

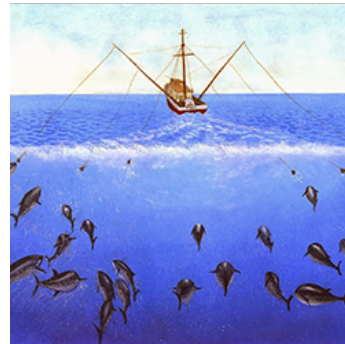


Meeting with AcqVA Aurora

16 June 2021

Helene N. Andreassen
& Philipp Konzett

University Library, UiT

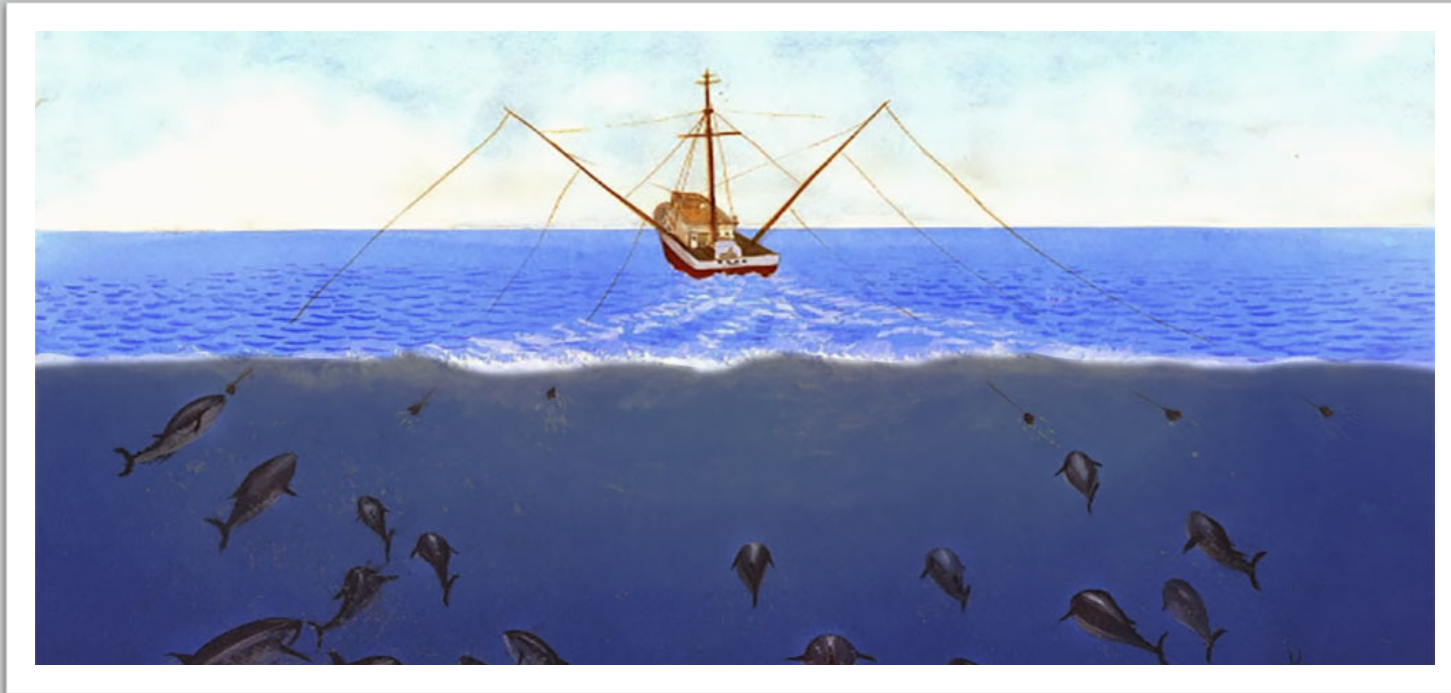


TROLLing

The Tromsø Repository
of Language and Linguistics

Meeting objectives

- Inform about TROLLing and how archiving in this repository may contribute to
 - the transparency and visibility of your research
 - meeting requirements from funders and scientific publishers
- Inform about and discuss the possibility of creating an AcqVA Aurora sub-collection within TROLLing



About TROLLing

The Tromsø Repository of Language and Linguistics

«It's very important to provide access to the data so that others can see how it was done and learn from it, or even try to replicate it. In this way, we support the scientific method and the integrity of our field overall.»



