UiT The Arctic University of Norway

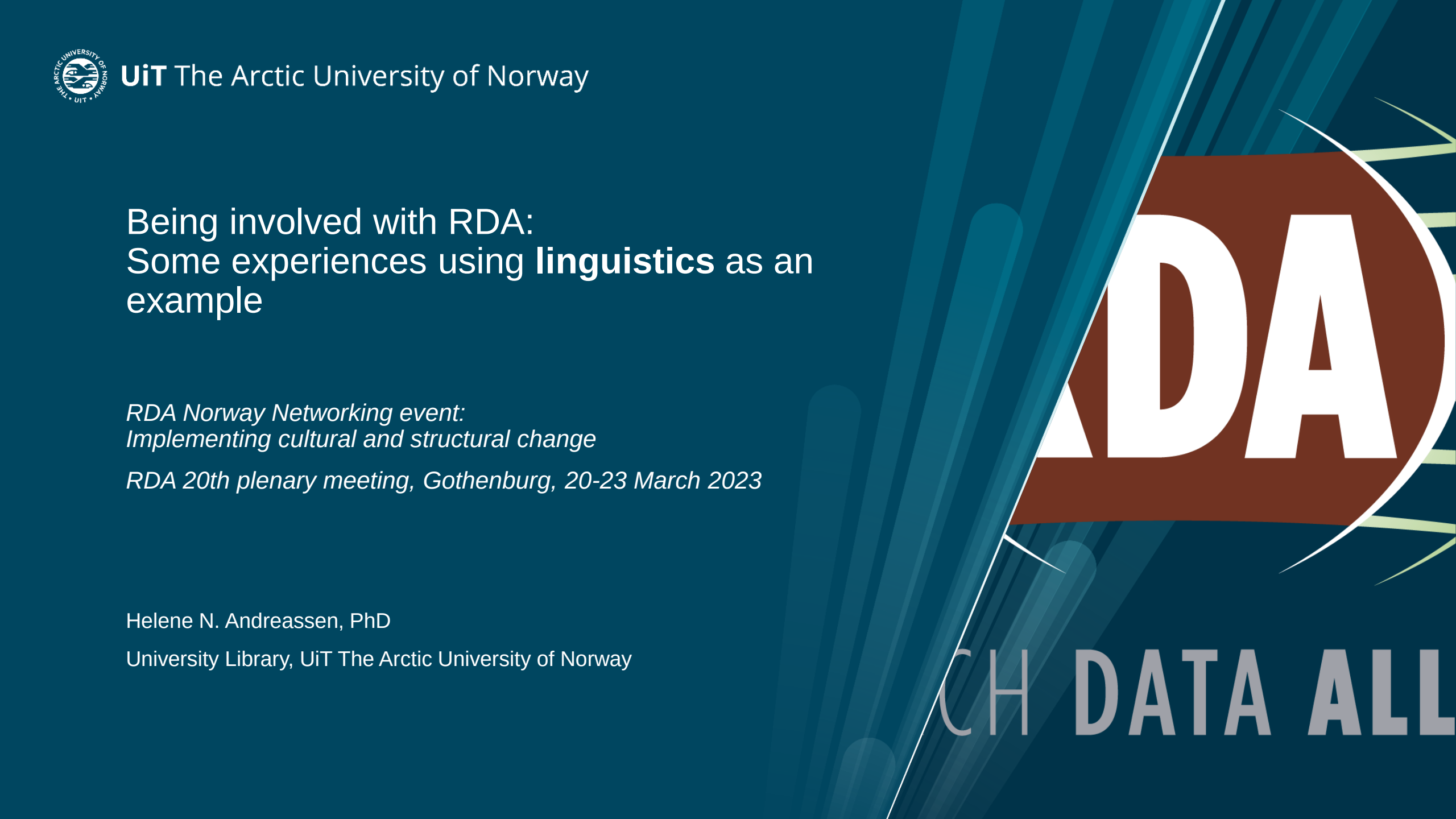
Being involved with RDA: Some experiences using **linguistics** as an example

