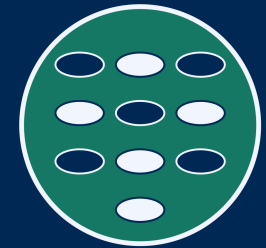


Norges miljø- og
biovitenskapelige
universitet

Institutt for datavitenskap

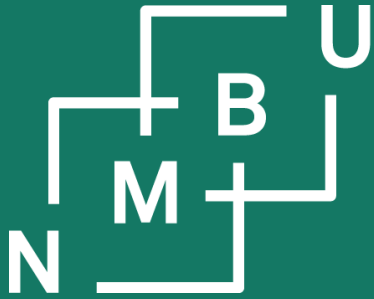


Digitalisering på Ås

DAT390

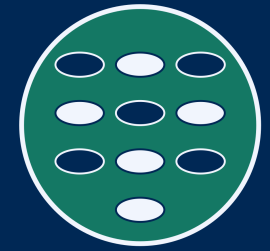
Data science seminar

- 3 Research methodology
 - 3.1 Going beyond the state of the art
 - 3.2 Research data management



Noregs miljø- og
biovitenskaplege
universitet

Institutt for datavitenskap

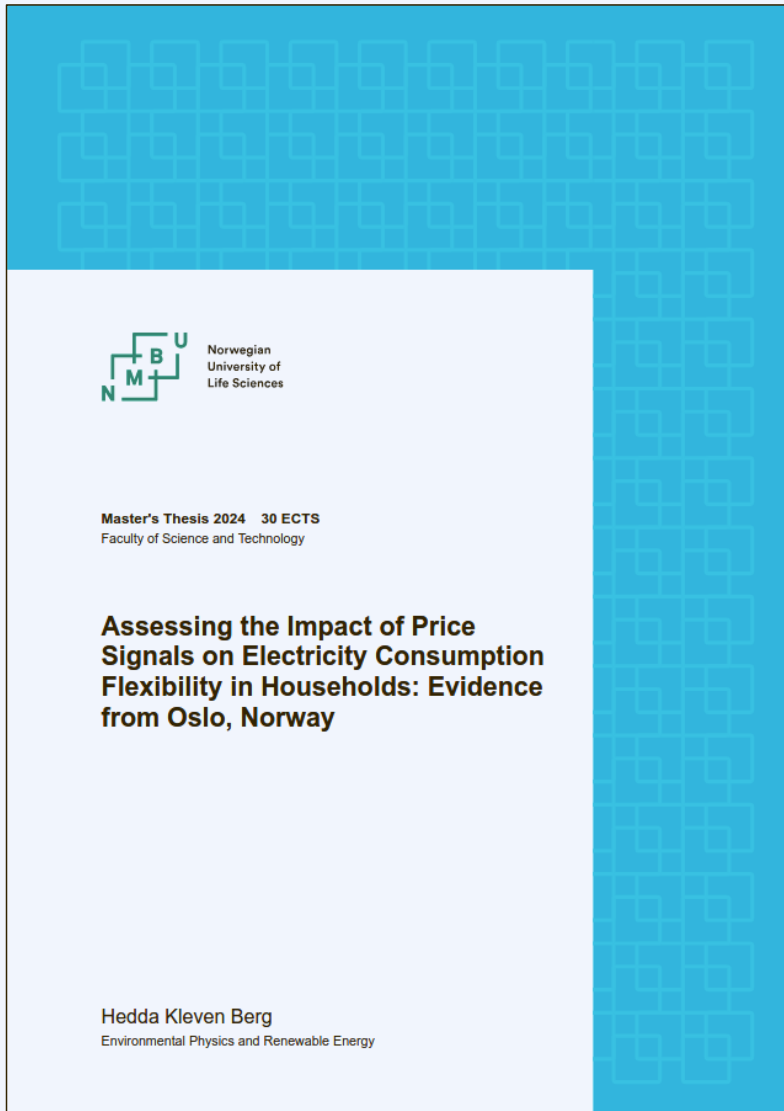


Digitalisering på Ås

3 Methodology

3.1 Beyond state of the art

How to be “beyond-the-art by design”



Be specific, work on a narrow topic.