European Open Data Champions

Inspiration from influential European academics and information professionals on Open Data

[Home](#) » [Champions](#) » By sharing our data, and doing this in an open, public, community fashion, we can determine the best practices for our field

Name: Prof Laura A. Janda

Position: Professor of Russian Linguistics

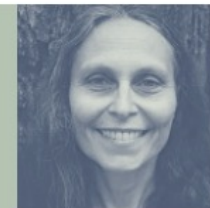
Institution: UiT The Arctic University of Norway

Country: Norway

More info: [Home Page](#); [Other](#)

ORCID ID: <http://orcid.org/0000-0001-5047-1909>

"By sharing our data, and doing this in an open, public, community fashion, we can determine the best practices for our field"



TROLLing: history

Fall 2013: The library was contacted by Laura Janda and Tore Nasset (ISK) to create a repository where linguists worldwide could archive and share their data and code

Establishment of working group and development of TROLLing; user guidelines, curation routines, outreach

June 2014: TROLLing was launched, as the first open repository for linguistic research data



TROLLing: scope

All subdisciplines of linguistics

The international community

All types of data (but open)

Raw data

Processed data

All types of supplementary material

Scripts

Experimental protocol

...



TROLLing: the infrastructure

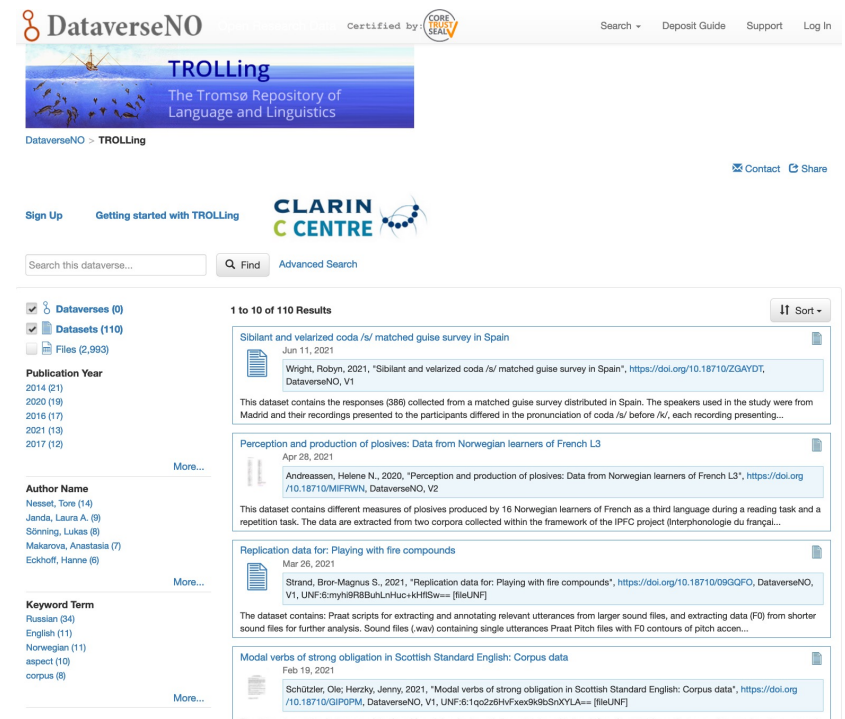
TROLLing – part of DataverseNO, a multi-institutional generic repository for open research data

Based on the international data platform Dataverse

Developed and operated at UiT by the University Library and the IT Department

10 institutional partners

Operated in line with the FAIR principles (Findable – Accessible – Interoperable – Reusable)



TROLLing: the infrastructure

Since 2020, DataverseNO is CoreTrustSeal certified as a sustainable and trusted research data repository




TROLLing: the infrastructure

Since 2018, TROLLing is a CLARIN C Centre, and harvested by the CLARIN Virtual Language Observatory (VLO)

The screenshot shows the Virtual Language Observatory (VLO) search results page. The header includes the VLO logo, navigation links (Search, Contributors, Help), and the CLARIN logo. The search bar contains the query 'westergaard'. Below the search bar, it indicates 'Showing 4 results for westergaard' and 'Results per page: 10'. On the left, there are filters for Language, Collection, Resource type, Modality, Format, Temporal Coverage, Availability, and Search options. The main content area displays three search results:

- Acquisition of definiteness marking in monolingual and bilingual Latvian-speaking children**
(Part of The Tromsø Repository of Language and Linguistics (TROLLing))
This is experimental data for the study investigating the acquisition of definiteness marking in monolingual and bilingual Latvian-speaking pre-school children, comparing them with monolingual adult controls. Pending the publication of the results, please contact the authors for the detailed description of the methodology.
- Adjectival gender agreement in monolingual and bilingual Latvian- and Russian-speaking pre-schoolers**
(Part of The Tromsø Repository of Language and Linguistics (TROLLing))
This is the experimental data for the study investigating the acquisition of adjectival gender agreement in monolingual and bilingual Latvian- and Russian-speaking children. We have investigated five groups of participants: monolingual Russian three-year-olds, monolingual Russian four-year-olds, monolingual Latvian thr...
Filters: Latvian, Russian
- Replication Data for: Universal linguistic hierarchies are not innately wired**

 **TROLLing**
The Tromsø Repository of
Language and Linguistics

DataverseNO > TROLLing > Replication Data for: Morphological variation and development in a Northern Norwegian role play register

[Contact](#) [Share](#)

