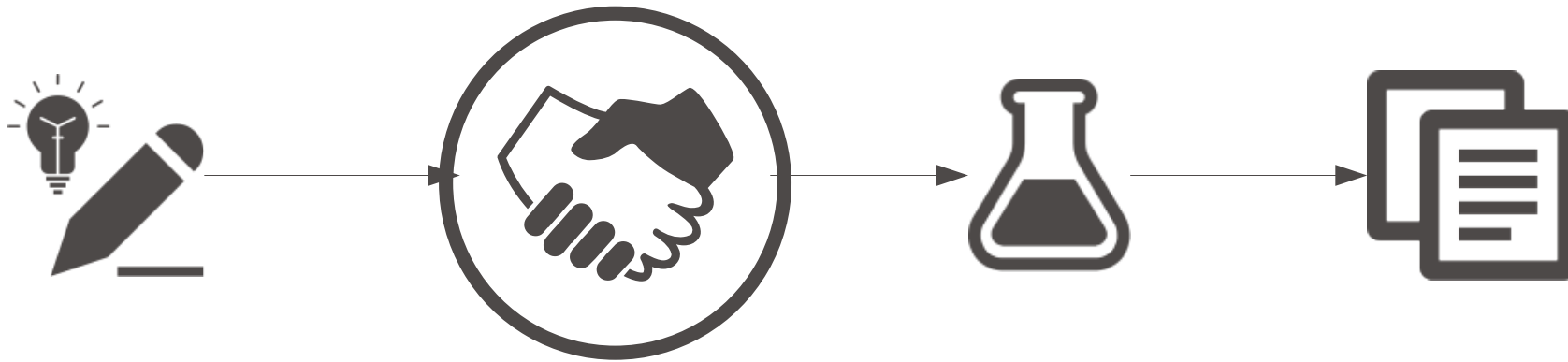
 <https://fairsharing.org/bsg-d000330>

Give the name of the database or dataset. You will be shown suggestions of data bases from FAIRSharing, but you can also type the name of a dataset that is not in FAIRsharing

☒ Desirable: Before Submitting the DMP



Forskningsrådet



DMP





Horizon 2020 → Horizon Europe

Open Research Data Pilot (ORD pilot)
Default since 2017 → mandatory 2021



DMP
outline




DMP





- After 6 month
- Update upon significant changes








 DSW ELIXIR-NO


 Knowledge Models


 Projects


 ELIXIR Norway


 How to - video


 How to





 Share


 Questionnaire

 TODOs

 Metrics

 Preview

 Documents

 Settings

Current Phase

Before Submitting the Proposal

Before Submitting the Proposal

Before Submitting the DMP

Before Finishing the Project

Ch

I. Administrative information1

II. Re-using data1

III. Creating and collecting data6

IV. Data management plan8

1 Will you be using a shared working space to work with your data?

Horizon 2020 DMP

Science Europe DMP

Will you be using a working space that is shared between all the people working on the data in the project? Sometimes such a system is called a *Virtual Research Environment*.

☒ Desirable: *Before Submitting the Proposal*

☐ a. No

elixir-no.ds-wizard.org



DATA STEWARDSHIP WIZARD

1.b.1 Will you be using any pre-existing data (including other people's data)?



Will you be referring to any earlier measured data, reference data, or data that should be mined from existing literature? Your own data as well as data from others?

☒ Desirable: *Before Submitting the Proposal*

Data Stewardship for Open Science: [ezi](#)

☐ a. No

☒ b. Yes

Clear answer

1.b.1.b.1 What reference data will you use?



Much of today's data is used in comparison with reference data. A genome for instance is compared with a reference genome to identify genomic variants. If you use reference data, there are several other issues that you should consider. What are the reference data sets that you will use?

☒ Desirable: *Before Submitting the DMP*

Data Stewardship for Open Science: [quc](#)

1.b.1.b.1.a.1 Reference data:



Banana Breeding Tracker Database

<https://fairsharing.org/bsg-d001258>

☒ Desirable: *Before Submitting the DMP*

1.b.1.b.1.a.2 Do you know where and how is it available?



Do you know where the reference data is available, what the conditions for use are, and how to reference it?

☒ Desirable: *Before Submitting the DMP*



elixir-no.ds-wizard.org

ds-wizard.org
github.com/ds-wizard

1.b.1.b.1.a.1

Reference database or dataset

Horizon 2020 DMP

Science Europe DMP

Banana Breeding Tracker Database

FAIRsharing

<https://fairsharing.org/bsg-d001237>

Give the name of the database or dataset. You will be shown suggestions of data bases from FAIRSharing, but you can also type the name of a dataset that is not in FAIRsharing

☒ Desirable: *Before Submitting the DMP*

Guidance on Metadata Standards & Repositories



Index of

... metadata standard

... data deposition repositories

... funder & publisher policies

fairsharing.org



Data Stewardship for Open Science: Chapter 2.5.1.2

With kind permission of



Can the original data be regenerated?

What's up?