RDA Norway Networking event:
Implementing cultural and structural change

RDA 20th plenary meeting, Gothenburg, 20-23 March 2023

Helene N. Andreassen, PhD

University Library, UiT The Arctic University of Norway

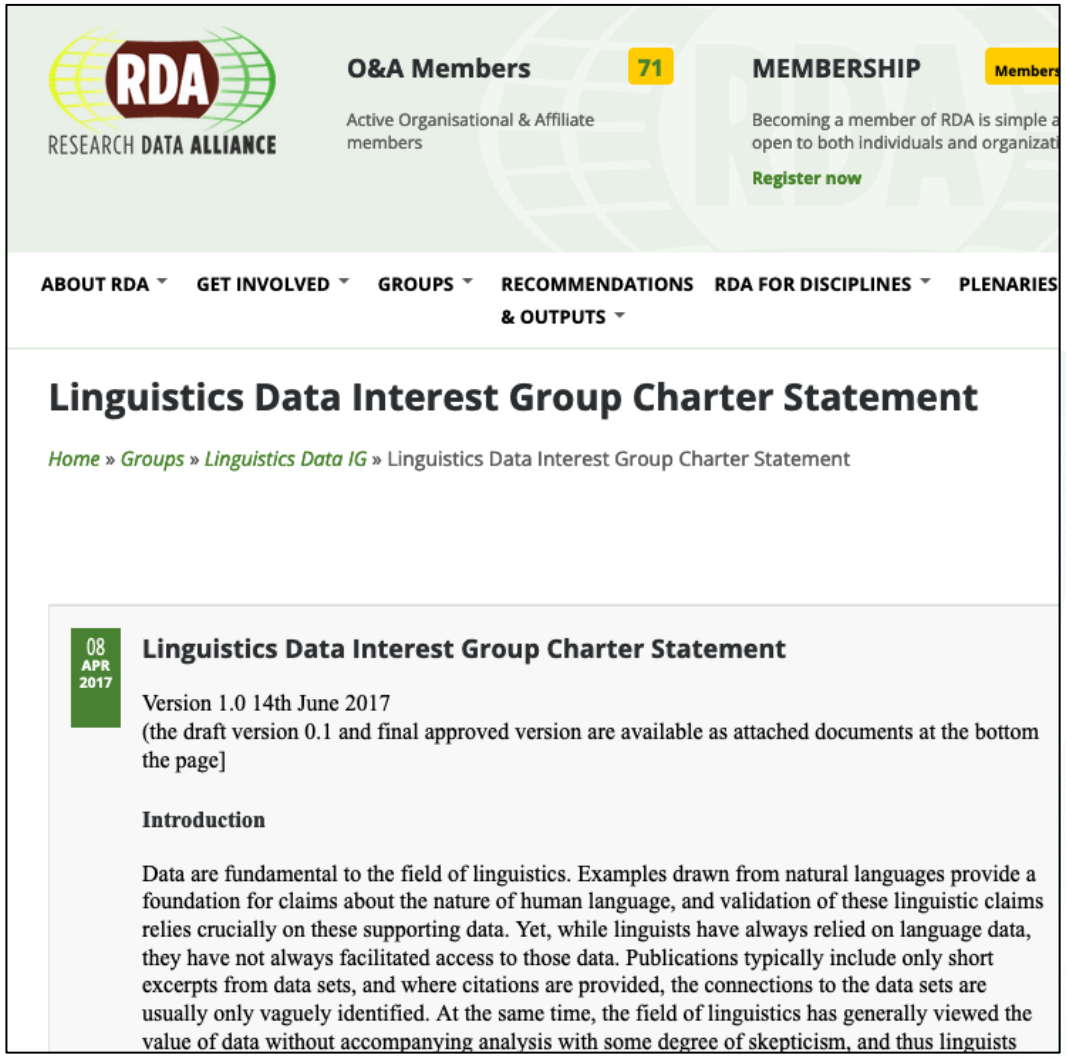


About me

- Head of teaching and learning support at the University Library (UiT) → strong focus on education & outreach.
- PhD in French linguistics.
- Worked with corpus data since 2001, as data collector, researcher, database manager and now co-leader of the project *Phonologie du français contemporain*.
- Curator in the international discipline-specific Tromsø Repository of Language and Linguistics (TROLLing).

Linguistics Data Interest Group (LDIG)

- In the last decade, many initiatives taken to improve data management practices in linguistics, to make it a more open and transparent science.
- Around 6-7 years ago, development of an international network in the US to discuss and work together on data management-related issues. TROLLing invited to attend.
- RDA Linguistics Data IG created in 2017.



LDIG deliverables

The Austin Principles of Data Citation in Linguistics

Set of guidelines to help linguists to make informed decisions regarding the accessibility and transparency of their research data.

Written and revised via asynchronous meetings. Working session & finalised during the RDA plenary in 2017.

LDIG deliverables

Tromsø Recommendations for citation of research data in linguistics

Citation model for in-text citations and bibliographic references, including commented examples and elaborated definitions.