Replication Data for: Morphological variation and development in a Northern Norwegian role play register

Version 1.0

Strand, Bror-Magnus S., 2020, "Replication Data for: Morphological variation and development in a Northern Norwegian role play register", <https://doi.org/10.18710/TU1GSY>, DataverseNO, V1, UNF:6:UPUQ/MQuotphpXCjM0n49w== [fileUNF]

[Cite Dataset](#) [Learn about Data Citation Standards.](#)

Dataset Metrics
17 Downloads

Description
The dataset contains: Matrix containing anonymised transcriptions and coding of spontaneous play among 7 children and R scripts used in data manipulation and to fit a binomial mixed effect model. (2020-08-20)
Abstract: This paper investigates the variation in and development of a set of morphological variables in a register known to be used by Norwegian children when engaging in role play. In this register they code-switch to something resembling the standard or Oslo variety for their in-character role utterances. The variation across variables, subjects, and age is demonstrated and discussed, and although most variables are used in the standard variants, their rates vary. A fitted binomial generalised mixed effect analysis on the most frequent variables shows that the rate of standard variants increases significantly as an effect of age. (2020-12-16)

Subject
Arts and Humanities

Keyword
Morphology, Pronouns, Nouns, Nominal inflection, Verbs, Verbal inflection, Norwegian, Child language, Play, Speech registers




Related Publication
Strand, B. (2020). Morphological variation and development in a Northern Norwegian role play register. Nordic Journal of Linguistics, 1-33. doi: 10.1017/S0332586520000219

Files Metadata Terms Versions

Search this dataset... [Find](#)

Filter by
File Type: All Access: All [Sort](#)

1 to 4 of 4 Files [Download](#)

| | | |
|---|--|--|
|  | 00_readme.txt Plain Text - 3.8 KB - Dec 19, 2020 - 3 Downloads MD5: 3b36834c6602f90320537ee6ae48a441 | Explore Download |
|  | 01_data_20200816.tab Tabular Data - 1.8 MB - Dec 19, 2020 - 5 Downloads 32 Variables, 21997 Observations - UNF:6:UPUQ/MQuotphpXCjM0n49w== | Explore Download |
|  | 02_variables_in_csv_file.txt Plain Text - 3.7 KB - Dec 19, 2020 - 3 Downloads | Explore Download |

Automatically generated
reference

Version control

Permanent identifier (DOI)

Files can be downloaded
individually or as part of a
dataset

Readme file is always on top

Standardised metadata

Open metadata, harvested by
search engines such as
Google Dataset Search

Other search engines:

<https://www.base-search.net/>
<https://search.datacite.org/>

Citation Metadata

Dataset Persistent ID

doi:10.18710/TU1GSY

Publication Date

2020-12-19

Title

Replication Data for: Morphological variation and development in a Northern Norwegian role play register

Author

Strand, Bror-Magnus S. (UiT – The Arctic University of Norway) - ORCID: 0000-0003-2381-1327

Contact

Use email button above to contact.

Strand, Bror-Magnus S. (Feide - Norwegian educational institutions)

Description

The dataset contains: Matrix containing anonymised transcriptions and coding of spontaneous play among 7 children and R scripts used in data manipulation and to fit a binomial mixed effect model. (2020-08-20)
Abstract: This paper investigates the variation in and development of a set of morphological variables in a register known to be used by Norwegian children when engaging in role play. In this register they code-switch to something resembling the standard or Oslo variety for their in-character role utterances. The variation and although most variables are used in the standard variants, on the most frequent variables shows that the rate of standard

Subject

Arts and Humanities

Keyword

Morphology
Pronouns
Nouns
Nominal inflection
Verbs
Verbal inflection
Norwegian
Child language
Play
Speech registers

Related Publication

Strand, B. (2020). Morphological variation and development in Linguistics, 1-33. doi: 10.1017/S0332586520000219 <https://doi.org/10.1017/S0332586520000219>

Language

English

Producer

UiT – The Arctic University of Norway (UiT)

Contributor

Research Group : AcqVA AURORA

Distributor

The Tromsø Repository of Language and Linguistics (TROLL)

Depositor

Strand, Bror-Magnus S.

Deposit Date

2020-08-16

Time Period Covered

Start: 2017 ; End: 2018

Date of Collection

Start: 2017 ; End: 2018

Kind of Data

experimental data; r-script

Software

ELAN, Version: several (4.9.4 to 5.7-AVFX)
R, Version: 1.3.1093
Microsoft® Excel for Mac, Version: 16.44

Google

speech registers

Last updated Download format Usage rights Topic Free

11 datasets found

PLOS

Sensitivity to linguistic register in 20-month-olds: Understanding the...

figshare.com

docx, xlsx

Updated Apr 10, 2018

PLOS

Infant Directed Speech Enhances Statistical Learning in Newborn Infants: An...

