DAT390
Data science seminar

4   Research impact and ethics

4.1   Plagiarism vs. copyright
4.2   Re-use of own material (so-called "self-plagiarism")

Highlight talk schedule for today

Monday, 13th November 2023

Peer feedback from …

15.15 – 15.19   #1   Alin Dak Al-Bab   15.19 – 15.21   #1   Amila Haputhanthri

Predicting customer churn levels (B2C) in the telecommunication industry

15.24 – 15.28   #2   Nivetha Suntharamoorthy   15.28 – 15.30   #2   Suhail Rauf

Effects of imputation techniques on time series classification methods

15.33 – 15.37   #3   Olutomi Okubadejo   15.37 – 15.39   #3   Martin Myklebust

AI and machine learning algorithms to detect and prevent network security threats

15.42 – 15.46   #4   August Noer Steinset   15.46 – 15.48   #4   Mats Hoem Olsen

Deep learning for direct DNA domain detection

15.51 – 15.55   #5   Eljar Alihosseinzadeh   15.55 – 15.57   #5   Kristoffer Lien

One-shot learning in business analytics

Highlight talk schedule for next week

Monday, 20th November 2023

Peer feedback from …

15.15 – 15.19 #1 Rusith C. Hathurusinghe   15.19 – 15.21 #1 Maryna Berg

Prediction and estimation of indoor radon based on sensory and meteorological data

15.24 – 15.28 #2 Bastian Undheim Øian   15.28 – 15.30 #2 Trishaban Jegatheeswaran

Declustering for spatial models

15.33 – 15.37 #3 Kristoffer Lien   15.37 – 15.39 #3 Vegard Molaug

Enhancing CRM/ERP data modelling efficiency by MC simulation: Minimum viable data

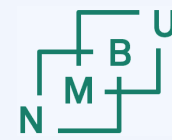15.42 – 15.46 #4 Min Jeong Cheon   15.46 – 15.48 #4 Hongpeng Zhang

Comparison of survival analysis algorithms on a breast cancer dataset

15.51 – 15.55 #5 Martin Myklebust   15.55 – 15.57 #5 Nivetha Suntharamoorthy

Explainable AI readiness of data and models in journalism

Visionary Mind #1 vs. *BANKTOG* #6

	Visionary Mind (#1)		BANKTOG (#6)	
Synergy	4		4	
Activity	4	4	3	3
Outcome	4	4	3	

Visionary Mind #1 vs. *BANKTOG* #6

	Visionary Mind (#1)	BANKTOG (#6)
Synergy	4	4
Activity	4　4　(avg: 4)	3　3　(avg: 3)
Outcome	4　4　(avg: 4)	3
Total	12 out of 15	10 out of 15

Feedback to the *Visionary Mind* (#1)

Synergy (4 points)

«The group identifies one main support mechanism which is […] group collaboration or open lines of communication/group chat. […] could have been more creative with ways to support each other and […] more options.»

Activity (4 points)

«**Data Science Symposium**, […] an event aimed at bringing together faculty and students […]. includes student presentations, guest speakers, and panel discussions. It encourages student engagement and provides opportunities for presenting and discussing research. […] more detail on the execution and management of such a large-scale event could enhance its feasibility.»

Activity (4 points)

«The symposium's objectives, including student presentations, guest speakers, and panel discussions, suggest a well-rounded […] approach. […] include […] details like the date, venue, and registration process […].»

Feedback to the *Visionary Mind* (#1)

Outcome (4 points)

«The symposium is expected to enhance **presentation skills**, provide exposure to **industry trends**, and facilitate **problem-solving** and **collaboration** among students. The outcomes are clearly aligned with the academic and professional development of students.
However, the proposal could be strengthened by including more specific metrics for measuring these outcomes.»

Outcome (4 points)

«Each part, like "Enhanced presentation skills," "Exposure to Industry Trends," and "Problem Solving and Collaboration," is clearly defined. […] they might make them even better by brainstorming together for more detailed examples. Talking about communication skills, real-world applications, and working together grabs the reader's attention, showing how everyone benefits. […] could make it even beler by adding more specific examples.»