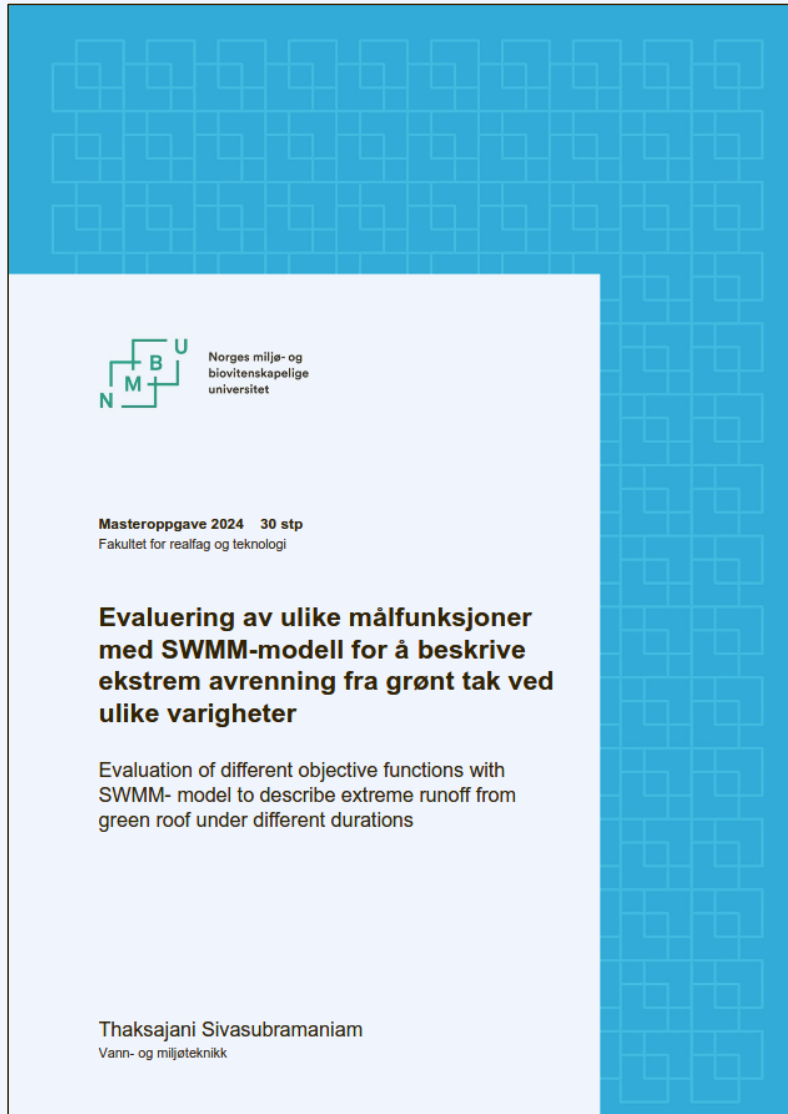
It is easy to confirm that nobody else ever worked on this exact topic.

Also good: It is best if already from the title itself it becomes clear in what way the work is novel.

However: Being specific is not enough, we must **learn something novel** from it.

This can be something generalizable, but here it could also mean learning what is special about people in Oslo.

How to be “beyond-the-art by design”

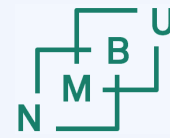


Use recently developed methods

Millions of people have done scientific research, how do you know yours is something nobody else did?

The SWMM model was developed in 2015, so it was possible for the author to evaluate the complete body of literature on it.

Your advisor should point you to recent methods or new approaches where it is very easy to do novel work that advances our body of knowledge.



About the 2nd “nearly finished” report

From the “nearly finished report” on (deadline 8th November), the work is expected to cover all aspects, not just reviewing the state of the art.

It must in this sense be complete.

What distinguishes a “nearly finished” DAT390 report from the master thesis?

- The **master thesis** needs to report on the finalized research and results.
- The **master thesis** is a comparably long document.



About the 2nd “nearly finished” report

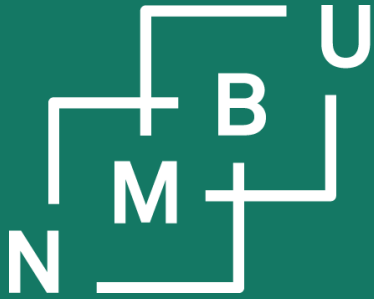
UHR Universitets-
og høgskolerådet

See the UHR’s recommended **standardized assessment form** for master theses.

What distinguishes a “nearly finished” DAT390 report from the master thesis?

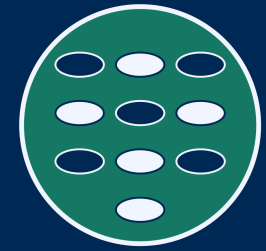
- The **master thesis** needs to report on the finalized research and results. The **DAT390 report** is not expected to contain any final results. However, **preliminary results** or a **feasibility study** are needed.
- The **master thesis** is a comparably long document. The **DAT390 report** is like a conference paper, up to 12 pages (+ literature), 11pt, A4 paper.
- There is only one difference between the final **DAT390 report** and the **“nearly finished” report**: The DAT390 report determines your character grade in DAT390. The “nearly finished” report is not graded.

What are the **criteria for evaluating a master thesis**? See UHR document.



Noregs miljø- og
biovitenskaplege
universitet

Institutt for datavitenskap



Digitalisering på Ås

3 Methodology

3.1 Beyond state of the art

3.2 Research data management

NMBU's guidelines for research data management

Research data management guidelines document

(Approved by NMBU's rector in 2018)

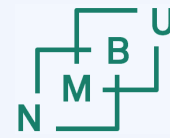
- Principle **"open as standard"**
- Comply with international standards for research data management
- Comply with the **FAIR principles**.
- **Research data/metadata** should be available, searchable, and reusable.
- Furthermore, "interoperable" means that both data and metadata must be **manageable for machines** and that a **consistent vocabulary** is used.