In some cases it might be cheaper (and acceptable) to regenerate data rather than storing them. Two examples: It may soon become cheaper to 're-sequence' a genome than to store it for 10 years. Also text mining the same massive corpus of text with the same tagger and the same thesaurus, should in principle give the exact same result when repeated at any time. However, in both examples, a number of assumptions would have to be made before a decision would be made to re-generate the data rather than storing the first version for extended period of time. First of all, the technology should not change; sequencers get more reliable by the day and therefore may give different results and the 'old sequencer' may not be in your possession anymore by the time you want to generate the results. Workflows are not necessarily stable but more importantly even 'stable' substrates (a genome of a living individual or a corpus of text) may not be as stable as you think. Changes to a text corpus may occur unbeknownst to you, but also, the somatic mutation rate in the genome of a living organism are not insignificant and therefore a new sample of cells to take DNA from may give different results and even if the DNA sample was stored in 'preserved state' there is no absolute guarantee that later re-sequencing of it will give exactly the same result. So in all cases, the decision to 'regenerate versus store' is a deep-scientific method discussion in the group and not a 'trivial decision'

Barend Mons

elixir-no.ds-wizard.org

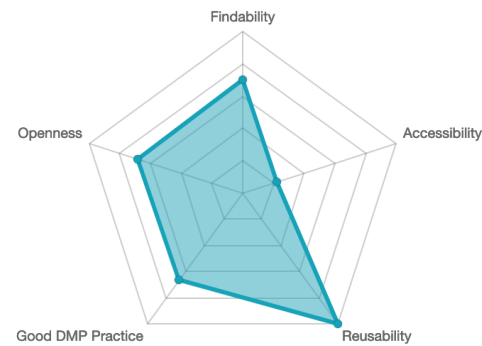
ds-wizard.org
github.com/ds-wizard

Information and insight

Answered: 12/17



Metric	Measure	
Findability	0.70	<div><div></div></div>
Accessibility	0.22	<div><div></div></div>
Reusability	1.00	<div><div></div></div>
Good DMP Practice	0.67	<div><div></div></div>
Openness	0.69	<div><div></div></div>



Metrics Explanation

F - Findability

The Findability metric describes how easily data can be located. The score associated with an answer will be higher if it makes it easier for humans or for computers to locate your data set, e.g. if it ends up in an index or has a unique resolvable identifier.

A - Accessibility

The Accessibility metric describes how well the access to the database is described and how easy it is to implement. The score associated with an answer will be higher if it makes it easier for humans and computers to get to the data. This is determined by e.g. the protocol for accessing the data or for authenticating users, and also by the guaranteed longevity of the repository. Note that this is different from the Openness metric!



elixir.no
digitallifenorway.org



contact@bioinfo.no
korbinian.bosl@uib.no



@elixirnorway
@DigitaltLiv
@BoslKorbinian



Horizon 2020
Research Infrastructure
Program



With funding from
**The Research
Council of Norway**



**CENTRE FOR
DIGITAL LIFE
NORWAY**



UNIVERSITY OF BERGEN



UiO : **University of Oslo**



UiT The Arctic University of Norway



NTNU

Norwegian University of
Science and Technology



Norwegian University
of Life Sciences

The Data Stewardship Wizard is provided by:

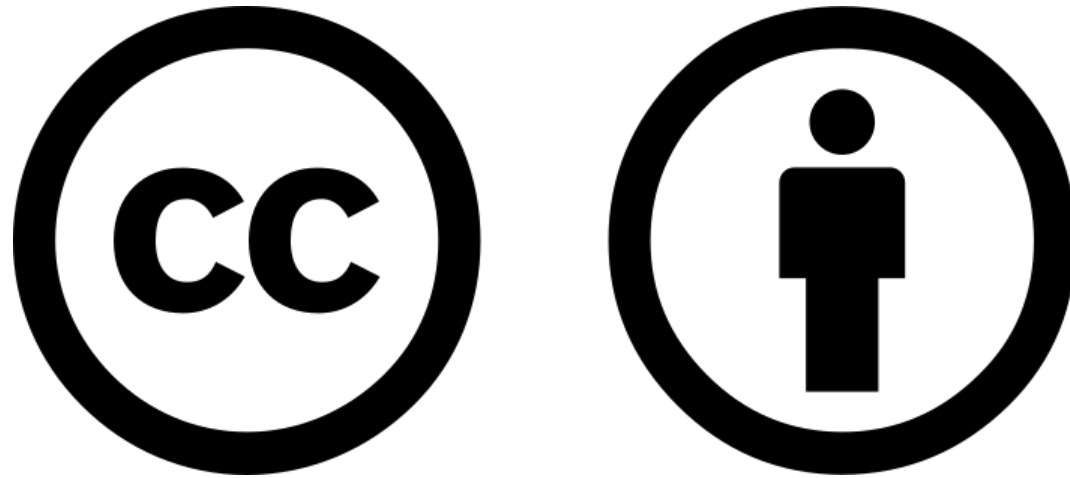


FACULTY
OF INFORMATION
TECHNOLOGY
CTU IN PRAGUE



ÚOCHB IOCB
PRAGUE





Except where otherwise noted, this work is licensed under:
<https://creativecommons.org/licenses/by/4.0/>