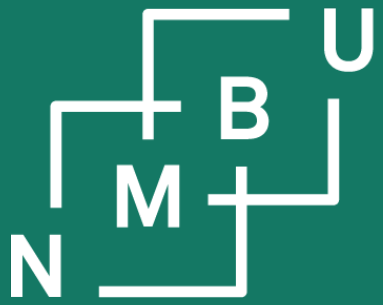
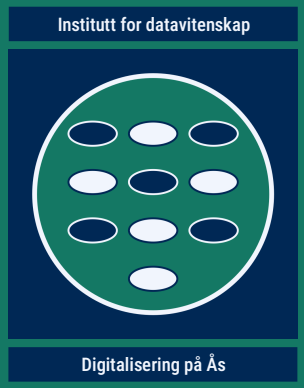


Norges miljø- og biovitenskapelige universitet



DAT390
Data science seminar

3 Methodology

3.2 Research data management (continued)
3.3 Group collaboration proposals
3.4 Reproducibility

Norwegian University of Life Sciences

Highlight talk schedule for today

Monday, 16th October 2023

				Peer feedback from …
15.15 – 15.19	#1	Awo Arab	15.19 – 15.21	#1 Sougata Bhattacharya
		Genomic prediction of complex traits in wheat using multispectral time-series data		
15.23 – 15.27	#2	Simen Holter	15.27 – 15.29	#2 Sujan Devkota
		(undeclared topic)		
15.31 – 15.35	#3	Julie Overrein	15.35 – 15.37	#3 Artush Mktrchyan
		Building yield prediction models with remote sensing and deep learning		
15.39 – 15.43	#4	Asim Rasheed	15.43 – 15.45	#4 Halvor Steffenssen
		EEM spectroscopy and PARAFAC modelling of water quality in nanofiltration		
15.47 – 15.51	#5	Mahrin Tasfe	15.51 – 15.53	#5 Isak Vartdal-Gjerde
		Deep learning identification and classification of paddy disease in precision agriculture		
15.54 – 15.58	#6	Ulrik Egge Husby	15.58 – 16.00	#6 Petter Bøe Hørtvedt
		Exploring the landscape of explainable AI models: An empirical study		

Highlight talk schedule for week 43

Monday, 23th October 2023

				Peer feedback from …	
15.15 – 15.19	#1	Eljar Alihosseinzadeh	15.19 – 15.21	#1	Ming Jeong Cheon
		One-shot learning in business analytics			
15.24 – 15.28	#2	Roy Granheim	15.28 – 15.30	#2	Rusith C. Hathurusinghe
		Impact of transfer learning on human activity recognition using kinect and Mediapipe			
15.33 – 15.37	#3	Sushant Kumar Srivastava	15.37 – 15.39	#3	Kristoffer Lien
		Automated interpretation of visual scenes for autonomous navigation			
15.42 – 15.46	#4	Joen Yacob Teklemariam	15.48 – 15.48	#4	Kim Son Ly
		Automated AI-based event detection systems with audio intensity			
15.51 – 15.55	#5	Baris Ustun	15.55 – 15.57	#5	Bastian Undheim Øian
		Data integration and presentation for drinking water works: Boost readability and reliability			

Highlight talk schedule for week 45

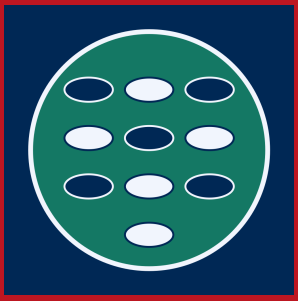
Norwegian University of Life Sciences

Monday, 6th November 2023

				Peer feedback from …
15.15 – 15.19	#1	Amila Haputhanthri	15.19 – 15.21	#1 Ole Benjamin Gauslaa
		Wheat yield prediction using weather, soil, and phenotype data		
15.24 – 15.28	#2	Martin Myklebyst	15.28 – 15.30	#2 Ulrik Egge Husby
		Explainable AI readiness of data and models in journalism		
15.33 – 15.37	#3	Mats Hoem Olsen	15.37 – 15.39	#3 Tonje Martine L. Kirkholt
		Soil temperature model assessment and validation by use cases from agriculture		
15.42 – 15.46	#4	Suhail Rauf	15.46 – 15.48	#4 Bikesh Shrestha
		(undeclared topic)		
15.51 – 15.55	#5	Halvor Steffensen	15.55 – 15.57	#5 Gurubaran Rajeshwaran
		Tsetlin-machine methods over gene data of deep-sea species		

