



# DISCIPLINARY CASE STUDY: THE TROMSØ REPOSITORY OF LANGUAGE AND LINGUISTICS (TROLLING)

Research Data Management (RDM) support for language and linguistics provided by the University Library at UiT The Arctic University of Norway.

*Philipp Konzett, Senior Research Librarian (University Library, UiT)<sup>1,2</sup>*

## 1 Introduction

**Which discipline is the focus?** To start with, UiT's primary focus in terms of RDM support has been on language and linguistics. In fact, the development of RDM support services at our library was kicked off by a request from faculty at our linguistics department to develop and run an open archive for linguistic data, where linguists around the world could publish and share their data. This request – and the following collaborative work between the UiT Library and linguists – resulted in the Tromsø Repository of Language and Linguistics (TROLLing), launched in 2014<sup>3</sup>.

**What questions do researchers ask, and which services or types of support do they expect?** We get all kinds of questions about RDM, mostly about persistent file formats and documentation. Researchers expect a reliable archive service for their data, guidance, advice on FAIR RDM according to international standards, and best-practice recommendations.

## 2 Institutional Services

**What types of discipline-focused services are offered by UiT?** Connected to the TROLLing repository, we offer RDM training in the form of online guides, individual consulting during the curation of datasets deposited to the repository (as well as through institutional Skype and face-to-face courses), and seminars given at national and international conferences and workshops. We are also planning webinars on RDM for linguistic data in collaboration with CLARIN<sup>4</sup>, the European Research Infrastructure for Language Resources and Technology.

TROLLing is also one of the initiators of, and contributors, to the RDA Linguistics Data IG, whose main goal is to develop the discipline-wide adoption of common standards for data citation and attribution, and to improve research data management training in the discipline<sup>5</sup>.

1. <https://orcid.org/0000-0002-6754-7911>

2. [https://en.uit.no/om/enhet/ansatte/person?p\\_document\\_id=43835](https://en.uit.no/om/enhet/ansatte/person?p_document_id=43835)

3. <https://info.trolling.uit.no>

4. <https://www.clarin.eu>

5. <https://rd-alliance.org/groups/linguistics-data-ig>





Based on the TROLLing archive and associated support services, the UiT Library has extended its RDM services by developing a general repository for open research data for researchers and students at the institution, UiT Open Research Data<sup>6</sup>. At the same time, the UiT Library engaged in a more formalised collaboration with the UiT research administration and the IT department, resulting in, among other things, the UiT Research Data Portal. This web portal provides information about RDM<sup>7</sup>, as well as RDM training on different subjects and at different levels.

*Is this activity included in the library's annual plan?* Providing support services for RDM, and developing them further, is highlighted as a core activity of the library in the university's 2019 annual plan. The plan for the library also includes specific activities, such as the development of RDM webinars for linguists mentioned above.

### 3 Needs Assessment

*What are the main needs and challenges in terms of RDM support?* The most urgent need across disciplines is to convince more researchers of the benefits of making data FAIR.

At a discipline-specific level, we need to promote standards and good practices for metadata description (e.g., the use of a core set of citation metadata as well as common keyword vocabularies). As for linguistics, the UiT Library is pursuing this goal through the participation in the RDA Linguistics Data IG, the collaboration with CLARIN, and our contact with the linguistic community at UiT.

*What additional needs and gaps have been identified?* In some fields, especially within the humanities, researchers are denying or are not aware that they are dealing with research data. We need to raise awareness of RDM in these fields. We also need to provide better support for storage and handling of (modestly) sensitive data in the active phases of the research data lifecycle.

In some fields,  
especially within  
the humanities,  
researchers are  
denying, or are not  
aware, that they  
are dealing with  
research data.

### 4 Policy Context

*Is there a data policy in place at the faculty or research group level?* I am not aware of any research data policy at faculty or research group level at our university. In terms of external policies, some linguistic journals recommend or require authors to make their data openly available upon submission or publication of their article (e.g., *Glossa*<sup>8</sup>, *Language*<sup>9</sup>, *Poljarnyj vestnik*<sup>10</sup>).

*Which general policies play a role?* Linguists have to follow general policies when applicable. For example, linguists employed/funded by UiT have to follow the UiT RDM policy<sup>11</sup>, to

<sup>6</sup> <https://dataverse.no>

<sup>7</sup> <https://uit.no/researchdata>

<sup>8</sup> <https://www.glossa-journal.org/about/research-integrity>

<sup>9</sup> <https://journals.linguisticsociety.org/language/index.php/language/Notes>

<sup>10</sup> <https://septentrio.uit.no/index.php/vestnik/about/submissions>

<sup>11</sup> [https://intranett.uit.no/Content/532111/cache=20170109133727/Principles%20and%20guidelines%20for%20research%20management%20at%20UiT\\_010917.pdf](https://intranett.uit.no/Content/532111/cache=20170109133727/Principles%20and%20guidelines%20for%20research%20management%20at%20UiT_010917.pdf)



our knowledge the very first of its kind in Norway. Other relevant general policies have been issued by The Research Council of Norway<sup>12</sup>, Horizon 2020<sup>13</sup>, and others.