Original Thesis Titans #4 vs. *Thesis Titans* #7

	Original Titans (#4)	**Titans *no*. Seven (#7)**
Synergy	4	5
Activity	4 5	3
Outcome	*has received reviews*	*has <u>not</u> received a review*

Original Thesis Titans #4 vs. *Thesis Titans* #7

	Original Titans (#4)	**Titans *no.* Seven (#7)**
Synergy	<u>4</u>	<u>5</u>
Activity	4 5 (avg: <u>4.5</u>)	<u>3</u>
Outcome	*has received reviews*	*has <u>not</u> received a review*
<u>Total</u>	8.5 out of 10	8 out of 10

Feedback to the *Original Thesis Titans* (#4)

Synergy (4 points)

«sharing ideas, checking each other's work, and helping each other with writing and […] to support each other emotionally. To make it even better, the plan could include ways to measure how well these ideas are working»

Activity (4 points)

«share ideas, hold brainstorming sessions and […] a peer review system. […] the description didn't include […] how the activities should be done […] "**Thesis Pomodoro Plan**" […] They have done a good job and made it easy to follow. Its rather easy to implement as well, being a **realistic** option.
On whether the activity would succeed involving students beyond the group itself, it would not be that ideal for that.»

Activity (5 points)

«It is realistically executable and seems engaging for students, potentially attracting a wide audience among final year master students in data science.»

Feedback to the *Original Thesis Titans* (#4)

Outcome

- «The proposed activity from the Original Thesis Titans is Thesis Pomodoro Plan.»
- «The group highlights the most essential aspects when considering the outcome of their proposal.»
- «They could be more detailed when looking at […] when it would start and how many weeks it would be hosted.»

- «Fun and easy activity to be a part of. There are very few "bad" outcomes from this kind of activity.»
- «The group did not mention […] who will set a date or find a room for everyone to sit in.»
- «This activity can be implemented both physically and virtually, although the group did not discuss this opportunity further.»
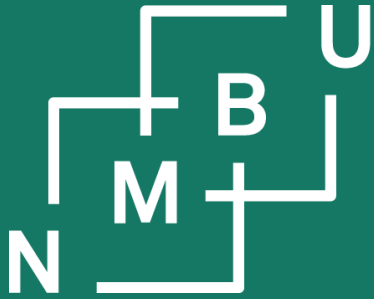
Revised proposals

Four groups proceed to the next round:

- Visionary Mind (#1)

- The Original Thesis Titans (#4)

- Genetic Prophesiers (#5)

- The Forecasters (#8)
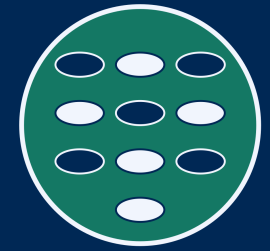
These groups are asked to:

- Look into reviewers' feedback and submit a revised proposal (17.11.).
- Make a very brief statement on what they improved, next week (20.11.).

Noregs miljø- og biovitskaplege universitet

4 **Impact and ethics**

4.1 **Plagiarism and copyright**

Plagiarism and copyright are different issues

It is possible to commit a copyright infringement without plagiarism.

- Example: *A* publishes an opinion piece on a streaming platform, under a license that does not permit commercial reuse. *B* has a monetized channel on that platform and publishes a reaction video that consists of *A*'s content and *B*'s trivial reactions. *B* does not obscure *A*'s authorship.

- Copyright infringement is a matter of penal law (strafferett) following åndverksloven §§ 79 and 80. It is a matter of civil law (privatrett), specially liability (erstatningsansvar), following åndverksloven § 81.

- You have the right to fair use for citation (sitatrett) of others' material:
  - As part of the public debate, including academic discussions, you can use direct quotes and reproduce part of others' work verbatim.
  - In the example, *B*'s use of *A*'s material would be legitimate if *B* was engaging in a genuine, non-trivial way with the original content.
  - The right to fair use for citation is grounded in åndverksloven § 29.

Plagiarism and copyright are different issues

It is possible to commit plagiarism without a copyright infringement.

– Example: *A* hires *B* as a contractor to write technical documentation for *A*'s code. The contract clarifies that A owns the documentation.
A publishes a paper in a scientific journal. Part of the documentation is included as an Appendix. It is not mentioned that *B* wrote that text.

– Depending on circumstances, plagiarism may be completely legal.
  • The above is probably illegal by being in violation of the terms and conditions of the journal's publisher, with whom *A* has a contract. But if there is no such contract, it is legal, but it is still plagiarism. Irrespective, it is no copyright violation, since *A* owns the copyright.
  • A can sell the code and documentation to C without mentioning B.