Noregs miljø- og biovitskaplege universitet

Institutt for datavitskap



Digitalisering på Ås

3 Methodology

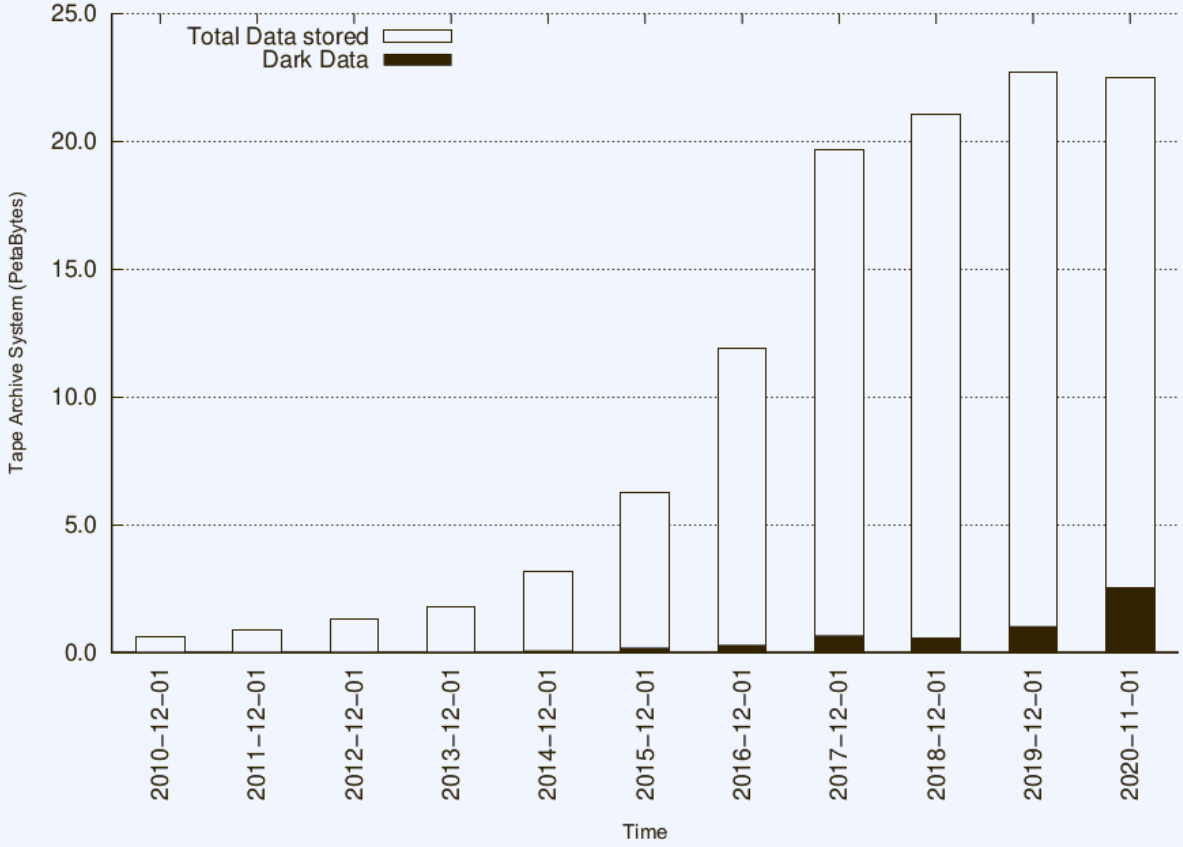
3.2 Data management (ctd.)

DAT390

16. oktober 2023

The challenge: Dark data

Dark data are data with an uncharacterized epistemic status.
In other words: We do not know what we know from and about the data.



dark data

Flood of dark data:
More and more data are accumulated, but are dark – and useless.

Source: Björn Schembera and work by Juan Durán and Björn Schembera.[1]

[1]B. Schembera, J. Durán, *Philos. Technol.* **33**: 93–115, doi:10.1007/s13347-019-00346-x, **2019**.

FAIR principles[1] in detail

persistent identifier

Findability

F1. Globally unique **persistent identifiers (PID)**
F2. **Enriched with metadata**
F3. Data identifier included in metadata
F4. **Registered** in **searchable platform**

Accessibility

A1. **Retrievable from PID** via a standard protocol
A1.1. Open and freely implementable protocol
A1.2. … **authentication/authorization** if necessary
A2. Metadata remain accessible (beyond data)

Interoperability

I1. **Formal language** used for **knowledge representation**
I2. Metadata use **vocabularies** that are themselves FAIR
I3. Semantic web principles, **data can refer to other data**

Reusability

R1. Metadata include a plurality of accurate and relevant attributes
R1.1. Release data and metadata with an accessible **data usage license**
R1.2. Data are annotated with a detailed **provenance description**
R1.3. Relevant **disciplinary and community standards** are fulfilled

[1]M. D. Wilkinson *et al.*, "The FAIR Guiding Principles …," doi:10.1038/sdata.2016.18, **2016**.

FAIR principles[1] in detail

persistent identifier

Findability

F1. Globally unique **persistent identifiers (PID)**
F2. **Enriched with metadata**
F3. Data identifier included in metadata
F4. **Registered** in **searchable platform**

Accessibility

A1. **Retrievable from PID** via a standard protocol
A1.1. Open and freely implementable protocol
A1.2. … **authentication/authorization** if necessary
A2. Metadata remain accessible (beyond data)

Interoperability

I1. **Formal language** used for **knowledge representation**
I2. Metadata use **vocabularies** that are themselves FAIR
I3. Semantic web principles, **data can refer to other data**

