

Indigenous rights in, over or to research?

Trond Trosterud & Linda Wiechetek & Maja Lisa Kappfjell

Divvun/Giellatekno centre for Sámi language technology
UiT, Norgga árktalaš universitehta

19.11.2025 at Munin Conference on Scholarly Publishing

(original picture Hilfling-Rasmussen, edited picture: Mali Brødreskift 2017, Lisens: CC-BY-SA)

Introduction

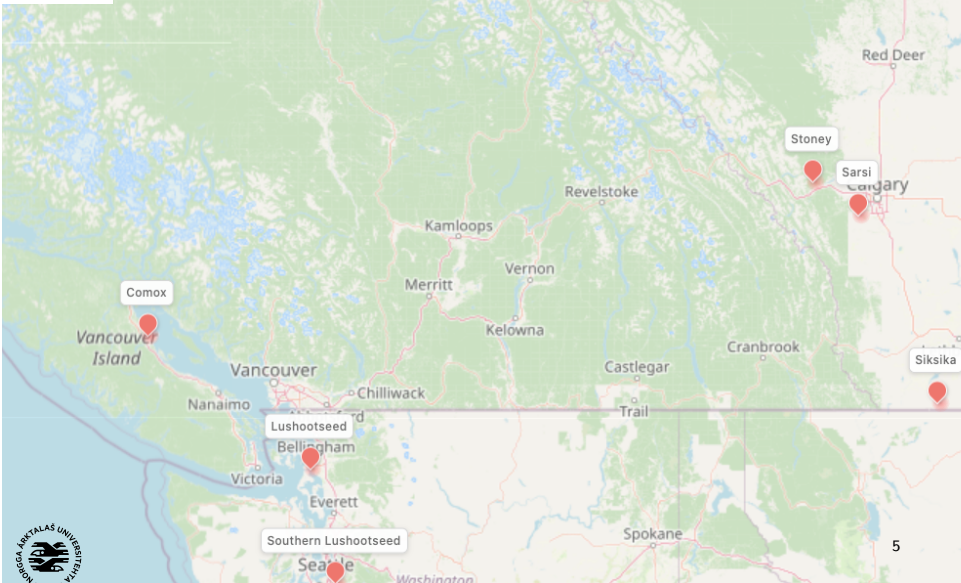
Who are we?

- Giellatekno and Divvun at UiT Norgga árkálaš universitehta
- Computational linguists and philologists
 - 50/50 Indigenous non-Indigenous group,
 - 85% (mostly North) Sámi speaking
 - In addition, we speak Lule, South, Inari Sámi
- We believe that linguistic rights are *universal human rights*
- We thus make the infrastructure GiellaLT and tools Indigenous language communities need
- Thus contributing to prevent **mother tongue illiteracy**



FAIR and CARE

The languages on the previous page are in **GiellaLT**



FAIR?

FAIR?

- The FAIR principles: Linguistic data should be **F**indable, **A**ccessible, **I**nteroperable, **R**eusable
 - Mark D. Wilkinson et al 2016: The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* volume 3, Article number: 160018
 - www.go-fair.org

FAIR?

- The FAIR principles: Linguistic data should be **F**indable, **A**ccessible, **I**nteroperable, **R**eusable
 - Mark D. Wilkinson et al 2016: The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* volume 3, Article number: 160018
 - www.go-fair.org
- FAIR was made to benefit transparent and traceable research

FAIR more detailed

	Findable	Accessible	Interoperable	Reusable
1	(Meta) data are assigned globally unique and persistent identifiers	(Meta)data are retrievable by their identifier using a standardised communication protocol	(Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation	(Meta)data are richly described with a plurality of accurate and relevant attributes
1.1		The protocol is open, free and universally implementable		(Meta)data are released with a clear and accessible data usage license
1.2		The protocol allows for an authentication and authorisation procedure where necessary		(Meta)data are associated with detailed provenance
1.3				(Meta)data meet domain-relevant community standards
2	Data are described with rich metadata	Metadata should be accessible even when the data is no longer available	(Meta)data use vocabularies that follow the FAIR principles	
3	Metadata clearly and explicitly include the identifier of the data they describe		(Meta)data include qualified references to other (meta)data	
4	(Meta)data are registered or indexed in a searchable resource			