plos.figshare.com

docx, tiff

Updated Sep 13, 2016

D

Replication Data for: Morphological variation and development in a Northern...

dataverse.no

tsv, txt

Updated Dec 19, 2020

Replication Data for: Morphological variation and development in a Northern Norwegian role play register

Related Article

Explore at DataverseNO

37 scholarly articles cite this dataset ([View in Google Scholar](#))

txt(3913), txt(3812), txt(21280), tsv(1933054)

Unique identifier

<https://doi.org/10.18710/TU1GSY>

Dataset updated

Dec 19, 2020

Dataset provided by

DataverseNO

License

CC0 1.0 Universal Public Domain Dedication

License information was derived automatically

Time period covered

2017 - 2018

Area covered





Tromsø

Files Metadata Terms Versions

Search this dataset... Find

Filter by
File Type: All - Access: All - File Tag: All -

1 to 4 of 4 Files

| | |
|---|--|
|  | <p>0_Readme_Plosives_Andreassen-2020.txt</p> <p>Plain Text - 10.2 KB - Dec 18, 2020 - 1 Download</p> <p>MD5: 398b6f7a10b504936c9cb7f849c3c2c5</p> <p>This file contains the necessary information to reproduce the study.</p> |
|  | <p>1_Wordlist_IPFC_Norway.pdf</p> <p>Adobe PDF - 85.6 KB - Dec 18, 2020 - 3 Downloads</p> <p>MD5: f34fb284654d8723e12e4764ddc832b1</p> <p>Word list used for all IPFC investigation points with a Norwegian L1, in a reading task and a repetition task. The data points in this dataset are extracted from these tasks.</p> |
|  | <p>2_Values_plosives.csv</p> <p>Comma Separated Values - 47.9 KB - Dec 18, 2020 - 0 Downloads</p> <p>MD5: 7e47fc1b8d1ad2bdc8c88dbabcb582b2</p> <p>Measures for plosives /p t k b d g/ produced in word-initial and word-medial context, by informants from the Tromsø and Oslo registers. The file also contains measures of plosives produced by a native speaker serving as model in the repetition task.</p> <p>Not available until 2021-04-01</p> |
|  | <p>3_Individual profiles.pdf</p> <p>Adobe PDF - 138.9 KB - Dec 18, 2020 - 0 Downloads</p> <p>MD5: c18499140c950f4ab8e403c812180740</p> <p>Mean VOT values and ranges of VOT values obtained for each informant in the word-initial context.</p> <p>Not available until 2021-04-01</p> |

owels in Norwegian speakers

orwegian speakers of French", <http://www.oxfordjournals.org/doi/10.1093/oxfordjournals/lingua.a111111>

om Tromsø (level A2) and 6 from Oslo
 onologie du français contemporain
 d from three registers: text reading,
 consists of four Excel-files, plus a read-
 nversation" contain all potential
 ross a word boundary) observed in the
 t in the various cases. The files
 occurrences of liaison observed in the
 and (non) enchainement in the various
 e and quality of the consonant

Publish Dataset

Edit Dataset ▾

Contact O

Dataset Metr

0 Views ?

0 Downloads

0 Citations ?

- Files (Upload)
- Metadata
- Terms
- Permissions
- Private URL
- Thumbnails + Widgets
- Delete Dataset

Possibility to temporarily restrict access to files (embargo), for example until article has been published

You can administer this when submitting data

Possibility to assign pre-publishing reading or editing access, for example to collaborators or peer reviewer