metadata

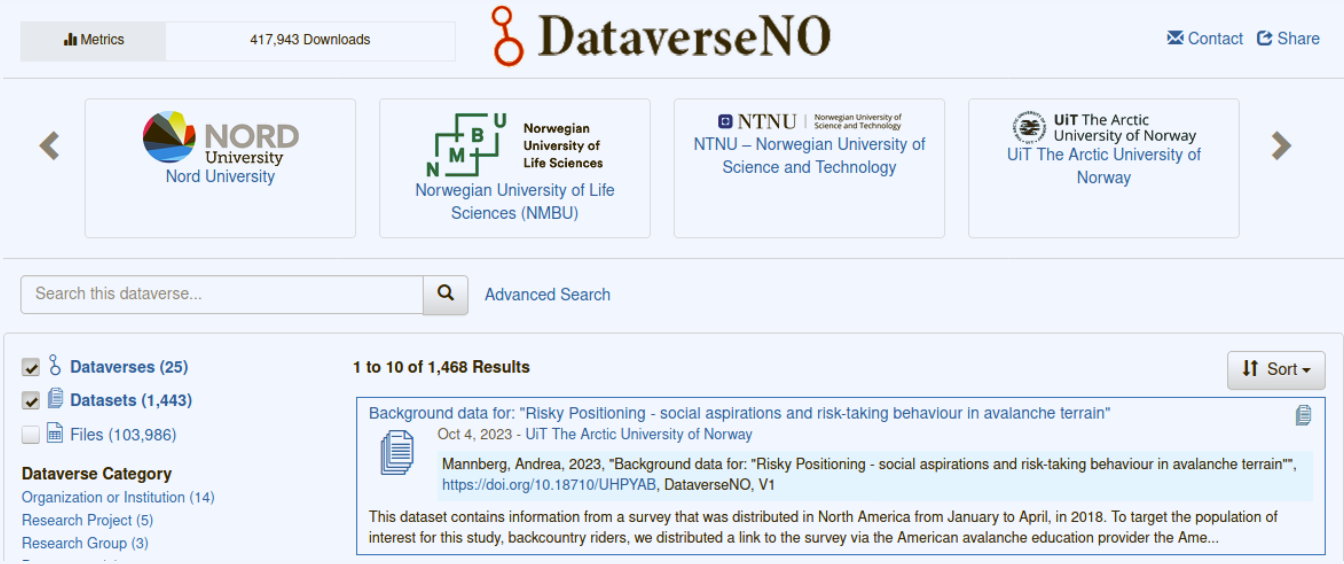
Reusability

R1. Metadata include a plurality of accurate and relevant attributes
R1.1. Release data and metadata with an accessible **data usage license**
R1.2. Data are annotated with a detailed **provenance description**
R1.3. Relevant **disciplinary and community standards** are fulfilled

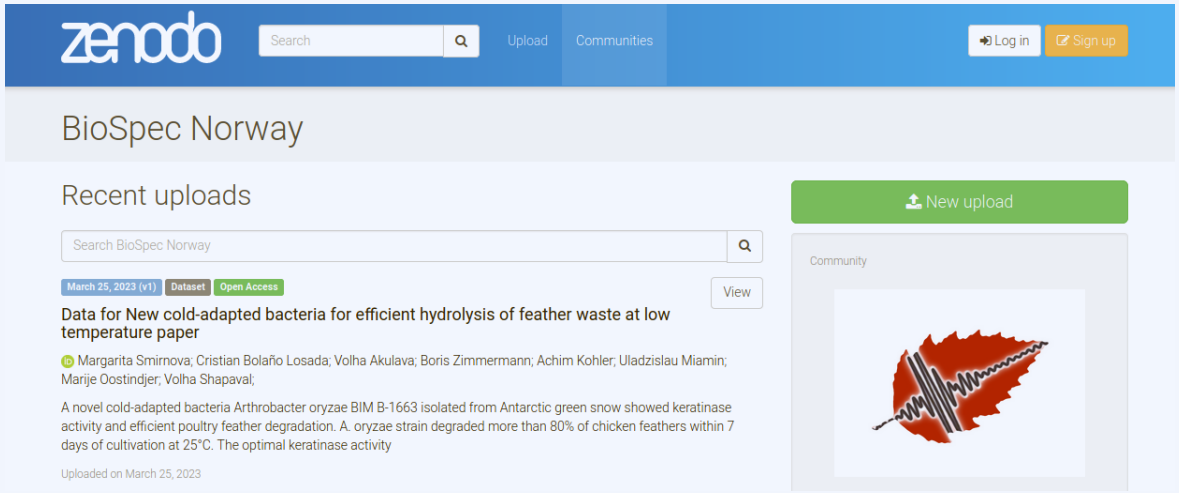
[1]M. D. Wilkinson *et al.*, "The FAIR Guiding Principles …," doi:10.1038/sdata.2016.18, **2016**.