FAIR in an Indigenous context — a dilemma

- For Indigenous language technology, paucity of data (and data creators!) is a central problem. Access to data is thus of *utmost importance*;
- Some Indigenous language communities *do not want to share* their data at all

How do we apply FAIR

- All our data, derivatives from the data and documentation thereof are openly accessible at Github ([giellalt.github.io](https://github.com/giellalt))
- **Exception:** Data restricted by copyright regulations or privacy concerns

CARE?

CARE?

- The CARE principles apply to use of data in research
- CARE requires research related to Indigenous peoples to be governed by principles of:
- **C**ollective benefit, **A**uthority to control, **R**esponsibility and **E**thics
 - Stephanie R. Carroll et al. 2020: The CARE Principles for Indigenous Data Governance. *Data Science Journal* 19:1-12

How do we apply CARE

- When choosing research topics and developing tools, we always ask *what will benefit the language community*
- Our goal is to strengthen the Indigenous language communities
 - cf. a parallel case: For most cancer research, the goal is to cure cancer
- Normal scientific methods still apply
- How to do this is the topic of the rest of the talk



Ethical considerations in Indigenous research

Gustaf Rezius (Karolinska institutet) and Hans Fjellstedt (Härjedalen) (Nordiska Museet: unknown
photographer, colored: Julius Jääskeläinen)

Ethical concerns

- Are there ethical concerns in research?

Ethical concerns

- Are there ethical concerns in research?
 - Obviously: Medical (or other!) research shall not hurt people

Ethical concerns

- Are there ethical concerns in research?
 - Obviously: Medical (or other!) research shall not hurt people
- Are there ethical concerns in the humanities?
 - Obviously: Soviet dialectologists studied Russian dialects in order to get rid of them on the way to communism
 - Today's dialectologists would be sceptical towards such a goal

Ethical concerns continued

- Are there other concerns for research on Indigenous people?
 - To a certain extent, yes:
 - To mainstream academia Indigenous people are *the other*
 - What is unproblematic for the majority may not be so for the minority

Ethical concerns continued



- Are there other concerns for research on Indigenous people?
 - To a certain extent, yes:
 - To mainstream academia Indigenous people are *the other*
 - What is unproblematic for the majority may not be so for the minority
 - Hippocrates' principles: (“do not harm; help”)
 - 1. “Do not harm”
 - Ethical violations may be due to lack of understanding
 - Challenge: to understand the consequences of the research
 - 2. “Help”
 - Understand what will help and what will not

Consequences of these ethical considerations

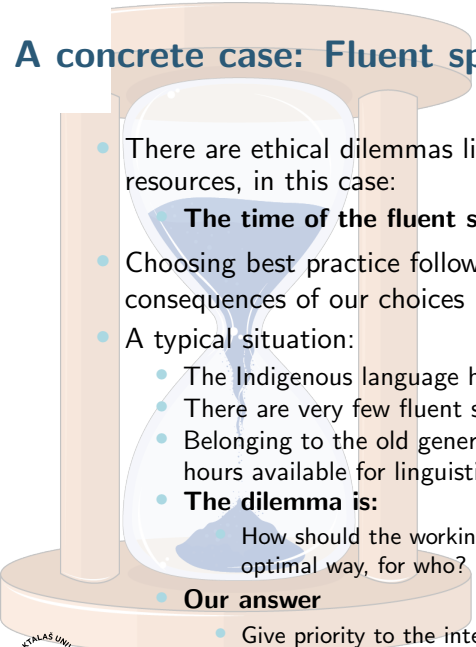
A concrete case: Fluent speakers as a resource