NMBU's guidelines for research data management

Research data management guidelines document

(Approved by NMBU's rector in 2018)

- 1) Research data must be **stored safely**.
- 2) Research data should not (only) be stored on a local storage.
- 3) Follow agreed **good practices**, regulations, guidelines, and the law.
- 4) Persistent long-term **data preservation** for a minimum of ten years.
- 5) The research data must be **annotated with metadata**.
- 6) A **data management plan** (DMP), describing how data will be managed.
- 7) Open and available (**open-access**) research data.
- 8) Protect **personal data**.



Compliance with GDPR

Concerning personal data, we need to comply with GDPR, and therefore:

- You need to make sure that there is a line of responsibility connecting your work to the **Data Protection Officer (DPO)**, also *personvernombod*, of the organization. [Hanne P. Gulbrandsen](#) is NMBU's DPO.

GDPR: General data protection regulation. (Link to PDF document.)

This is implemented in Norway through personopplysningsloven (see document).

SIKT - Kunnskapssektorens tjenesteleverandør must be notified about use of personal data in your research project.

¹<https://gdpr.eu/data-protection-impact-assessment-template/?cn-reloaded=1>

Data protection impact assessment

Concerning personal data, we need to comply with GDPR, and therefore:

- You need to make sure that there is a line of responsibility connecting your work to the **Data Protection Officer (DPO)**, also *personvernombod*, of the organization. [Hanne P. Gulbrandsen](#) is NMBU's DPO.
- Your work may require a **Data Protection Impact Assessment (DPIA)** ...
 - «if you're using new technologies»,
 - «data [...] used to make automated decisions about people»,
 - «if you're tracking people's location or behavior», «monitoring a publicly accessible place» or «processing children's data», etc.
- You need **freely given, specific, informed**, and **unambiguous consent**.
 - Be aware of simultaneous requirements from NMBU's RDM policy.

SIKT - Kunnskapssektorens tjenesteleverandør must be notified about use of personal data in your research project.

¹<https://gdpr.eu/data-protection-impact-assessment-template/?cn-reloaded=1>

Data protection impact assessment

Criteria from NTNU's knowledge base:¹

- 1) "Evaluation or assigning a point score, included profiling and prediction [...]"
- 2) Automated decisions with legal or similarly significant consequences [...]"
- 3) Systematic monitoring/observation
- 4) Special categories of personal data [... e.g.] of a highly personal nature [...]"
- 5) Personal data are processed at a large scale [...]"
«It is unclear what constitutes large scale.»
- 6) Matching or integration of datasets [...]"
- 7) Personal data on vulnerable subjects [...]"
- 8) Innovative use or application of a new technical or organizational approach [...]"
«in relation to the previously reached level of technical knowledge»
- 9) Whenever the processing *«deprives those registered of the possibility to exercise a right or make use of a service or an agreement.»* This includes processing that aims at allowing, changing, or denying the registered person access to a service or enter an agreement."

¹<https://i.ntnu.no/wiki/-/wiki/Norsk/Vurdere+personvernkonsekvenser>

Data protection impact assessment

Checklist from GDPR.EU



Sample DPIA template

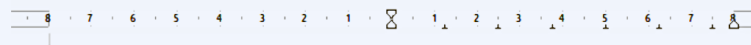
This template is an example of how you can record your DPIA process and outcome. It follows the process set out in our DPIA guidance, and should be read alongside that guidance and the [Criteria for an acceptable DPIA](#) set out in European guidelines on DPIAs.

You should start to fill out the template at the start of any major project involving the use of personal data, or if you are making a significant change to an existing process. The final outcomes should be integrated back into your project plan.

Step 1: Identify the need for a DPIA

Explain broadly what project aims to achieve and what type of processing it involves. You may find it helpful to refer or link to other documents, such as a project proposal. Summarise why you identified the need for a DPIA.

DPIA template from NTNU



2. Behandlings art

Behandlingens iboende karakteristik og hvordan behandlingsaktivitetene skal foregå. Beskrivelser av hva dere planlegger å gjøre med personopplysningene.

| | |
|---|---|
| Hvordan skal personopplysningene samles inn? | Samles de f.eks. inn fra den registrerte eller fra andre? |
| Hvordan skal personopplysningene lagres? | |
| Hvordan skal personopplysningene brukes? | |
| Hvem skal ha tilgang til personopplysningene? | |
| Hvem skal det samles inn personopplysninger om? | F.eks. ansatte i egen virksomhet, elever/studenter, pasienter, kunder/klienter osv. |
| Hvordan kan den registrerte utøve sine rettigheter? | |
| Vil det være systematisk behandling av personopplysninger? | Ja/nei |
| Brukes det ny teknologi eller ny bruk av eksisterende teknologi hvor personvermkonsekvenser ikke har blitt vurdert? | Ja/nei |

3. Behandlings omfang

| | |
|--|---|
| Kategorier av personopplysninger som behandles | Behandles særskilte kategorier av personopplysninger eller andre typer personopplysninger som kan oppleves som private? |
| Antall registrerte involvert i behandlingen | I tall eller evt. prosentandel av et utvalg |
| Datavolum | Antall variabler, detaljeringsgrad |
| Behandlingsfrekvens | Hentes opplysningene inn én gang, flere ganger, regelmessig, kontinuerlig osv. |
| Lagringstid for personopplysningene | Kort, tidsavgrenset, permanent |
| Geografisk omfang | Lokalt, regionalt, nasjonalt, internasjonalt |