Platforms for data storage and preservation

Dataverse.NO

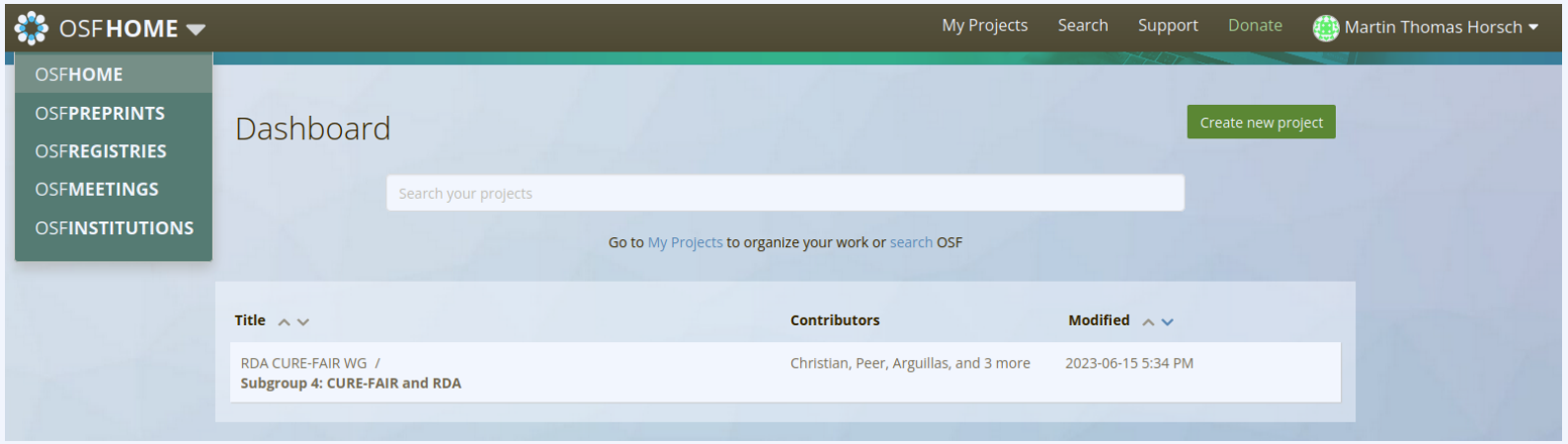


Zenodo

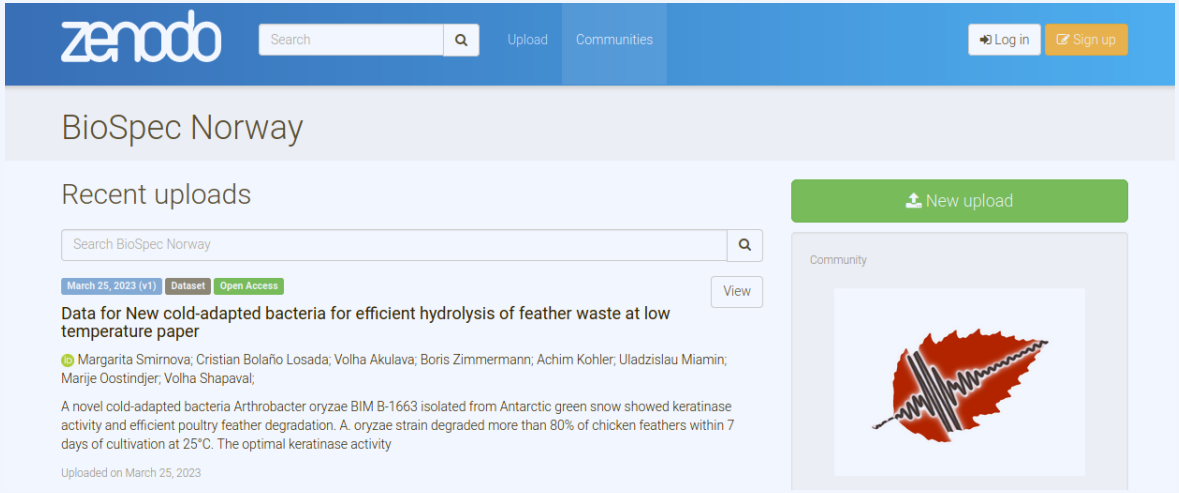


Platforms for data storage and preservation

Open Science Framework (OSF)

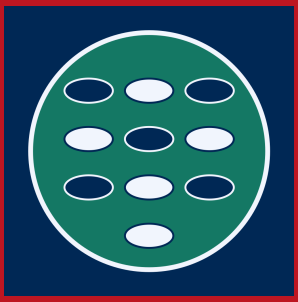


Zenodo



Noregs miljø- og biovitskaplege universitet

Institutt for datavitskap



Digitalisering på Ås

3 Methodology

3.2 Data management (ctd.)

3.3 Group proposals

DAT390

16. oktober 2023

Group collaboration plans and proposals

The group submission by 20th October is a mandatory activity.

It should have three elements (each up to two pages, 11pt font, A4 format):

- Synergy from a **planned group collaboration** beyond this semester.
  (Only involving the group members, how can you support each other?)
- **Proposed student-organized activity** for the coming semester.
  (This would reach out to more than the group, but we will do only one.)
- **Outcome and impact** of the proposed activity.
  (How will you and the community benefit if the activity is selected?)

You can cite references. The literature does not count toward the page limit.

Indicate all who co-authored your document. Keep in mind that for all those who do not contribute, the necessary consequence is that they fail DAT390.

The proposals will be evaluated by all the students together, in a competitive process. However, the result will have no impact on your character grade.

Group collaboration plans and proposals

The group submission by 20th October is a mandatory activity.

It should have three elements (each up to two pages, 11pt font, A4 format):

- **Synergy** from a planned group collaboration beyond this semester.
  (Only involving the group members, how can you support each other?)
- Proposed student-organized **activity** for the coming semester.
  (This would reach out to more than the group, but we will do only one.)
- **Outcome** and impact of the proposed activity.
  (How will you and the community benefit if the activity is selected?)

You can cite references. The literature does not count toward the page limit.

Indicate all who co-authored your document. Keep in mind that for all those who do not contribute, the necessary consequence is that they fail DAT390.

The proposals will be evaluated by all the students together, in a competitive process. However, the result will have no impact on your character grade.

Group proposal competition: Week 43

The eight groups will each describe an idea for joint activities next semester (as a **proposal**, to be realized if selected) and their potential for synergy and collaboration among themselves (as a **plan**, to be realized irrespective).

1. **Visionary Mind**
2. **The Second Land Pirate**
3. **The Overfitters**
4. **Thesis Titans IV**
5. ***Group 5***
6. **BANKTOG**
7. **Thesis Titans VII**
8. **The Forecasters**

In the **first elimination round**, they will be evaluated by individual blind peer review in 1:1 competitions. Before review, the proposals will be presented at seminar time.

Monday, 23rd October 2023 (week 43)

14.20 – 14.40	The Overfitters ./. The Forecasters
14.40 – 15.00	The Second Land Pirate ./. Group 5

Group proposal competition: Week 45

The eight groups will each describe an idea for joint activities next semester (as a **proposal**, to be realized if selected) and their potential for synergy and collaboration among themselves (as a **plan**, to be realized irrespective).

1. **Visionary Mind**
2. **The Second Land Pirate**
3. **The Overfitters**
4. **Thesis Titans IV**
5. ***Group 5***
6. **BANKTOG**
7. **Thesis Titans VII**
8. **The Forecasters**

In the **first elimination round**, they will be evaluated by individual blind peer review in 1:1 competitions. Before review, the proposals will be presented at seminar time.

Monday, 6th November 2023 (week 45)

14.20 – 14.40	Thesis Titans IV ./. Thesis Titans VII
14.40 – 15.00	Visionary Mind ./. BANKTOG

How to present the proposals?