Only the data set curator can administer this

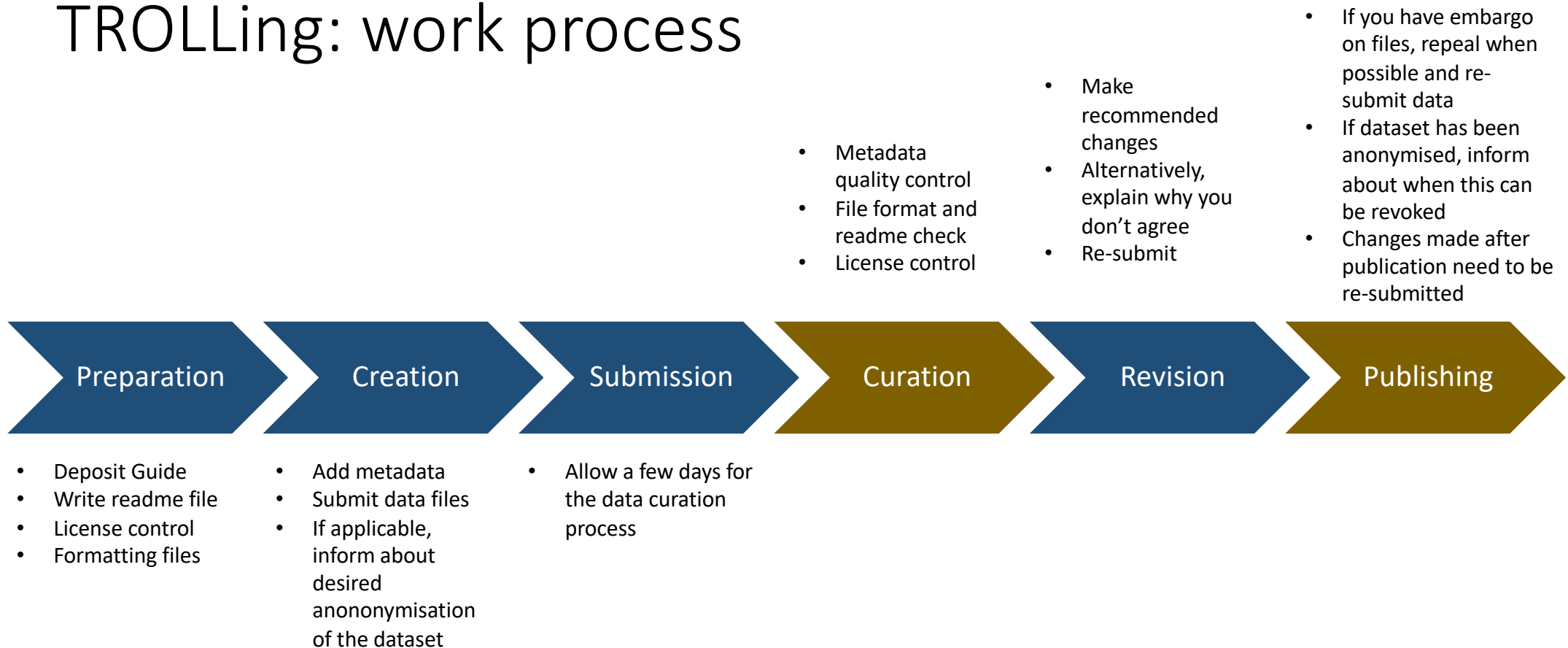
TROLLing: the contributors

- All who wish to contribute with data need to be granted status as contributor
- The contributor is responsible for ensuring that the data uploaded to TROLLing can be shared openly
 - GDPR
 - third-party data



Photo: "[Group Discussion - Flipped Learning](#)" by [mikecogh](#) is licensed under [CC BY-SA 2.0](#)

TROLLing: work process



To assist you in the process

info.dataverse.no

info: DataverseNO

About Deposit Address

Deposit Guidelines Prepare your data

Prepare your data

Deposit your data

Refer to your data

Prepare your data

Before depositing your data in DataverseNO (including the different collections, e.g. UiT Open) have to make sure your dataset(s) comply with our guidelines below. DataverseNO accepts a brief, good practice for preparing research data for archiving may be summarized as follows:

- Use consistent and comprehensible file names (see section 1 below).
- Save your data in a preferred file format(s) (see section 2 below).
- Describe your data in a ReadMe file (see section 3 below).

For more detailed guidelines, see below:

- ✓ 1 File naming
- ✓ 2 Preferred file formats
- ✓ 3 How to describe your data
- ✓ 4 File size
- ✓ 5 References

For questions, comments or suggestions, see our [support page](#).

<Help text is included in angle brackets and should be deleted before saving.>

<DataverseNO README File Template --- General --- Version: 2.1 (2021-01-07)>

This README file was generated on [YYYY-MM-DD] (YYYY-MM-DD) by [NAME].
Last updated: [YYYY-MM-DD].

GENERAL INFORMATION

// Title of Dataset:
// DOI:

// Contact Information

<The person to be contacted for questions about the dataset>

// Name:
// Institution:
// Email:
// ORCID:

<Whenever applicable, the following information should be registered in the metadata. remove fields/lines that are not applicable, and leave the rest unchanged. >

// Contributors: See metadata field Contributor.
// Kind of data: See metadata field Kind of Data.
// Date of data collection/generation: See metadata field Date of Collection.
// Geographic location: See metadata section Geographic Coverage.
// Funding sources: See metadata section Grant Information.

// Description of dataset:

<(Short) description of what the dataset is about, including reference to related projects. Should correspond to the information entered in the metadata fields Description and

METHODOLOGICAL INFORMATION

<Note! If the documentation referred to is not openly available through a persistent file(s) to the dataset.>

<Note! It may generally be considered appropriate to have a link to the methods section of the original article. See Committee on Publication Ethics (COPE) guidance: <https://publicationethics.org/resources/guidelines-new/text-recycling-guidelines-ed>

TROLLing: curators

Helene N. Andreassen

PhD in French Phonology

Responsible for the UiT training programme in research data management

Co-chair of the Linguistics Data Interest Group
(Research Data Alliance)