4. Behandlings formål

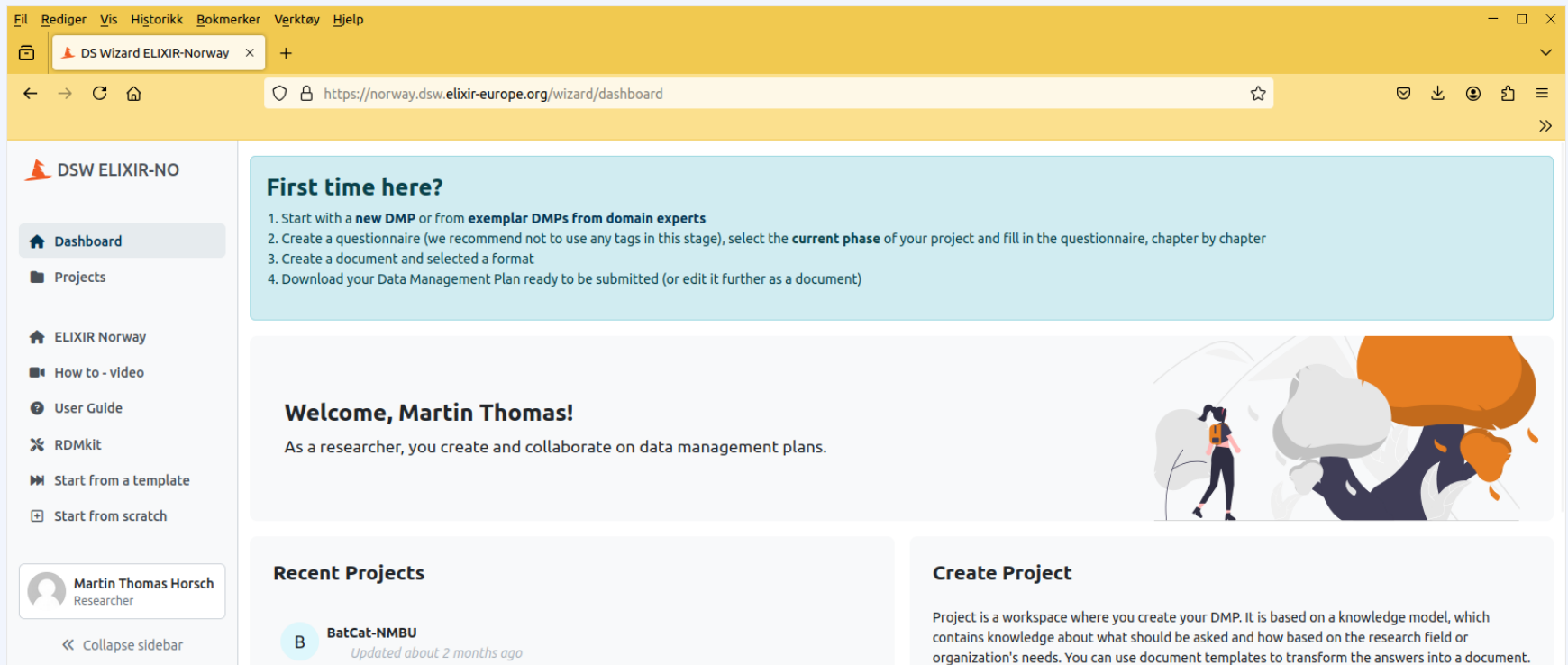
| | |
|------------------------------|--|
| Behandlings formål | |
| Vil det være kontrollformål? | |

Tools for drafting a data management plan (DMP)

Institutional templates for DMPs

Example: DMP template for Horizon Europe projects

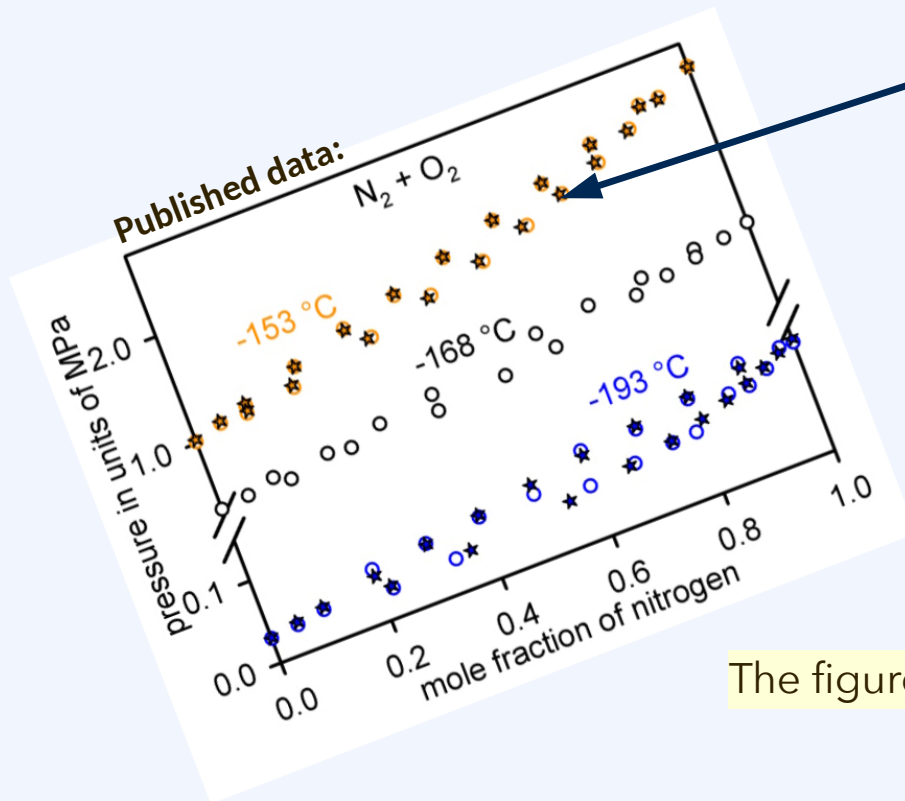
Data Stewardship Wizard (DSW)