The submitted documents will be shared among the students on Canvas.

The proposals will also be presented in 1-on-1 competitions at seminar time:

- Six minutes for the **first proposal**, presented by three group members
  - **Synergy**
  - **Activity**
  - **Outcome**
- Six minutes for the **second proposal**, presented by three group members
  - **Synergy**
  - **Activity**
  - **Outcome**
- Six minutes **discussion**

(Plus two minutes reserve for technical issues)

Group proposal evaluation: Synergy

In the first round, plans/proposals are evaluated on a scale from **0 to 15 points**, with 0 to 5 from each of the three aspects **synergy**, activity, and outcome. If you are selected to evaluate two plans on **synergy**, ask yourself questions such as:

"A" 5 points
"B" 4 points

"C" 3 points
(should be the norm)

"D" 2 points
"E" 1 point

"F" 0 point
(only if not addressed)

Does the group identify one or multiple **common research question**(s) of interest to them? Is this concrete and specific enough? Is it **realistic** that some new insight can be gained on this by the group's joint effort?

Does the group identify one or multiple **mutual support mechanism**(s) that will help them become stronger together? Is it **realistic**, without requiring too much effort?

Group proposal evaluation: Activity

In the first round, plans/proposals are evaluated on a scale from **0 to 15 points**, with 0 to 5 from each of the three aspects synergy, **activity**, and outcome. If you are selected to evaluate proposals on **activity**, ask yourself questions such as:

"A" 5 points
"B" 4 points

"C" 3 points
(should be the norm)

"D" 2 points
"E" 1 point

"F" 0 point
(only if not addressed)

Does the group concretely describe one or multiple **student-driven activities**? In the proposal, is it **clearly described** what exactly is to be done? Can it **realistically** be done?

Would the proposal, if realized, succeed at **involving students beyond the group** itself? To how much of the *community of final year master students in data science* would it be of interest? Is the involvement of other students **clearly described**, and is it **realistic**?

Group proposal evaluation: Outcome

In the first round, plans/proposals are evaluated on a scale from **0 to 15 points**, with 0 to 5 from each of the three aspects synergy, activity, and **outcome**. If you are selected to evaluate proposals on **outcome**, ask yourself questions such as:

"A" 5 points
"B" 4 points

"C" 3 points
(should be the norm)

"D" 2 points
"E" 1 point

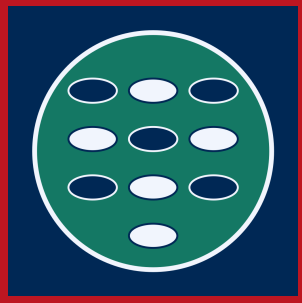
"F" 0 point
(only if not addressed)

Does the proposal concretely describe the potential **benefit to the community** from realizing the proposed activities? Does it take into account the **community's needs**?

Are the outcomes described in a **measurable and verifiable** way, *e.g.*, using key performance indicators and **SMART objectives**? Is it convincingly explained how the proposed activities would **realistically** lead to reaching the promised outcomes?