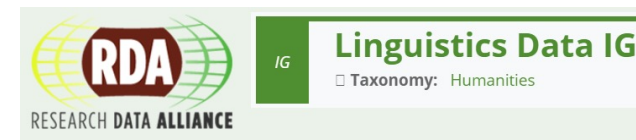
Philipp Konzett

MA in Nordic Linguistics

Part of the repository management of DataverseNO

Member of the Steering Committee of the Global
Dataverse Community Consortium





TROLLing collaboration

CLARIN – Common Language Resources and Technology Infrastructure, a European Research Infrastructure Consortium (ERIC)

COST – European Cooperation in Science and Technology: European network for Web-centred linguistic data science

SSHOC – Social Sciences and Humanities Open Cloud – a Horizon 2020 project

RDA – Research Data Alliance Linguistics Data Interest Group



TROLLing: numbers

(Jan 30, 2021)

Contributors

82 contributing authors

Representing a total of 42 research organizations

From 17 countries in 4 continents



TROLLing: numbers

(Jan 30, 2021)

Data

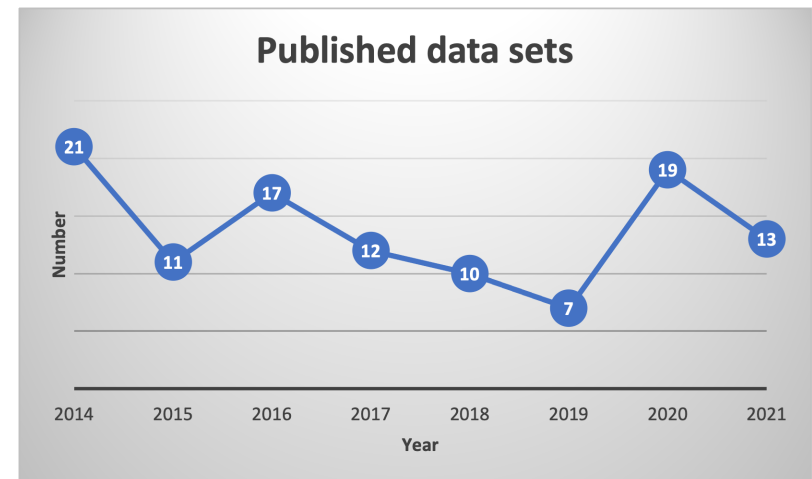
100 datasets containing 2 901 files

36 languages represented

Mostly replication data (articles and books)

Data from PhD and MA dissertations

2020: several datasets anonymised and shared with editors/peer reviewers together with a journal or book manuscript