The screenshot shows the DSW ELIXIR-Norway dashboard in a web browser. The browser's address bar displays the URL <https://norway.dsw.elixir-europe.org/wizard/dashboard>. The dashboard features a yellow header with navigation menus: 'Fil', 'Rediger', 'Vis', 'Historikk', 'Bokmerker', 'Verktøy', and 'Hjelp'. The main content area is divided into several sections:

- DSW ELIXIR-NO**: A sidebar menu on the left with options: Dashboard, Projects, ELIXIR Norway, How to - video, User Guide, RDMkit, Start from a template, and Start from scratch.
- First time here?**: A light blue box containing a 4-step guide:
 1. Start with a **new DMP** or from **exemplar DMPs from domain experts**
 2. Create a questionnaire (we recommend not to use any tags in this stage), select the **current phase** of your project and fill in the questionnaire, chapter by chapter
 3. Create a document and selected a format
 4. Download your Data Management Plan ready to be submitted (or edit it further as a document)
- Welcome, Martin Thomas!**: A personalized greeting with the text: "As a researcher, you create and collaborate on data management plans." An illustration of a person carrying a large orange bag is visible to the right.
- Recent Projects**: A section showing a project named "BatCat-NMBU" with a blue circular icon containing the letter 'B' and the text "Updated about 2 months ago".
- Create Project**: A section with the text: "Project is a workspace where you create your DMP. It is based on a knowledge model, which contains knowledge about what should be asked and how based on the research field or organization's needs. You can use document templates to transform the answers into a document."

Why do we need good practices?



What values did x and p have?

How was the data point obtained?

What is the margin of error, how was the error defined, and what software (or experimental setup) was used?

The figure is not enough - **metadata** are needed!

Good practice in managing research data:

Make all data **findable**, **accessible**, **interoperable**, and **reusable** (FAIR).

FAIR principles¹⁻³

Summary from the guidance document by Direktoratet for e-helse:³

Findability: «That there is information in the form of **metadata**, making it simple for both humans and machines to find **datasets**.»

Accessibility: «That the **data and metadata** are arranged in a way that makes it simple for humans and machines **to obtain datasets**.»

Interoperability: «That **metadata and variables** are described in accordance with international **standards**, terminologies, and classifications.»

Reusability: «That there is information in the form of **metadata** that describe the **conditions for reuse** of the dataset, how the dataset has been edited, and the quality of the data.»

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

²<https://www.uio.no/for-ansatte/arbeidsstotte/forskningsstotte/forskningsdata/tema/fair.html> (norsk versjon, UiO).

³Direktoratet for e-helse, Veileder for bruk av FAIR-prinsippene for helsedatakilder, **2020**.

FAIR principles in detail¹⁻³

persistent
identifier

Findability (gjenfinnbarheit)

F1. Globally unique **persistent identifiers (PID)**

F2. **Enriched with metadata**

F3. **Data identifier included in metadata**

F4. **Registered in searchable platform**

Accessibility (tilgjengelegheit)

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 (samhandlingsevne)

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 (gjenbruksevne)

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

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

²<https://www.uio.no/for-ansatte/arbeidsstotte/forskningsstotte/forskningsdata/tema/fair.html> (norsk versjon, UiO).

³Direktoratet for e-helse, Veileder for bruk av FAIR-prinsippene for helsedatakilder, **2020**.

What are metadata?

Dataverse.NO

Metrics 417,943 Downloads Contact Share

Search this dataverse... Advanced Search

Datasets (25) Datasets (1,443) Files (103,986)

Dataverse Category
Organization or Institution (14)
Research Project (5)
Research Group (3)

1 to 10 of 1,468 Results Sort

Background data for: "Risky Positioning - social aspirations and risk-taking behaviour in avalanche terrain"
Oct 4, 2023 - UiT The Arctic University of Norway
 Mannberg, Andrea, 2023, "Background data for: "Risky Positioning - social aspirations and risk-taking behaviour in avalanche terrain"", <https://doi.org/10.18710/UHPYAB>, DataverseNO, V1

This dataset contains information from a survey that was distributed in North America from January to April, in 2018. To target the population of interest for this study, backcountry riders, we distributed a link to the survey via the American avalanche education provider the Ame...

