

Leif Longva, Stein Høydalsvik, Helene Andreassen, Philipp Conzett, Odu Obiajulu

UiT The Arctic University of Norway

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

/ 1 HIGHLIGHTS

- An institutional archive for open research data
- A general service for the researchers at the institution, provided and managed by the University Library and the ITC-department
- Built on Dataverse, a SW originated from Harvard University
- The metadata scheme complies with the DataCite requirements, and the service is indexed by DataCite
- Assigning DOI (permanent ID) to each dataset
- Datasets curated by the library before published
- CC0 waiver as default
- UiT Open Research Data is still in its initial phase and improvements and developments in several areas of the service will be added.



/ 2 BACKGROUND

- Successful research is a co-operative process building on dialogue and sharing of insight, data and results of earlier work
- Increased focus on open access to scientific information, including research data (cf. The Research Council of Norway, EU/Horizon2020, OECD, and UNESCO)
- An institutional need for infrastructure and support services for research data management
- UiT has a longstanding commitment and support of open access to publications and are now ready to move on to open data.
- The University Library in cooperation with the ITC-department has established UiT Open Research Data, an open access digital platform for archiving, sharing, citing and reusing research data | opendata.uit.no
- A first version of the service was officially launched June 2014: *TROLLing – The Tromsø Repository of Language and Linguistics*, an international archive of linguistic data and statistical code

/ 3 THE ARCHIVING PROCESSES – AND USE

Researcher	Library	Community
Create account and Log in		
Create a dataset - enter metadata - upload files		
	Curation of dataset - controll of metadata - controll of filetypes	
	Publish the dataset	Search for datasets
Cite the dataset		Cite a dataset
		Resuse a dataset

/ 4 COMMUNITY

The research community:

- ❖ Many face requirements to archive their research data
- ❖ Some fear their research data will be lost after they retire
- ❖ The initiative and conceptual input for our first service on research data came from professors at Department of Language and Linguistics, UiT: *TROLLing*
 - Paved the way to develop a service for any research subject

The different roles:

- ❖ Researchers upload datasets and enter their metadata
- ❖ The library
 - creates guidelines and instruction videos to assist researchers and offers training and support
 - Curators at the library check and improve the quality of the metadata
 - Curators at the library check and, if needed, advice researcher to convert data files to persistent archival formats
 - Curators publish the uploaded and checked data sets
- ❖ The ITC-department runs the application, the user authentication service, monitor the servers and manage the back ups

/ 5 TRAINING AND SUPPORT

- ❖ There is a strong need for raising awareness among the researchers, academic community, research advisors and students, of the benefits of open science
- ❖ Information, training and support are important tasks for a successful implementation
- ❖ The library will develop, organize and implement these services at the institution
- ❖ Reaching researchers with guidance and support is best done individually and ad hoc – Good descriptions of the data, good metadata for best retrieval, persistent file formats.
- ❖ Teaching curriculum for best practices in research data management is developed and very well received among the PhD students
- ❖ We are right now finishing short, specialized training programs in different aspects of data managements, designed for experienced researchers, covering different phases of the research process like:
 - Data management plans – why and how to comply with funders requirement
 - Manage your data files – organize, give name to and select a long lasting file format
 - How to describe your datasets? – about metadata standards, use of Reade-me files
 - A license for your dataset – why you should add a license and pro/con for various licenses
 - Archive your datasets – why and how to comply with funders requirement
 - Publish your datasets – maximum dissemination and impact for your data, how to cite your data, the advantage (and disadvantage?) of sharing

/ 6 ORGANIZING AND PROMOTING THE ARCHIVE

- Datasets are published at the root level, as default
- Research projects or research groups may ask to have their own collection (sub-dataverse), for their common datasets
- TROLLing is an example of such a sub-dataverse
- Datasets are retrieved through searchable metadata and tags, rather than through browsing
- Efforts to raise awareness of the service among UiT researchers is needed
- The University will likely establish policies for the research data produced at the institution
- The service is in operation but will be formally launched **September 1, 2016**

/ 7 CONTACT

The service: <http://opendata.uit.no> – opendata@ub.uit.no

The institution: UiT The Arctic University of Norway, Library – <https://en.uit.no/ub>

“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



“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



“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