– Primarily, plagiarism is a matter of research ethics, not the law.
Norway, however, also has a law specifically about research ethics.
Forskningsetikkloven § 8 mentions[1] plagiarism as form of "uredelighet."

[1]Unfortunately, the law only mentions plagiarism, but does not define it.

Plagiarism and copyright are different issues

It is possible to commit plagiarism without a copyright infringement.

- Example: *A* hires *B* as a contractor to write technical documentation for *A*'s code. The contract clarifies that A owns the documentation. *A* publishes a paper in a scientific journal. Part of the documentation is included as an Appendix. It is not mentioned that *B* wrote that text.

- Depending on circumstances, plagiarism may be completely legal.
  - The above is probably illegal by being in violation of the terms and conditions of the journal's publisher, with whom *A* has a contract. But if there is no such contract, it is legal, but it is still plagiarism. Irrespective, it is no copyright violation, since *A* owns the copyright.
  - A can sell the code and documentation to C without mentioning B.

- Primarily, plagiarism is a matter of research ethics, not the law.

- Plagiarism in an **examination** setting is a case of misconduct (**fusk**). It is regulated as a matter of administrative law (forvaltningsrett).

NMBU's regulations on academic misconduct (fusk)

The relevant document is called "Retningslinjer - behandling av mistanke om fusk eller annen uredelig opptreden ved NMBU" (last changed in June 2015).

Defines fusk as "enhver handling som sikter mot å gi studenten et uberettiget studieresultat eller et uberettiget fortrinn ved evaluering av studieprestasjon".

The legal basis for this is given in **universitets- og høyskoleloven (uhl.)**:
- uhl. § 4-7 (1) b states that exam results etc. can be annulled in case of fusk or an attempt at fusk.
- uhl. § 4-8 (3) states that whoever acts as described in § 4-7 (1) *or contributes to it* can be expelled (utestengt) for up to one year.
- No definition of "fusk" is given in the law. The word is just used as is.

It is interesting that the law assumes that "har forsøkt å fuske" is different from "har fusket," as it mentions both separately. NMBU however gives a definition according to which the attempt to "fusk" already is a "fusk."

NMBU's regulations on academic misconduct (fusk)

The relevant document is called "Retningslinjer - behandling av mistanke om fusk eller annen uredelig opptreden ved NMBU" (last changed in June 2015).

Defines fusk as "enhver handling som sikter mot å gi studenten et uberettiget studieresultat eller et uberettiget fortrinn ved evaluering av studieprestasjon". This definition under point 3.1 is only followed by "examples" under point 3.2:

«[…] f. Fusk kan bestå i at en innlevert oppgave er **utarbeidet av en annen person** *enn den som står oppført som eksaminand.*

g. **Plagiat er fusk.** *Eksempler på plagiering: Gjengivelse eller sitater fra bøker, artikler, internettsider,* **egne eller andres oppgaver***, bruk av bilder, grafer o.l. uten kildehenvisning, sitattegn eller annen tilkjennegivelse i teksten/bildet/tegningen som viser hvor materialet er hentet fra.*

h. Fusk kan bestå i at en innlevert besvarelse er **brukt av eksaminanden selv til en tidligere eksamen***, med mindre slik bruk er avtalt med faglærer. […]»*

It seems that point 3.2 is only intended as an illustration of point 3.1.

Plagiarism versus "plagiarism detection software"

Plagiarism detectors can help detect cheating. They don't define plagiarism.

- You can have no similarity at the text level, but still commit plagiarism.
    - "I have to rewrite or paraphrase the text" (maybe using ChatGPT?).
    - No! Others' academic works don't become yours by paraphrasing.
- High similarity scores in no way mean that plagiarism really occurred.
    - Maybe you are using a form or template that many are using.
- Even reusing text verbatim can be legitimate, distinct from plagiarism, and not requiring direct quotation marks or even a citation.
    - Example: *A* uses *B*'s text to learn Norwegian, taking over many Norwegian idioms from *B*'s text. None of the scientific content is taken from *B*, and *A*'s work is on a different topic. *B*'s work was only used as a language learning resource. It does not need to be cited.

NMBU's plagiarism detector is called "Ouriginal."

(https://secure.urkund.com/account/en-US/auth/login) and (https://www.ouriginal.com/guides-tutorials/)

Plagiarism versus "plagiarism detection software"

Plagiarism detectors can help detect cheating. They don't define plagiarism.