- There are ethical dilemmas linked to the use of scarce resources, in this case:
 - **The time of the fluent speaker**
- Choosing best practice follows from analysing the consequences of our choices

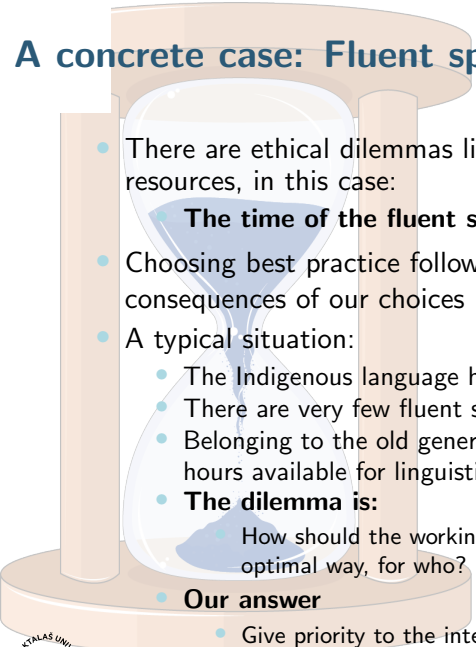
A concrete case: Fluent speakers as a resource

- There are ethical dilemmas linked to the use of scarce resources, in this case:
 - **The time of the fluent speaker**
- Choosing best practice follows from analysing the consequences of our choices
- A typical situation:
 - The Indigenous language has undergone a language shift
 - There are very few fluent speakers
 - Belonging to the old generation, they have a limited number of hours available for linguistic work before they die
 - **The dilemma is:**
 - How should the working hours of these speakers be used in an optimal way, for who?

A concrete case: Fluent speakers as a resource

- 
- There are ethical dilemmas linked to the use of scarce resources, in this case:
 - **The time of the fluent speaker**
 - Choosing best practice follows from analysing the consequences of our choices
 - A typical situation:
 - The Indigenous language has undergone a language shift
 - There are very few fluent speakers
 - Belonging to the old generation, they have a limited number of hours available for linguistic work before they die
 - **The dilemma is:**
 - How should the working hours of these speakers be used in an optimal way, for who?
 - **Our answer**
 - Give priority to the interests of the language community as a whole, rather than to the interest of the researcher

A concrete case: Fluent speakers as a resource

- 
- There are ethical dilemmas linked to the use of scarce resources, in this case:
 - **The time of the fluent speaker**
 - Choosing best practice follows from analysing the consequences of our choices
 - A typical situation:
 - The Indigenous language has undergone a language shift
 - There are very few fluent speakers
 - Belonging to the old generation, they have a limited number of hours available for linguistic work before they die
 - **The dilemma is:**
 - How should the working hours of these speakers be used in an optimal way, for who?
 - **Our answer**
 - Give priority to the interests of the language community as a whole, rather than to the interest of the researcher

Bad and better practices in language research

Bad and better practices in language research

- **Bad practice:**
 - Native language speakers working as informants for linguists doing research relevant to theoretical linguistics only
 - Making closed source learning resources for commercial e-learning companies
 - Working in platforms getting outdated so that the work is lost

Bad and better practices in language research

- **Bad practice:**
 - Native language speakers working as informants for linguists doing research relevant to theoretical linguistics only
 - Making closed source learning resources for commercial e-learning companies
 - Working in platforms getting outdated so that the work is lost
- **Better practice:**
 - Let native speakers tell or write stories, thereby giving information on both the language and on traditional life
 - Participate in work on basic descriptive grammatical and lexicographic work
- Indigenous language data are important. Collecting them should be carefully planned and they should be reused



Who gets to decide — and on what?

source: Sámediggi Sametinget, Creative Commons Attribution 2.0

Decisions on academic research

- University-financed research
 - Professors decide what to research during their allocated research time
- Externally finances research (Research councils etc.)
 - Political bodies decide on research politics and the Research Council decides on the details
- The tendency is that the role of the research funds becomes more important