Files Metadata Terms Versions

DataverseNO

Citation Metadata

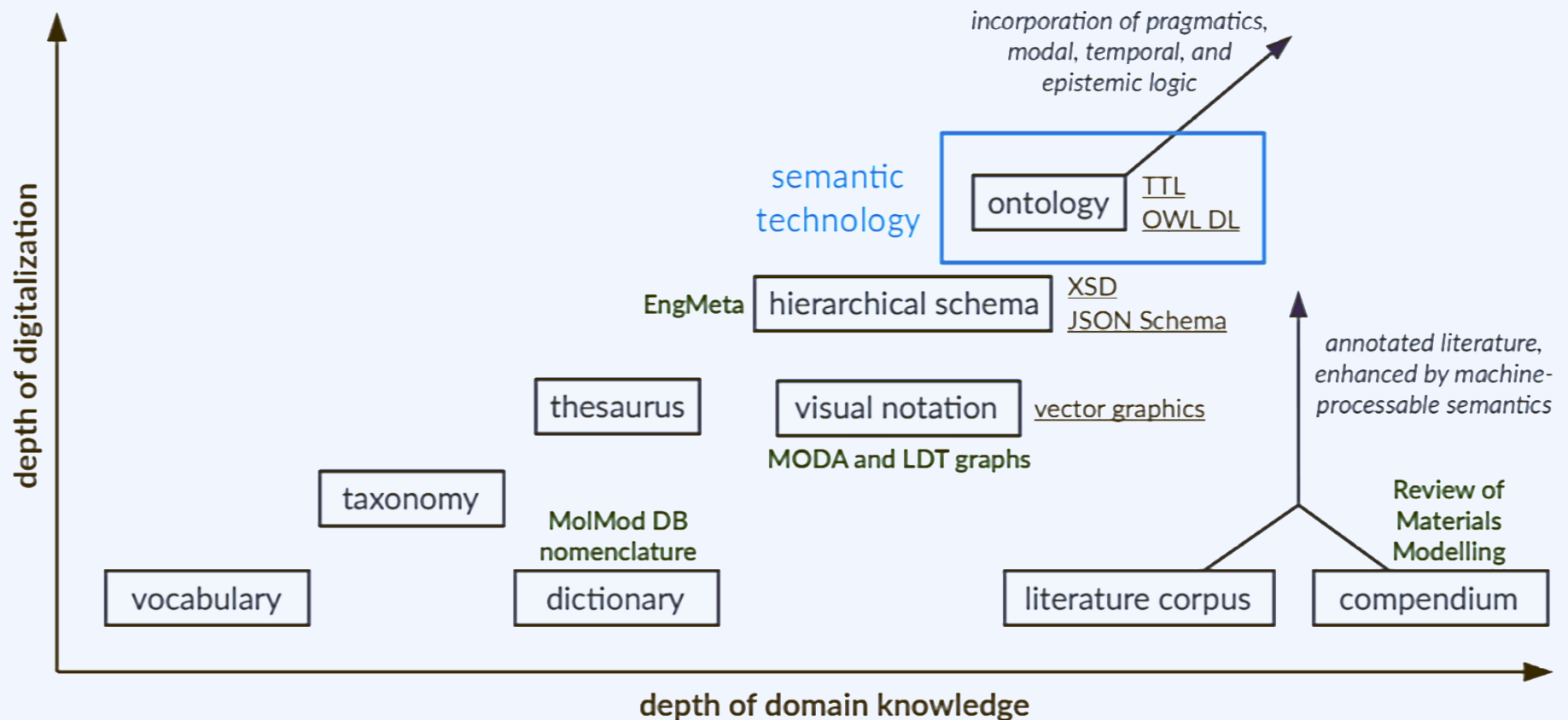
| | |
|--------------------------------|--|
| Persistent Identifier ? | doi:10.18710/BXFHQ1 |
| Publication Date ? | 2023-09-04 |
| Title ? | Replication Data for: Red and melanized focal changes in white skeletal muscle in Atlantic salmon (<i>Salmo</i> analysis of farmed, wild, and hybrid fish reared under identical conditions. |
| Author ? | Brimsholm, Malin (Norwegian University of Life Sciences (NMBU)) - ORCID: 0009-0002-2724-3822 |
| Point of Contact ? | Use email button above to contact. Brimsholm, Malin (Norwegian University of Life Sciences (NMBU)) |
| Description ? | Focal melanization in the skeletal muscle of farmed Atlantic salmon is a great welfare and economic problem in the aquaculture industry. The focal discoloration is a result of a chronic inflammation in the skeletal muscle. This condition has not been observed in |

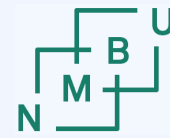
Export Metadata

- Dublin Core
- DDI
- DataCite
- DDI HTML Codebook
- JSON
- OAI_ORE
- OpenAIRE
- Schema.org JSON-LD

Agreed metadata by standardization

Types of **semantic artefacts**, also referred to as **metadata standards**:





Persistent identifiers (PIDs)

The issuer of a PID must guarantee that it remains in place for the foreseeable future, and that metadata for the referenced object also remain in place.