Written and revised via asynchronous meetings. Working session and a one-day meeting during the RDA plenary in 2019. Finalised during the Munin conference in 2019.

RDA supporting output.

Tromsø recommendations for citation of research data in linguistics

RDA
RESEARCH DATA ALLIANCE

DOI: 10.15497/rda00040

Authors: Helene N. Andreassen, Andrea L. Berez-Kroeker, Lauren Collister, Philipp Conzett, Christopher Cox, Koenraad De Smedt, Bradley McDonnell, and the Research Data Alliance Linguistic Data Interest Group.

Published: December 4th, 2019

Abstract: Language and linguistics datasets are often not cited, or cited imprecisely, because of confusion surrounding the proper methods for citing them. For the use of researchers and scholars in the field working with datasets, we propose components of data citation for referencing language data, both in the bibliography and in the text of linguistics publications.

The intended audience for these recommendations consists of i) academic publishers, ii) resource providers (e.g. repositories, archives), iii) researchers citing data, and iv) researchers making data management plans, developing and depositing data, and preparing metadata. Academic publishers will have the opportunity to add these recommendations to their author guidelines for citation. Resource providers will learn which metadata and citation elements are crucial in order for data to be properly citable. Researchers using data in their publication will know how to cite these data in case publisher guidelines are underspecified. Researchers depositing data in a repository will know more about what metadata they should provide in order to publish their data so as to make them citable.

As each journal may have its own stylistic conventions, we do not address specific formats or citation styles, but rather elements of citations; however, for journals or repositories seeking to update their data citation guidance, we hope this document will be helpful. Furthermore, these recommendations are intended to be only guidelines, as we cannot account for every possibility here. This guidance is based on the Austin Principles, the FORCE11 and Research Data Alliance Joint Declaration of Data Citation Principles, and the Reproducible Research in Linguistics position statement.

Keywords: RDA Supporting Output; Linguistics Dataset; Linguistics Data Citation.

Language: English

License: Attribution 4.0 International (CC BY 4.0)

Citation and Download: Andreassen H. N., Berez-Kroeker A. L., Collister L., Conzett P., Cox C., De Smedt K., McDonnell B., Research Data Alliance Linguistic Data Interest Group. (2019). Tromsø recommendations for citation of research data in linguistics. DOI: https://doi.org/10.15497/rda00040

Related project

The Open Handbook of Linguistic Data Management

Practical guide to the principles and methods for the management, sharing and citing of linguistic data.

Published at MIT Press in 2022, with open access.

127 international authors.

The EOSC Future project is co-funded by the European Union Horizon Programme call INFRAEOSC-03-2020 - Grant Agreement Number 101017536

Current project

RDA / EOSC Future Ambassador: Broadly train linguists in the methods of open science

What:

In my period as RDA Domain Ambassador, I will work with the RDA Linguistics Data Interest Group (LDIG) to develop a Needs Analysis survey, as a first step in determining which educational efforts are needed to broadly train linguists in the methods of open science. In addition to the Needs analysis, I will work together with LDIG to promote some of our work from the last few years in the greater European linguistics community, in particular the Tromsø Recommendations for Citation of Research Data in Linguistics, which was adopted as an output of the RDA in 2019 (Andreassen et al., 2019). Another goal of the Ambassadorial period is to host a webinar in which linguistics experts discuss how FAIR relates to their own work. This webinar will then be edited into a vlog to be shared in the community.

https://www.rd-alliance.org/rda-disciplines/rda-europe-ambassadors-2022

Knowledge of a **scientific community,** its **needs** and point of **progress** towards Open Science

RESEARCH DATA ALLIANCE

... builds the **social** and **technical bridges** to enable the open **sharing** and **re-use** of data

EOSC Future

... creates a trusted platform with **open** and **FAIR data**, **resources** and **services** for all scientific disciplines

DOMAIN AMBASSADOR

RESEARCH DATA ALLIANCE

EOSC Future

Two way communication between the scientific community and RDA / EOSC Future

Peer insight and deeper **understanding** of how to promote **Open Science** and **data management**

Outreach to promote **RDA outputs** and **EOSC offerings**

Key output of domain ambassador period

- Together with LDIG co-chairs (Andrea Berez-Kroeker, U. of Hawai'i and Lindsay Ferrara, NTNU), develop and pilot a Needs Analysis survey, as a first step in determining which educational efforts are needed to broadly train linguists in the methods of open science.

- **Take home message: Having a structure such as RDA (and funding possibilities linked to this) facilitates knowledge enhancement and networking/collaboration, which again speeds up the progress of the work.**