- You can have no similarity at the text level, but still commit plagiarism.
  - "I have to rewrite or paraphrase the text" (maybe using ChatGPT?).
  - No! Others' academic works don't become yours by paraphrasing.
- High similarity scores in no way mean that plagiarism really occurred.
  - Maybe you are using a form or template that many are using.
- Even reusing text verbatim can be legitimate, distinct from plagiarism, and not requiring direct quotation marks or even a citation.
  - Example: *A* uses *B*'s text to learn Norwegian, taking over many Norwegian idioms from *B*'s text. None of the scientific content is taken from *B*, and *A*'s work is on a different topic. *B*'s work was only used as a language learning resource. It does not need to be cited.

NMBU's plagiarism detector is called "Ouriginal." Do not use it to "find out if you committed plagiarism." This does not make sense. You already know.

What even is plagiarism?

Plagiarism detectors can help detect cheating. They don't define plagiarism.

OK, but what is plagiarism then, really?

NMBU's retningslinjer om fusk 3.1g do not define plagiarism, but give a series of examples:

g. **Plagiat er fusk.** *Eksempler på plagiering: Gjengivelse eller sitater fra bøker, artikler, internettsider,* **egne eller andres oppgaver**, *bruk av bilder, grafer o.l. uten kildehenvisning, sitattegn eller annen tilkjennegivelse i teksten/bildet/tegningen som viser hvor materialet er hentet fra.*

h. *Fusk* kan *bestå i at en innlevert besvarelse er* **brukt av eksaminanden selv til en tidligere eksamen**, *med mindre slik bruk er avtalt med faglærer. […]»*

This is not a definition, but makes it clear that "self-plagiarism" is considered plagiarism, and that this includes reuse of work from previous theses/exams. However, it is then not clear why points *g* and *h* are listed separately.

What even is plagiarism?

UHR: Universitets- og høgskolerådet[1] (2012)

UHR's first definition: «Hva er plagiering? <u>Meget forenklet</u> kan vi si at **å plagiere er å presentere andres arbeid, herunder tanker og ideer, som sitt eget**.»

However, UHR does not continue to rely on this "very simplified" definition.

The same report goes on to write: «Studenten kan også plagiere seg selv: Når studenten helt eller delvis bruker egne tidligere innleverte arbeider i nye innleveringer, uten å oppgi det, kalles det gjerne selvplagiering.»

This is followed by a list of six examples for plagiarism. The sixth one is "self-plagiarism," which is then subsumed under plagiarism. However, no new, better definition of plagiarism is given instead of the first "very simplified" one.

[1]UHR, *Plagiering i universitets- og høgskolesektoren: Felles problem, felles ansvar*, **2012**.

What even is plagiarism?

Research ethics guidelines for natural science and technology[1] (2015)

These Norwegian guidelines,[1] developed by the NENT committee, assert:
«**Å plagiere innebærer å framstille andres ideer eller forskning som sitt eget.**»

This is inconsistent with the idea of "self-plagiarism" as a kind of "plagiarism."

[1]NENT, *Forskningsetiske retningslinjer for naturvitenskap og teknologi*, FEK, **2015**.

What even is plagiarism?

Project RINO: Research Integrity in Norway[1] (2018)

The RINO project was carried out by *De nasjonale forskningsetiske komiteene (FEK)* in collaboration with UiB and HVL. It was centered on research integrity as the opposite of "forgery, fabrication, and plagiarism" (FFP).

Plagiarism is defined in this project by: "**Å fremstille andres arbeid (ideer, materiale, tekst) som sitt eget ved å utelate henvisning til opphavskilden**".[1]

Their report[1] finds that 90.1% of Norwegian research institutions' employees regard plagiarism, as defined above, as a "very problematic" behaviour. The fraction who find it "very problematic" *or "quite problematic"* is 98.5%.

(Remark: This is probably only because "copying your own work" is excluded.)