Good PIDs are **GUPRIs**: Globally unique, persistent, resolvable identifiers.

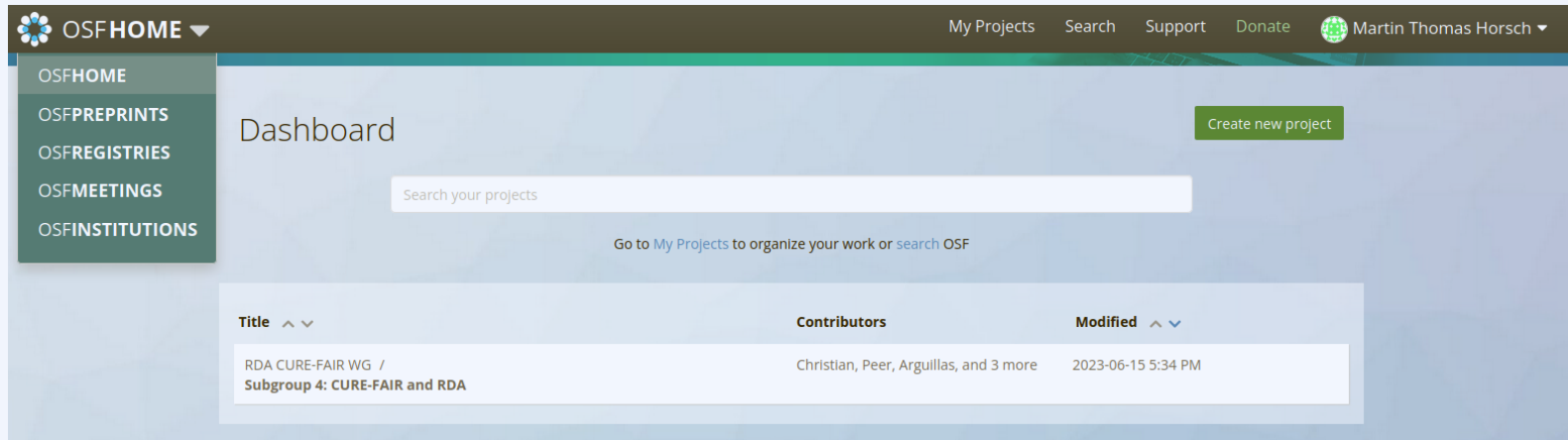
Examples:

- For researchers: **ORCID**.
- For books: **ISBN**. For journals: **ISSN**; however, these are not resolvable.
- For organizations: **ISNI** and **ROR**.
- Redirects to any digital artefacts, e.g., software development projects: Persistent URLs (**PURLs**); however, these are not globally unique.
- For documents, datasets, or any other digital objects: **DOI**.

Everybody engaged in research should have an ORCID.

Platforms for creating digital object identifiers

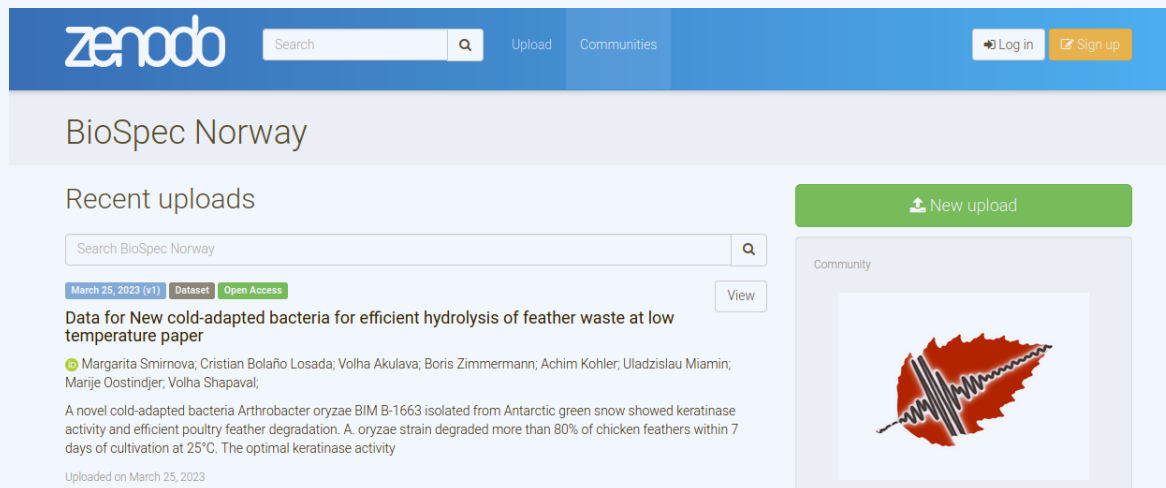
Open Science Framework (OSF) – see also the OSF hands-on guide