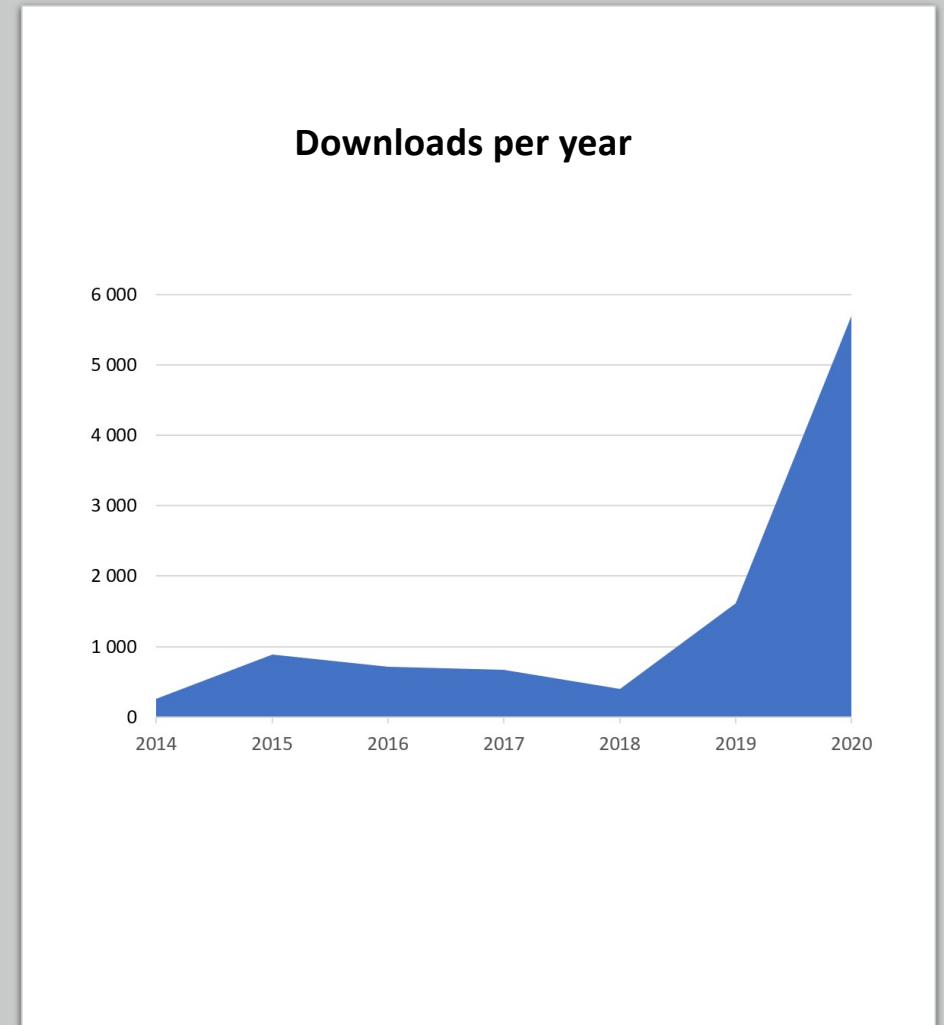
TROLLing: numbers

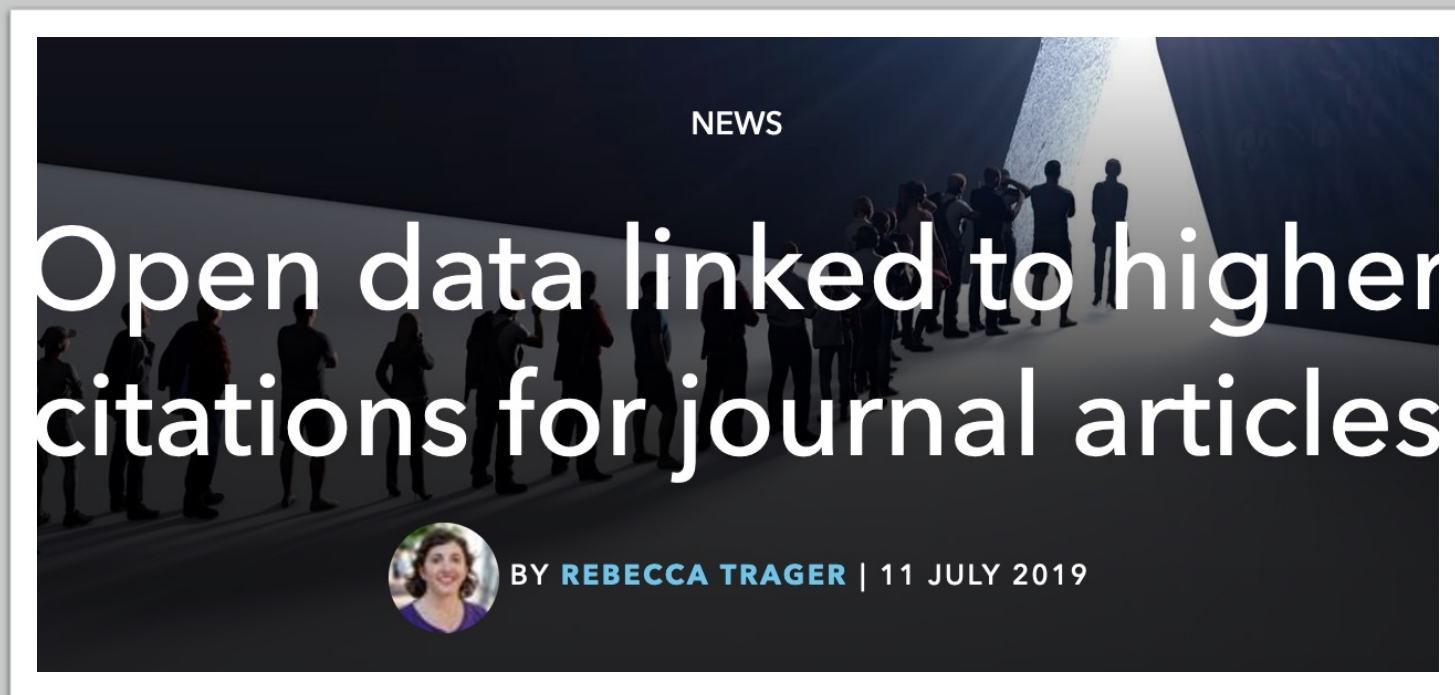
(Jan 30, 2021)

Usage

11 082 downloads

At average 111 downloads per dataset





chemistryworld.com

Citations and credit

Linking to an archived data set in your publication may increase its citation rate

AcqVA Aurora and research data

Principles and guidelines for RDM at UiT

The researcher shall ensure that the research data is securely stored, backed up, and archived, in accordance with the Information security management system at UiT (§4.3)

The researcher shall make the research data openly available for future use by all relevant users, providing this is not prevented by any legal, ethical, security, or commercial reasons. (§4.4)

[Principles and guidelines for research data management at UiT](#), effective as of 01.09.2017



Domain-specific repositories

“We recommend that researchers deposit their data in a **domain-specific repository** whenever possible. Because domain-specific repositories are based on research communities, they are most likely to provide data curation services that will enhance the value of the data for future reuse. [...] We also advocate ‘**trusted digital repositories**’ that adhere to standards for data discovery, documentation, and preservation.”