[1] J. Hjellbrekke *et al.*, *Etikk og integritet i forskning: Resultater fra en landsomfattende undersøkelse*, **2018**.
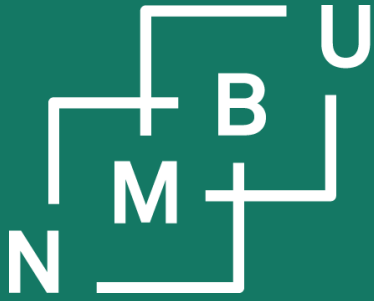
Noregs miljø- og biovitskaplege universitet
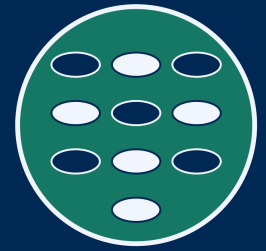
4   Impact and ethics

4.1   Plagiarism and copyright

4.2   <u>Re-use of own material</u>
(so-called "self-plagiarism")

Nightmare in Norway

Begge kopierte mye fra egen tekst. Den ene slapp straff, den andre ble utestengt i ett år

En student som kopierte 27 prosent av en tidligere innlevering, ble utestengt i to semestre. Men ved et annet universitet fikk det ingen konsekvenser at en student kopierte 40 prosent.
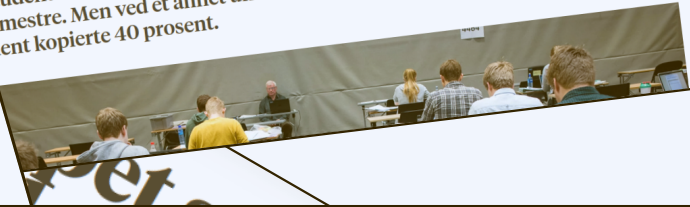
SOMMER I KHRON...

— **B...** en tils...ikelse

— Det er helt greit å gjenbruke... akademiske artikler og bø...

...et selvplagiat er

...er av populærvitenskapelige essays i ...nvendt, mener Gisle Selnes.

RETTSSAK

Ble anklaget for å skrive av sin egen eksamen — tapte i retten

Studenten gikk til sak da Høgskolen i Innlandet utestengte henne i to semestre for å ha brukt deler av egen eksamen på en konteeksamen.

Nightmare in Norway

Guideline for legitimate reuse of own material

Why is it necessary?

- Sanctions against "self-plagiarism" in Norway haven been excessive.
- The problem is being discussed as if it was a kind of plagiarism.
  - Reusing own material in an exam is very different from plagiarism.
- The real questions to be answered are:
  - Under what conditions can own material be reused in an exam?
  - How and to what degree does the reuse need to be documented?
  - How should legitimate reuse be taken into account for grading?

What is the aim of the ongoing drafting process?

- Contribute to a more viable treatment of so-called "self-plagiarism."
- Provide a document that students can propose to the main advisors of their master theses for signature in advance of the thesis work.
- It is a guideline which the main advisors are not required to sign.
  - By signing it, they would follow a recommendation from DAT390.
  - By not signing it, they still contribute to clarity about expectations.

Guideline for legitimate reuse of own material

What are the steps in the drafting process?

- Monday, 13th November 2023: Basic idea introduced in DAT390.
- Wednesday, 15th November 2023: Contribute issues/ideas by e-mail.
- Thursday, 16th November 2023: Discussion meeting of institute faculty.
  - Students are welcome; but send at most one delegate from each of the eight groups in DAT390. The meeting is at 15.45 in TF2-323b.
- Monday, 20th November 2023, in DAT390:
  - Summary of the discussions from week 46 and the way forward.
  - Unclear points taken up through further discussion if needed.
- Objective by start of the Spring 2024 semester:
  - Students have communicated about this with their main advisors.
  - The guideline document can and will be signed by a majority of the master thesis main advisors for students from data science.
- In the long run:
  - No "scary stories" about students being expelled over minor issues.
  - Data science at NMBU remains attractive for students.

29

Draft orientation for the guideline

Basic proposition 1: We accept the **definitions of plagiarism** endorsed by De nasjonale forskningsetiske komiteene (FEK), such as the one from NENT's Forskningsetiske retningslinjer for naturvitenskap og teknologi:

«Å plagiere innebærer å framstille andres ideer eller forskning som sitt eget.»

… and the one from the RINO project:

«Å fremstille andres arbeid (ideer, materiale, tekst) som sitt eget ved å utelate henvisning til opphavskilden».

Basic proposition 2: Reuse of own material cannot be plagiarism, as this would contradict the definition of plagiarism. Therefore, **there is no self-plagiarism**.