Noregs miljø- og biovitskaplege universitet

Institutt for datavitskap



Digitalisering på Ås

3 Methodology

3.2 Data management (ctd.)

3.3 Group proposals

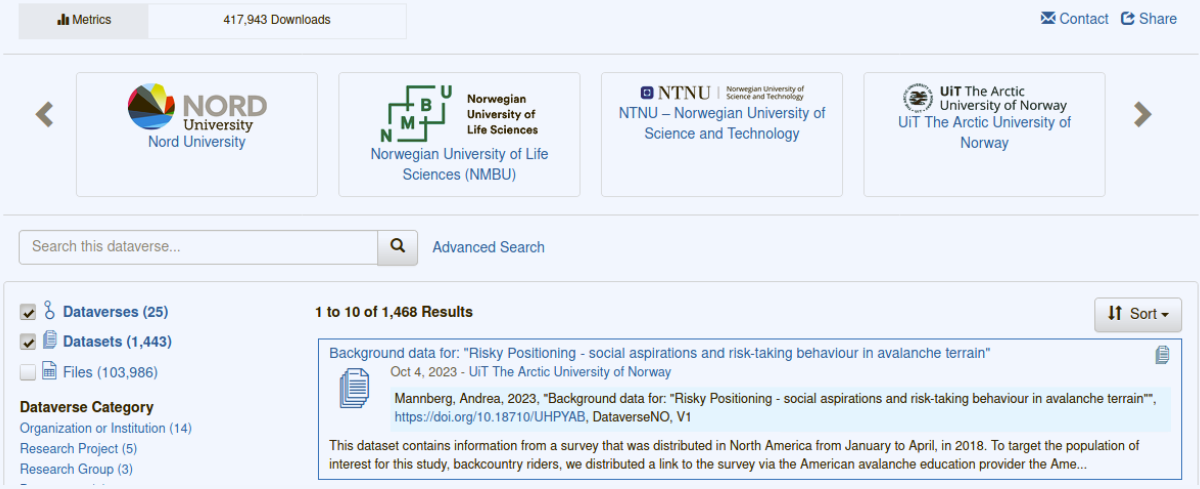
3.4 Reproducibility

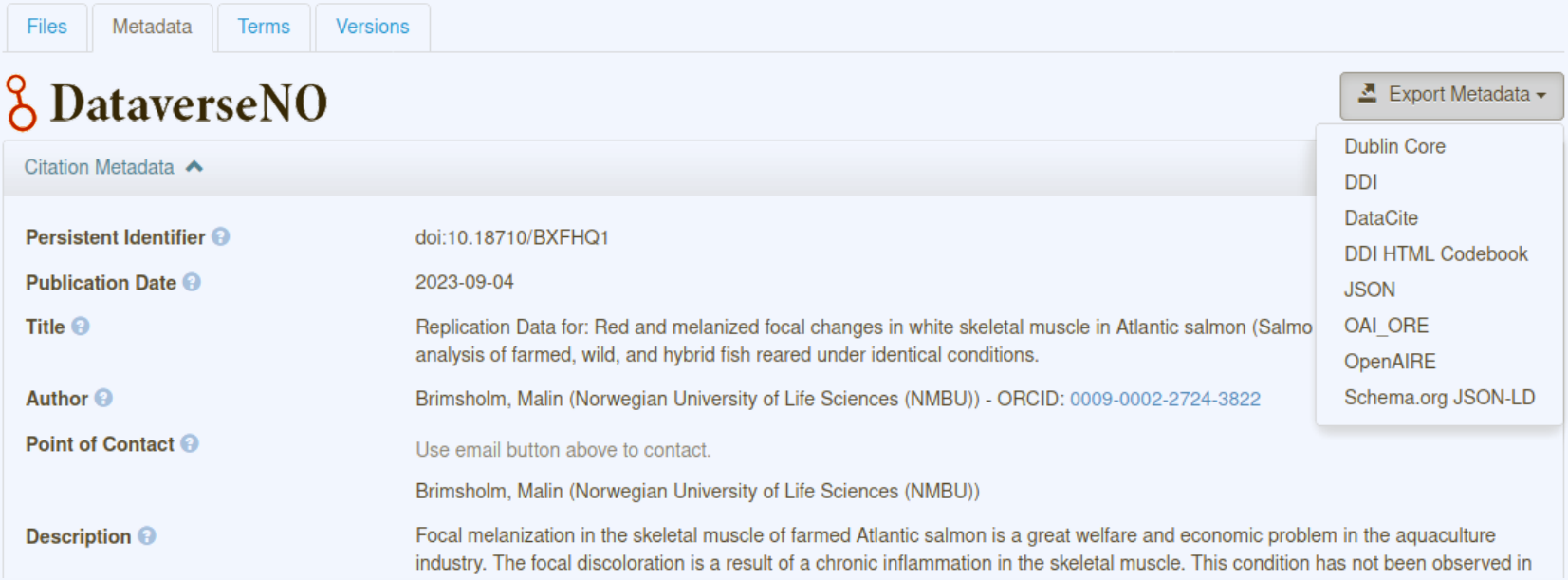
DAT390

16. oktober 2023

Data for replication of scientific findings

Dataverse.NO





Reproducibility, verification, and falsification

reproducibility

There are many definitions of reproducibility and replicability; see the review by Hans Ekkehard Plesser.[1]

1) Researcher *a* did κ and found φ.

2) Researcher *b* did γ, which is very similar to κ, and found ζ, not very similar to φ.

3) Nobody disputes *a*'s integrity. Nobody disputes that *a* did κ and found φ.

Reproducibility claim

«Whenever the research process κ'' is carried out, it **must** lead to the outcome φ''.»

[1]H. E. Plesser, *Frontiers Neuroinform.* **11**: 76, doi:10.3389/fninf.2017.00076, **2018**.

Reproducibility, verification, and falsification

Common formulation and schema for reproducibility claims (RCs):

«Whenever research process κ'' is carried out, it must lead to the outcome φ''.»

1) Researcher *a* did κ and found φ.

   Here, *a* also made a **positive reproducibility claim ψ**.

2) Researcher *b* did γ, **consistent with κ''**, and found ζ, **inconsistent with φ''**.

   Here, *b* made the **negative reproducibility claim $\neg\psi$**.

3) What is relevant there is the **contradiction between ψ and $\neg\psi$**.

provenance metadata κ
provenance paradata κ'

---

provenance orthodata $\kappa'' = \kappa - \kappa'$

«repeat κ, but no need to retain κ'»

knowledge claim metadata φ
knowledge claim paradata φ'

---

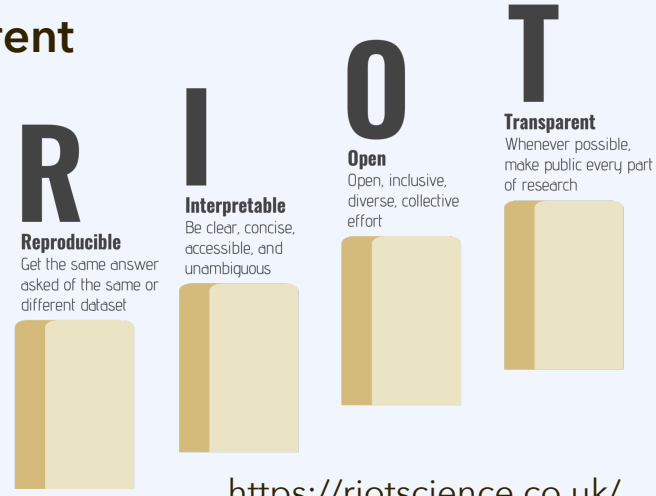
knowledge claim orthodata $\varphi'' = \varphi - \varphi'$

«obtain φ again, except for φ' maybe»

Good practices beyond FAIR

RIOT:[1] Reproducible, interpretable, open, transparent

- Origin: UK Reproducibility Network (UKRN)
- UKRN encouraged foundation of the other reproducibility networks, such as NORRN, the Norwegian Reproducibility Network
- Local "RIOT science clubs" were founded



https://riotscience.co.uk/

CARE:[2] Collective benefit, authority to control, responsibility, ethics

- Origin: Global Indigenous Data Alliance
- Uptake supported by the Research Data Alliance
- Orientation: Sovereignty and epistemic justice