Alter, G., & Gonzalez, R. (2018). Responsible practices for data sharing. *American Psychologist*, 73(2), 146-156.
<https://doi.org/https://doi.org/10.1037/amp0000258>

“When data preservation standards or norms exist in your discipline, these should be followed. Your research funder may recommend a data centre or self-deposit archive. In order of preference: 1) Use an **external data archive or repository already established for your research domain** to preserve the data according to recognised standards in your discipline. 2) If available, use an **institutional research data repository**, or your research group’s established data management facilities. 3) Use a cost-free data repository such as Zenodo.”

OpenAIRE. (n.d.). How to select a data repository?
<https://www.openaire.eu/opendatapilot-repository-guide>

AcqVA Aurora data as a sub-collection within TROLLing

a possibility

An example



UiB Open Research Data

DataverseNO > UiB Open Research Data



Search this dataverse... [Advanced Search](#)

☒  **Dataverses (1)**

☒  **Datasets (86)**

☐  **Files (66,263)**

Dataverse Category
Research Group (1)

1 to 10 of 87 Results

 **Replication Data for: The potential for cleaner fish driven environmental control of pigmentation?**

May 19, 2021

Hamre, Lars Are; Oldham, Tina; Nilsen, Frank; Glover, I
in the salmon louse *Lepeophtheirus salmonis*: genetic
DataverseNO, V1



BIRKELAND CENTRE FOR SPACE SCIENCE

DataverseNO > [UiB Open Research Data](#) > [UiB Global Navigation Satellite System Data](#)



[Contact](#) [Share](#)

The University of Bergen Global Navigation Satellite System Data Collection contains data from four global navigation satellite system receivers at four locations (Ny-Ålesund, Longyearbyen, Hopen, Bjermøya). At each location, signals from three sets of satellite constellations are recorded: GPS, GLONASS and Galileo. Data has been recorded since 2013.

The collection is managed by the Birkeland Centre for Space Science (BCSS) at the University of Bergen. For more information about us, visit our webpage [birkeland.uib.no](#).

For more information about the collection, please see the documentation data set:

Oksavik, Kjellmar, 2020, "Documentation of GNSS Total Electron Content and Scintillation Data (60 s) at Svalbard", DataverseNO, <https://doi.org/10.18710/EA5BYX>

To refer to the whole collection, please use the following information:

Oksavik, Kjellmar, 2020. *The University of Bergen Global Navigation Satellite System Data Collection*. DataverseNO. <https://doi.org/10.18710/AJ4S-X394>.

Search this dataverse... [Advanced Search](#)

☒  **Dataverses (0)**

☒  **Datasets (65)**

☐  **Files (65,382)**

Publication Year

2020 (57)

2021 (8)

Series Name

Bjermøya OBS 60 s (8)

Bjermøya TEC 60 s (8)

Hopen OBS 60 s (8)

Hopen TEC 60 s (8)


Longyearbyen OBS 60 s (8)

1 to 10 of 65 Results

 Sort

 **GNSS Total Electron Content Data (60 s) at Ny-Ålesund in 2020**


Mar 8, 2021

 Oksavik, Kjellmar, 2021, "GNSS Total Electron Content Data (60 s) at Ny-Ålesund in 2020", <https://doi.org/10.18710/YQH715>, DataverseNO, V1

This data set contains Total Electron Content data at 60 seconds time resolution at Ny-Ålesund, Svalbard. The measurements were collected by the University of Bergen using a NovAtel GPStation-6 global navigation satellite system receiver. The measurements include signals from GPS...

 **GNSS Total Electron Content Data (60 s) at Longyearbyen in 2020**

Mar 8, 2021

 Oksavik, Kjellmar, 2021, "GNSS Total Electron Content Data (60 s) at Longyearbyen in 2020", <https://doi.org/10.18710/6PWRX0>, DataverseNO, V1

TROLLing and the road ahead

Move the repository to its own installation (outside of DataverseNO)

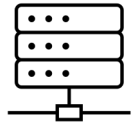
Market it to scientific publishers as a trustworthy, discipline-specific repository

Establish closer collaboration with the research communities in Norway



Repository

trolling.uit.no



User guide

info.dataverse.no



Info

info.trolling.uit.no



Support

researchdata@hjelp.uit.no



*You may also contact
[Helene](#) or [Philipp](#) directly.*

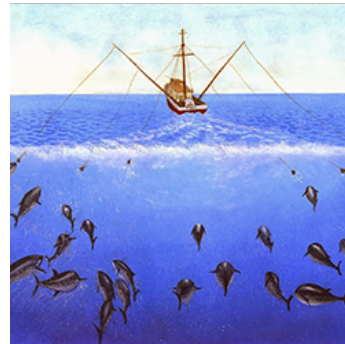


Meeting with AcqVA Aurora

16 June 2021

Helene N. Andreassen
& Philipp Konzett

University Library, UiT



TROLLing

The Tromsø Repository
of Language and Linguistics