## 5 Implementation of Data Management

**What data and support services are currently tailored to disciplines, and who provides these services?** Our most advanced research data support services are given to linguistic research. These include a repository and associated support services such as data curation and RDM training and guidance. The library is also running a project aimed at senior researchers approaching retirement. The aim of this project is to assist researchers across disciplines in preparing their data for archiving and sharing according to best practice.

**Is there a strategy or work plan in place to sustain or expand these activities?** One of the main priorities in the UiT Library's strategic plan for the coming years is the scaling of support services in order to cover all disciplines at UiT. Subject librarians act as the primary contact and curatorial staff within their respective disciplines.

**How do these link up to disciplinary infrastructures at the national/international level?** In both its current and future strategic plans, the UiT Library commits to providing best-practice RDM support to UiT researchers, and aims to be a driving force in the advancement of open data at a national and international level. Within linguistics, the library is participating in the future work of the RDA Linguistics Data IG, and in a proposal for a Norwegian CLARIN follow-up infrastructure. Representing CLARIN, TROLLing also participates in a sub-task of the SSHOC project<sup>14</sup> which aims at improving Dataverse<sup>15</sup>, the repository software used for TROLLing. At a general level, the UiT Library is involved in the planning of a national collaboration forum for RDM, as well as a national node within RDA.

**What level of interactivity do these services offer?** TROLLing is collaborating with other stakeholders, such as the RDA Linguistics Data IG and CLARIN. TROLLing metadata are harvested by CLARIN's Virtual Language Observatory (VLO)<sup>16</sup>. Currently, one of our main priorities is to reach out to journal editors to promote TROLLing as a suitable archive for publishing underlying data for scholarly publications.

One of the main priorities in the UiT Library's strategic plan for the coming years is the scaling of support services in order to cover all disciplines at UiT.

## 6 Training & Recognition

**Is discipline-focused data management training provided?** Our training efforts are described above.

<sup>12</sup> <https://www.forskingsradet.no/om-forskingsradet/publikasjoner/2018/open-access-to-research-data2/>

<sup>13</sup> [http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-pilot-guide\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf)

<sup>14</sup> <https://sshopencloud.eu/>

<sup>15</sup> <https://dataverse.org>

<sup>16</sup> <https://site.uit.no/trolling/2018/05/23/trolling-recognised-as-a-clarin-c-centre>



**What rewards and/or recognition are offered?** The library promotes newly published datasets in TROLLing through several social media channels, e.g. Twitter<sup>17</sup>, blog<sup>18</sup> and Facebook<sup>19</sup>. UiT was also the first university in Norway to endorse the San Francisco Declaration on Research Assessment (DORA), including, among other things, research data and citations as one of the main factors to be considered in the evaluation of scholarly output<sup>20</sup>. In terms of recognition for our own efforts, one of the initiators of the TROLLing archive was nominated (and elected) as a SPARC Europe Open Data Champion<sup>21</sup>.

## 7 Outlook

**What are the next steps for your institution?** The next steps within RDM support for linguistics are to:

- Get the TROLLing archive certified through the CoreTrustSeal<sup>22</sup>;
- Develop webinars on RDM of linguistic data in collaboration with CLARIN;
- Contribute to the development of a style sheet for research data citation within the RDA Linguistic Data IG;
- Promote TROLLing among linguists, linguistic journals and other stakeholders, aiming at TROLLing to be recognised as a recommended archive for research data underlying scholarly publications within linguistics.

*This case study was carried out by LIBER's Research Data Management Working Group<sup>23</sup>.*

<sup>17</sup>. <https://twitter.com/TROLLingRepo>

<sup>18</sup>. <http://info.trolling.uit.no>

<sup>19</sup>. <https://nb-no.facebook.com/TromsoRepositoryofLanguageandLinguistics>

<sup>20</sup>. <https://sfdora.org>

<sup>21</sup>. <https://site.uit.no/trolling/2017/11/15/laura-a-janda-new-european-open-data-champion>

<sup>22</sup>. <https://www.coretrustseal.org>

<sup>23</sup>. <https://libereurope.eu/strategy/research-infrastructures/rdm>