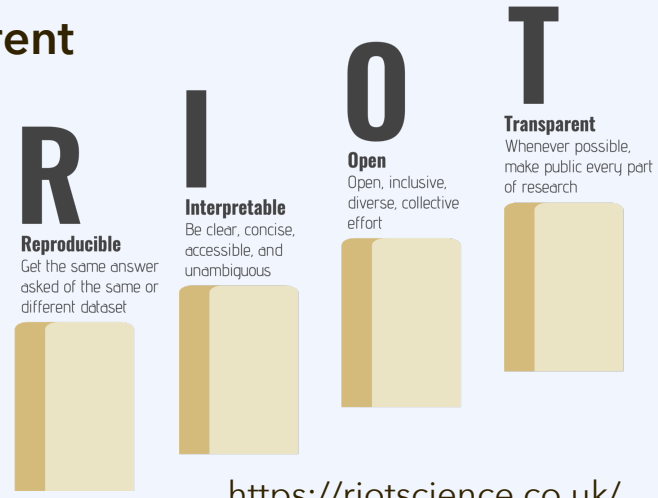
https://www.gida-global.org/care/

[1]E. Ganley *et al.*, *BMC Res. Notes* **15**: 51, doi:10.1186/s13104-022-05932-5, **2022**.
[2]S. Russo Carroll *et al.*, *Sci. Data* **8**: 108, doi:10.1038/s41597-021-00892-0, **2021**.

Good practices beyond FAIR

RIOT:[1] Reproducible, interpretable, open, transparent

- Origin: UK Reproducibility Network (UKRN)
- UKRN encouraged foundation of the other reproducibility networks, such as NORRN, the Norwegian Reproducibility Network
- Local "RIOT science clubs" were founded



https://riotscience.co.uk/

CARE:[2] Collective benefit, authority to control, responsibility, ethics

- Origin: Global Indigenous Data Alliance
- Uptake supported by the Research Data Alliance
- Orientation: Sovereignty and epistemic justice

https://www.gida-global.org/care/

[1]E. Ganley *et al.*, *BMC Res. Notes* **15**: 51, doi:10.1186/s13104-022-05932-5, **2022**.
[2]S. Russo Carroll *et al.*, *Sci. Data* **8**: 108, doi:10.1038/s41597-021-00892-0, **2021**.

Norwegian Reproducibility Network

Norwegian University of Life Sciences

NORRN

Our Mission

The Norwegian Reproducibility Network (NORRN) is a peer-led network that aims **to promote and enable rigorous, robust and transparent research practices in Norway**. We attempt to achieve this goal by establishing appropriate training activities, designing, and evaluating research improvement efforts, disseminating best practices, and working with stakeholders to ensure coordination of efforts across the sector. NORRN's activities span multiple levels, including researchers, librarians, institutions, and other stakeholders (e.g., funders and public authorities).

Researchers

We **support researchers** in educating themselves about open science practices, and founding local open science communities.

Initiatives

We **connect Reproducibility Initiatives** to a national network, and foster connections between them.

Institutions

We **advise institutions** on how to embed open science practices in their work.

Stakeholders

We **represent the open science community** toward other stakeholders in the wider scientific landscape.

Norwegian Reproducibility Network

Norwegian University of Life Sciences

NORRN

Our Mission

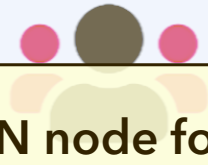
The Norwegian Reproducibility Network (NORRN) is a peer-led network that aims **to promote and enable rigorous, robust and transparent research practices in Norway**. We attempt to achieve this goal by establishing appropriate training activities, designing and evaluating research improvement efforts, disseminating best practices, and working with stakeholders to ensure coordination of efforts across the sector. NORRN's activities span multiple levels, including researchers, librarians, institutions, and other stakeholders (e.g., funders and public authorities).

Norwegian Reproducibility Network: https://www.norrn.no/

Schedule a meeting for all those interested in creating a NORRN node for Ås:

https://terminplaner6.dfn.de/p/491179ac2aaea450dd6f015191d08269-429068



Researchers

We **support researchers** in educating themselves about open science practices, and founding local open science communities.

Initiatives

We **connect Reproducibility Initiatives** to a national network, and foster connections between them.

Institutions

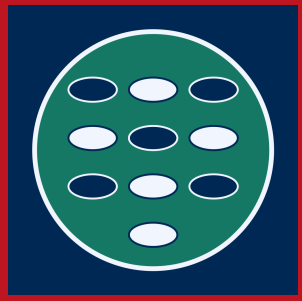
We **advise institutions** on how to embed open science practices in their work.

Stakeholders

We **represent the open science community** toward other stakeholders in the wider scientific landscape.

