

Helene Andreassen, Philipp Conzett, Stein Høydalsvik, Laura A. Janda, Leif Longva, Tore Nettet, Odu Obiajulu

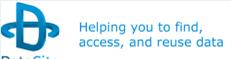
UiT The Arctic University of Norway

helene.n.andreassen@uit.no, philipp.conzett@uit.no, stein.hoydalsvik@uit.no, laura.janda@uit.no, leif.longva@uit.no, tore.nettet@uit.no, obiajulu.odu@uit.no

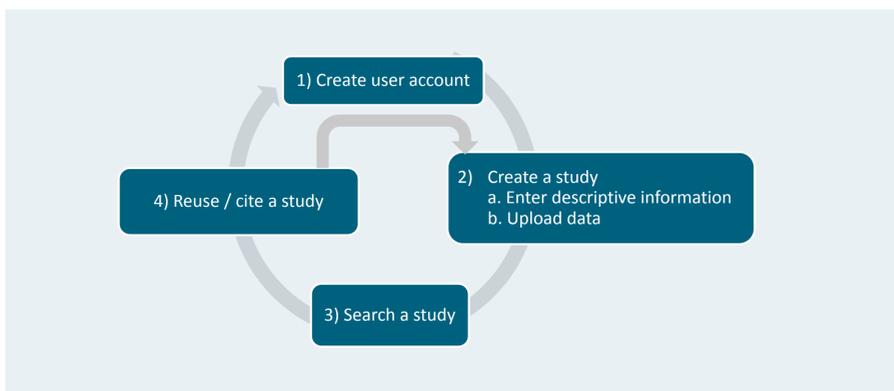
/ 1 BACKGROUND

- Successful research is a co-operative process building on dialogue and sharing of insight, data and results of earlier work
- Increasing focus on open access to scientific information, including research data (cf. guidelines issued by e.g. The Research Council of Norway, EU, OECD, and UNESCO)
- Need for digital infrastructure for research data management
- Request from academic community for digital tool for sharing linguistic research data
- University Library at UiT The Arctic University of Norway (UBT) establishes **UiT Open Research Data**, an open access digital platform for archiving, sharing, citing and reusing research data | opendata.uit.no
- First research community represented the data archive: **TROLLing – The Tromsø Repository of Language and Linguistics**, an international archive of linguistic data and statistical code
- Officially launched on 18 June 2014 | opendata.uit.no/trolling/

/ 2 HIGHLIGHTS

- Built on the Dataverse Network, a SW originated from Harvard University 
- The Dataverse metadata scheme complies with the scheme from DataCite – International Data Citation Initiative, a global association of member institutions supporting data archiving 
- Compliant with CLARIN (Common Language Resources and Technology Infrastructure in the EU), the EU research infrastructure for language-based resources 
- Assigning a permanent URL to each post (handle, being replaced by DOI) 
- Using metadata standard that ensures visibility and retrieval through international services

/ 3 PROCESSES



/ 4 PEOPLE

- **Research community:** Initiative and conceptual input
- Prof. Laura A. Janda, prof. Tore Nettet (both Department of Language and Linguistics, UiT)
- **System:** project conduct, development and management by UBT
- Stein Høydalsvik, Leif Longva, Odu Obiajulu
- **Curators:** approving new users and uploaded studies; ensuring correct registration of metadata and persistence of data format; providing user support by creating guidelines and instruction videos and by taking care of user questions and requests; provided by UBT
- Helene N. Andreassen, Philipp Conzett
- **Scientific advisory board:** Advising on board matters and strategic issues
- Hans C. Boas (University of Texas at Austin), Dagmar Divjak (University of Sheffield), Maria Polinsky (Harvard University)

/ 5 SYSTEM FEATURES

- The system is based on Dataverse Network (DVN)
- The DVN software is a multi-tier Java enterprise application that makes use of Java Server Faces in the User Interface layer, Enterprise Java Beans in the middle tier, and PostgreSQL as the database
- Cataloging fields and descriptions are indexed using Lucene Index Server, and are searchable across all dataverses and within a dataverse. Customizable metadata support
- DVN generates a formal data citation with a persistent identifier (DOI or handle) and Universal Numerical Fingerprint (UNF) for verification and validation of the dataset.
- Files uploaded in SPSS, R, STATA or GraphML offer additional sub setting and analysis services, and can be downloaded in multiple formats.
- Version control for datasets and citation to previous version(s) allowed.
- DVN can be both an OAI server and client, defining sets and harvesting from external sources.
- Supports exporting the metadata in various formats, like DDI, Dublin Core, FGDC, MARC.
- Tool for end-users to export metadata in formats like DDI, Dublin Core, FGDC, MARC.

/ 6 OUTLOOK

- UiT Open Research Data is still in its initial phase
- Several areas of improvement and development:
- Technical issue 1: Facilitate access to the data archive by providing standard log-in solutions (e.g. FEIDE, EduGain...?)
- Technical issue 2: Providing better procedures for ensuring the persistence of uploaded data
- Expanding scope to new scientific fields
- Raising awareness of the benefits of sharing scientific information among academic community, research advisors and students
- Developing teaching curriculum for best practices in research data management

/ 7 APPRAISAL



"In the age of Big Data, the creation of a general repository of datasets and statistical models for linguistic research is a welcome development. It will stimulate more research and new analyses." -- Maria Polinsky, Director of the Polinsky Language Sciences Lab at Harvard University



"TROLLing will revolutionize research in linguistics and drive the discipline forward: making data publicly available significantly reduces the risk of bogus results, avoids duplication of efforts and facilitates large-scale analysis of meticulously annotated datasets." -- Dagmar Divjak, Reader, Russian and Slavonic Studies, University of Sheffield



"TROLLing is crucial for the field of linguistics as it takes the next steps towards becoming more empirical. For the first time, it will be possible for researchers to deposit their primary linguistic data (the foundation for all research) in a central freely accessible on-line repository so that colleagues around the world have access to the same data. This invaluable resource will promote on-going academic exchange on an empirical basis." -- Hans Boas, Professor, Department Germanic Studies and the Department of Linguistics, University of Texas at Austin



"TROLLing is exactly what our field needs - with the potential to become the most useful data resource in linguistics." -- Marit Westergaard, Professor, Center for Advanced Study of Theoretical Linguistics, UiT

"I would like to recommend that scholars deposit their data at TROLLing. I strongly believe that sharing of data and methods for analysis can play a key role in the growth of cognitive linguistics. It will be beneficial for the community of linguists to have a single searchable repository rather than having data scattered about in many places." -- Laura Janda, Professor, Center for Advanced Study of Theoretical Linguistics, UiT



/ 8 GET STARTED!

<http://site.uit.no/trolling/getting-started/>

- Promotional video
- Instructional videos
- User guide
- TROLLing banner

UiT Open Research Data

Share, Cite, Reuse, Archive Research Data
Scientific data for reproducible research