The screenshot shows the OSFHOME dashboard. At the top, there is a navigation bar with 'OSFHOME' and a dropdown menu, 'My Projects', 'Search', 'Support', 'Donate', and a user profile for 'Martin Thomas Horsch'. A left sidebar contains links for 'OSFHOME', 'OSFPREPRINTS', 'OSFREGISTRIES', 'OSFMEETINGS', and 'OSFINSTITUTIONS'. The main content area is titled 'Dashboard' and features a 'Create new project' button, a search bar for projects, and a link to 'Go to My Projects to organize your work or search OSF'. Below this is a table listing projects.

| Title | Contributors | Modified |
|---|---------------------------------------|--------------------|
| RDA CURE-FAIR WG / Subgroup 4: CURE-FAIR and RDA | Christian, Peer, Argullas, and 3 more | 2023-06-15 5:34 PM |

Zenodo




The screenshot shows the Zenodo website for the BioSpec Norway community. The header includes the Zenodo logo, a search bar, and links for 'Upload', 'Communities', 'Log in', and 'Sign up'. The main content area is titled 'BioSpec Norway' and features a 'Recent uploads' section. A search bar is present, and a 'New upload' button is visible. The featured upload is titled 'Data for New cold-adapted bacteria for efficient hydrolysis of feather waste at low temperature paper' by Margarita Smirnova, Cristian Bolaño Losada, Volha Akulava, Boris Zimmermann, Achim Kohler, Uladzislau Miamin, Marije Oostindjer, and Volha Shapaval. The abstract describes a novel cold-adapted bacterium, *Arthrobacter oryzae* BIM B-1663, isolated from Antarctic green snow, showing keratinase activity and efficient poultry feather degradation. The upload date is March 25, 2023.

Check published data for FAIRness


Self-evaluation form issued by Direktoratet for e-helse¹

| | | |
|---|--|----------------------------|
| | A | |
| 1 | | |
| 2 | FAIR Evalueringsskjema | ☰ Direktoratet for e-helse |
| 3 | | |
| 4 | Velkommen | |
| 5 | <p>Bakgrunn FAIR-prinsippene er et sett med veiledende arkitekturprinsipper som skal tilrettelegge for deling og gjenbruk av data gjennom at dataene er søkbare (Findable), tilgjengelige (Accessible), understøtter interoperabilitet (Interoperable) og er gjenbrukbare (Reusable).</p> | |

F-UJI:² Automated FAIR data assessment tool



Automated FAIR Data Assessment Tool

F-UJI is a web service to programmatically assess FAIRness of research data objects at the dataset level based on the FAIRsFAIR Data Object Assessment Metrics 

[Click here to assess a dataset](#)

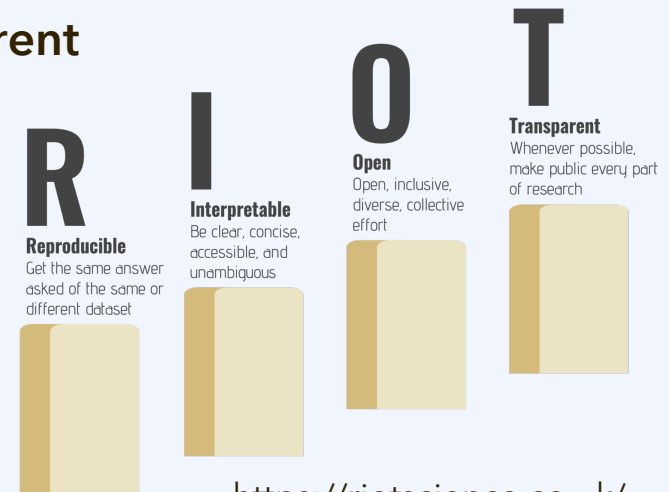
¹FAIR evaluation form (Norwegian), Direktoratet for e-helse, 2020.

²<https://www.f-uji.net/>

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

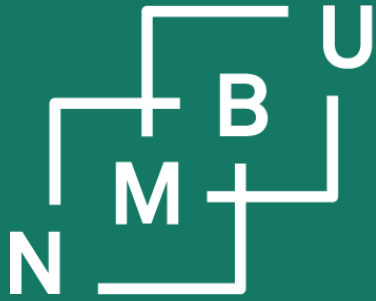
- Origin: Global Indigenous Data Alliance
- Uptake supported by the Research Data Alliance
- Orientation: Sovereignty and epistemic justice



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

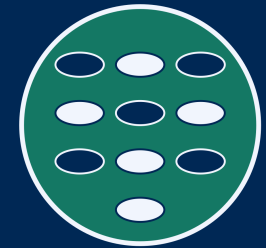
¹E. Ganley *et al.*, *BMC Res. Notes* **15**: 51, doi:10.1186/s13104-022-05932-5, **2022**.

²S. Russo Carroll *et al.*, *Sci. Data* **8**: 108, doi:10.1038/s41597-021-00892-0, **2021**.



Norges miljø- og
biovitenskapelige
universitet

Institutt for datavitenskap



Digitalisering på Ås

DAT390

Data science seminar

- 3 Research methodology
 - 3.1 Going beyond the state of the art
 - 3.2 Research data management