Dilemmas for Indigenous research

- Should Indigenous people be able to restrict research topics?
 - A democratic society allocates its resources by means of political processes
 - It follows from this that Indigenous people should participate in the decisions on research topics
 - This will probably also give more relevant, and thus better, research

Dilemmas for Indigenous research

- Should Indigenous people be able to restrict research topics?
 - A democratic society allocates its resources by means of political processes
 - It follows from this that Indigenous people should participate in the decisions on research topics
 - This will probably also give more relevant, and thus better, research
- Are we free to conclude as our research leads us to do, or should we also here ask the Indigenous people?
 - This is a different situation
 - Disagreement, both *with* research results and *within* research, should be solved by more research, not by censoring research results

If the majority in the Sámi gets to decide

- An example from outside research
 - Should an existing South Sámi language nest open up for Lule and North Sámi children?
 - It depends on the original motivation:
 - If the goal is to ensure employment for teachers from all language groups, and to ensure all children a social arena, then it should
 - If the goal is to ensure South Sámi revitalisation, then this should **by no means** be open
 - It thus boils down to a choice
- Researchers may explain the consequences, but not make the decisions

The FAIR dilemma

- Why are so many Indigenous language communities opposed to sharing data?
 - A central reason is one of historical distrust:
 - “When nothing good has come out of contact with the colonisers so far, a natural conclusion is that no good will come out of it now either”
 - Digital colonisation is just a follow-up to land and cultural colonisation
 - “Collecting Indigenous stories is as bad as mining minerals”
 - FAIR is dependent upon CARE:
 - When sharing data damages the language community, sharing is a bad idea
 - When CARE is not obeyed, then FAIR should not be obeyed either

The CARE dilemma

- A central part of CARE is the **A**: *Authority to control*
- But who shall have authority when there is disagreement within the Indigenous people?
- As researchers, our role should be to investigate and explain the consequences of choosing the different alternatives

Questions where Indigenous people disagree

- Disagreement on the importance of language: It is ..
 - *crucial* and should be the vehicle of communication
 - important in principle, but may be kept on a symbolic level
 - irrelevant, a lost case, ...

Questions where Indigenous people disagree

- Disagreement on the importance of language: It is ..
 - *crucial* and should be the vehicle of communication
 - important in principle, but may be kept on a symbolic level
 - irrelevant, a lost case, ...
- Disagreement on language data:
 - Share *everything* OR *some* OR *nothing*!

Questions where Indigenous people disagree

- Disagreement on the importance of language: It is ..
 - *crucial* and should be the vehicle of communication
 - important in principle, but may be kept on a symbolic level
 - irrelevant, a lost case, ...
- Disagreement on language data:
 - Share *everything* OR *some* OR *nothing*!
- Disagreement on AI:
 - It will *destroy* OR *rescue* OR *remain irrelevant* our language community

A short note on AI

- The texts are taken by Big Tech already
- Focus should be on
 - evaluating, understanding and preventing negative consequences,
 - but also take advantage of possible positive outcomes

Conclusion: in, over and to

Inari Sami language nest in Ivalo (source: [Wikimedia Commons](#), Muotâ)

ANARÁŠKIELÄ SÉRVI RY.

Conclusion

- Returning to our title: *Indigenous rights in, over or to research?*
 - rights **in** research:
 - Indigenous people must participate in principled decisions on research topics
 - rights **over** research:
 - Here we are sceptical: Researchers must be free to conclude in line with their research results
 - rights **to** research:
 - Indigenous people should participate in Indigenous research — **only in this way will we get a true insider perspective**

Conclusion

- Returning to our title: *Indigenous rights in, over or to research?*
 - rights **in** research:
 - Indigenous people must participate in principled decisions on research topics
 - rights **over** research:
 - Here we are sceptical: Researchers must be free to conclude in line with their research results
 - rights **to** research:
 - Indigenous people should participate in Indigenous research — **only in this way will we get a true insider perspective**

Gæjhtoe — Gijtto — Giitu — Takkâ !