Noregs miljø- og biovitskaplege universitet

Institutt for datavitskap



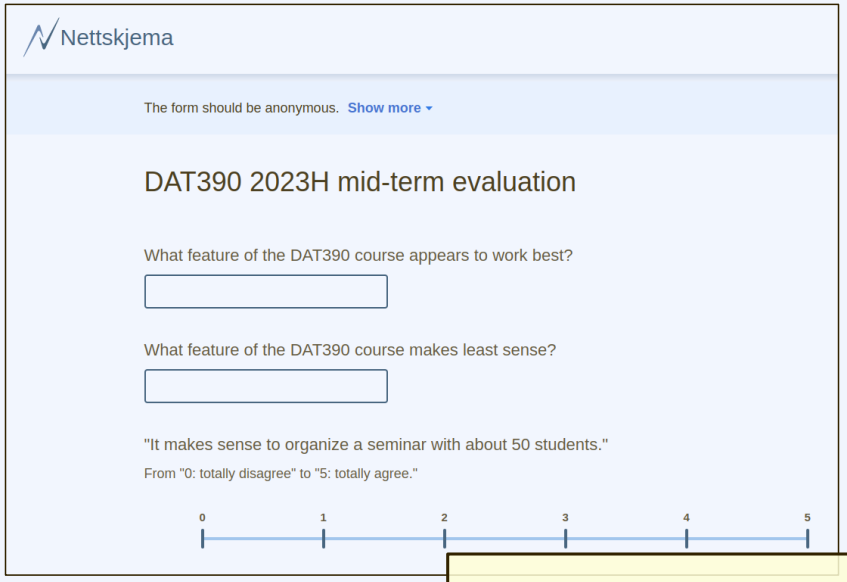
Digitalisering på Ås

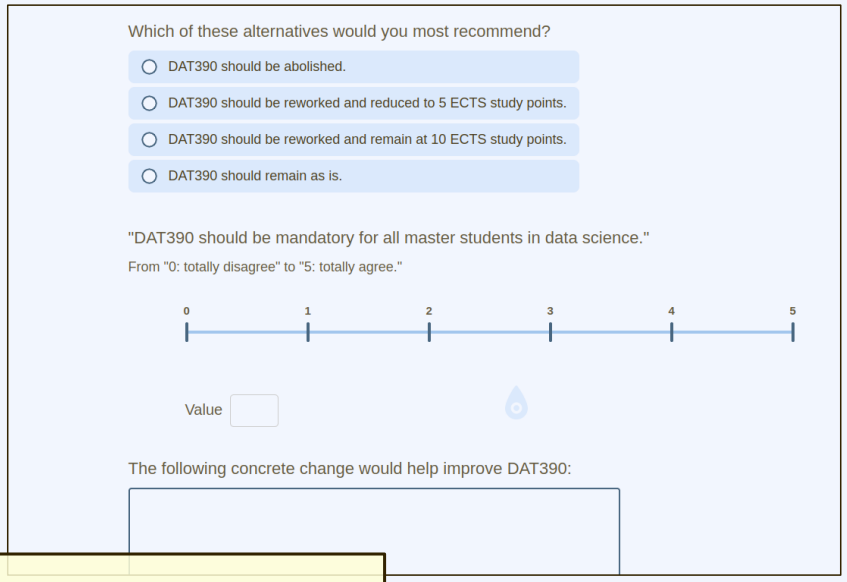
Mid-term evaluation (underveisevaluering)

DAT390

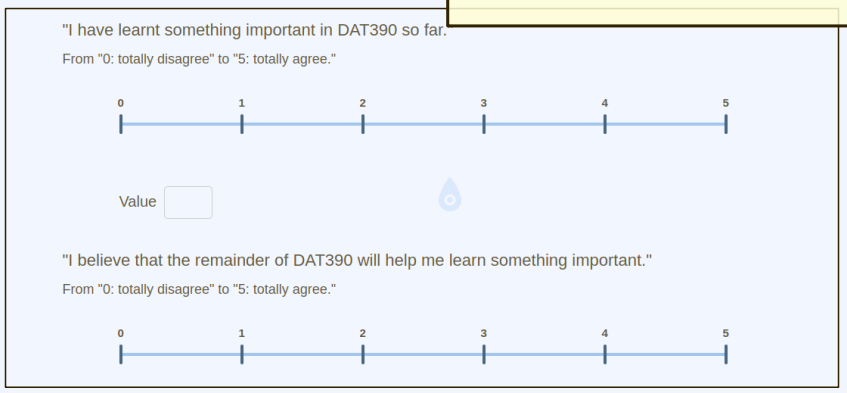
16. oktober 2023

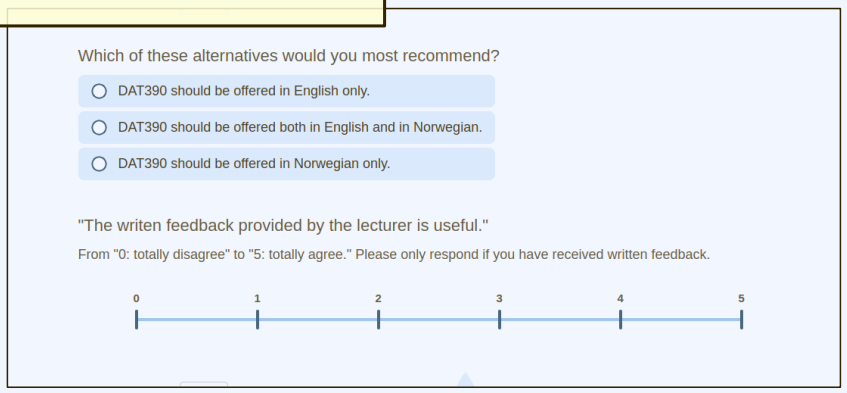
Mid-term evaluation

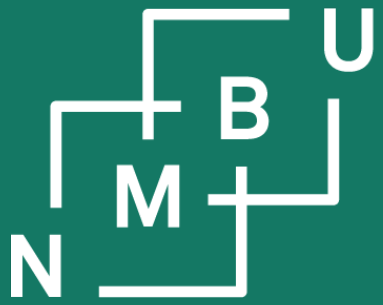




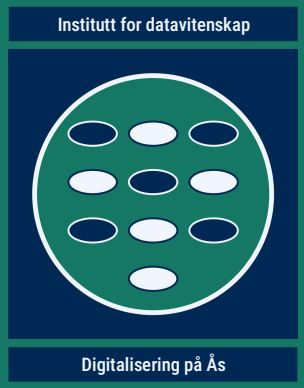
https://nettskjema.no/a/dat390







Norges miljø- og biovitenskapelige universitet



DAT390
Data science seminar

3 Methodology

3.2 Research data management (continued)
3.3 Group collaboration proposals
3.4 Reproducibility