Draft orientation for the guideline

Scenario:

Researcher *A* submits a paper *P* to a scientific journal, and it is eventually accepted for publication, and published. Paper *P* contains text and ideas from term paper *Q*, which *A* wrote when studying; the paper *Q* was then only sent to his lecturer, corrected and graded by the lecturer, and sent back to *A*.
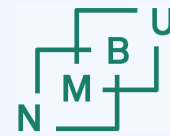
Is this plagiarism?
Is it illegitimate reuse of own material?
Should *A* have included a citation to "secret" term paper *Q* in journal paper *P*?

Recall NMBU's document:
g. **Plagiat er fusk.** *Eksempler på plagiering: Gjengivelse eller sitater fra bøker, artikler, internettsider,* **egne eller andres oppgaver**, *bruk av bilder, grafer o.l. uten kildehenvisning, sitattegn eller annen tilkjennegivelse i teksten/bildet/tegningen som viser hvor materialet er hentet fra.*

Draft orientation for the guideline

Scenario:

Researcher *A* submits a paper *P* to a scientific journal, and it is eventually accepted for publication, and published. Paper *P* contains text and ideas from term paper *Q*, which *A* wrote when studying; the paper *Q* was then only sent to his lecturer, corrected and graded by the lecturer, and sent back to *A*.

Is this plagiarism?
Is it illegitimate reuse of own material?
Should *A* have included a citation to "secret" term paper *Q* in journal paper *P*?

No – this is all nonsense.

Basic proposition 3a: It **can be legitimate to reuse own material** from exams and term papers. It can also be legitimate to reuse other own material.
Basic proposition 3b: It is **not in general obligatory to include a reference** to previous use of the same own material in an exam or term paper.

Draft orientation for the guideline

Reflection:

- Why do we conduct exams?
- Why is it a meaningful process to submit and defend a thesis?

Assessment is meant to be constructively aligned with the learning outcomes. Assessment is meaningful as a way to demonstrate that A has competency C.

If A through the same own work P can demonstrate both competency C and D, the university needs a good reason to reject using P twice for this purpose.

Basic proposition 4a: Reuse of own material from one assessment (exam, term paper, thesis) in another assessment detracts from the meaningfulness of the process **only if it interferes with showing that learning outcomes are reached**.
Basic proposition 4b: That will as a rule be the case if own content is **reused after a long time** (rule of thumb: over three years) without explicit permission.
Basic proposition 4c: Where there is no such interference, it is unproblematic.

What should we ask advisors to sign?

(3a) It **can be legitimate to reuse own material** from exams and term papers. It can also be legitimate to reuse other own material.

(3b) It is **not in general obligatory to include a reference** to previous use of the same own material in an exam or term paper.

(4a) Reuse of own material from one assessment (exam, term paper, thesis) in another assessment detracts from the meaningfulness of the process **only if it interferes with showing that learning outcomes are reached**.

(4b) That will as a rule be the case if own content is **reused after a long time** (rule of thumb: over three years) without explicit permission.

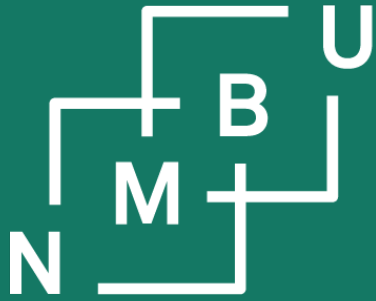**(4c)** Where there is no such interference, it is **unproblematic**.

(5a) Pursuant to NMBU's retningslinjer "behandling av mistanke om fusk," point 3.1h,[1] **permission to reuse own material** is hereby granted to student _____ for a thesis on _____.

(5b) This extends to all own material **created less than three years before** thesis due date.
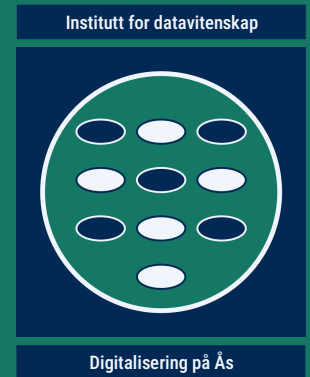
(5c) In particular, content and ideas from **the DAT390 report can be reused** for the thesis.

(6) The decision how or whether to include a source for own reused material among the cited literature references is subject to the student's discretion and academic freedom.

[1]This point states that reuse of own material can be agreed with the instructor ("avtalt med faglærer").

DAT390

Data science seminar

4 Research impact and ethics

4.1 Plagiarism vs. copyright

4.2 Re-use of own material (so-called "self-plagiarism")