Highlights
2018
Content

1. Updates to the Archive 1

2. Collections 2

3. Collaborations 3

Facts & Figures 4

4. Reports 5

5. Workshops & Events 7
1. Updates to the Archive

1.1 Licence Information

During the process of uploading data to 4TU. ResearchData, depositors can now choose a licence from a predefined list.

4TU.ResearchData offers the full range of Creative Commons licences for datasets, and specifically for software and code, three popular open source licences are supported.

4TU.ResearchData has adopted CC0 (Creative Commons Zero) as the default means for researchers to share their datasets to make its reuse as easy as possible without any legal barrier.

If there are reasons or circumstances when data can’t be shared with a CC0 licence, depositors can choose another, more appropriate licence for their data. [Guidance on all licence types offered can be found on the website](#).

1.2 Funder Information

To link datasets in a more structured way to funding, the archive now allows for making funding information available in dedicated metadata fields.

Depositors are asked to submit the name(s) of the funder and grant number as part of the standard metadata when they submit their dataset.

The funding information is displayed on the public dataset landing page, which includes the funder identifier from the Funder Registry.

Benefits:

- Funding organizations are able to better track the published results of their grants
- Greater transparency on who funded the research
- Research institutions are able to monitor the published outputs of their employees
1.3 Enhanced data discoverability

4TU.ResearchData has embedded now schema.org metadata in the datasets landing pages. Developed by Google, Microsoft and others, schema.orgelp search engines understand the content of web pages. This means, that datasets deposit in 4TU can now be indexed in Google Dataset Search and also follow our community recommendations for embedding machine-readable metadata in landing pages in landing pages.

1.4 Project Hartog

This year 4TU.ResearchData has started a tender process, internally known as Project Hartog, with the goal to award a contract to one supplier for the provision, implementation, maintenance and support of a new research data repository solution that will replace the current interface of the 4TU.ResearchData archive. In preparation of this process, the agreement of the 4TU.Centre for Research Data Consortium has been renewed.

2. Collections

2.1 Uploads

This year the archive expanded with 468 new datasets, 2.53 TB byte of data.

Number of datasets per institution:

<table>
<thead>
<tr>
<th>Institution</th>
<th>Number of Datasets</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>TU Delft</td>
<td>333</td>
<td>(2.24 TB)</td>
</tr>
<tr>
<td>TU/e</td>
<td>16</td>
<td>(1.72 GB)</td>
</tr>
<tr>
<td>UTwente</td>
<td>19</td>
<td>(123 GB)</td>
</tr>
<tr>
<td>Other institutes</td>
<td>99</td>
<td></td>
</tr>
</tbody>
</table>

Other depositing institutes in the Netherlands included Universiteit Leiden, NIOZ, and Wageningen University & Research. International depositors came from Italy, Serbia, Poland, Austria, Germany, Mongolia and China, amongst others.

2.2 Popular Datasets

We did some investigation into the most used datasets. As with previous years, datasets from TU Eindhoven were the most popular - they are used as part of Business Process Challenge competition, analysing event logs in institutions such as banks and hospitals.

The 5 most viewed datasets of the year were:

1. The BPI Challenge 2017 (2,111)
2. The BPI Challenge 2018 (1,566)
3. The BPI Challenge 2012 (also 1,566)
4. Real-life event logs - Hospital log (1,455)
5. Hospital Billing - Event Log (1,206)

2.3 Visits to Data Archive

2018 was the most popular year for the website (see graphs further in this document) Over 25,000 users visited the the pages of the data archive.
3. Collaborations

3.1 LIBER

4TU.ResearchData launched the LIBER (Association of European Research Libraries) Data Management Plan Catalogue. The catalogue has an initial offering of 7 data management plans that have been made publicly accessible and reviewed so to act as guidance for other researchers (and support staff) who are writing DMPs for the first time.

Additionally, over 100 people attended the LIBER-organised webinar by TU Delft Data Stewardship Coordinator Marta Teperek on data stewardship. The recording is available on Vimeo.

3.2 RDNL / Dutch Data Prijs

The bi-annually Dutch Data Prize is organized by Research Data Netherlands (RDNL consisting of DANS, SURFsara and 4TU.Centre for Research Data).

RDNL awarded, for the fifth time, a researcher/research group for making a special contribution to science by making high-quality research data accessible for new or additional research.

Prizes were awarded in 3 categories: humanities & social sciences; exact & technical sciences and medical & life sciences. The prize consists of a sculpture and € 5,000 to make the data (even) more accessible.

The three winners were
- **Portable Antiquities of the Netherlands PAN**, an online platform that makes citizens’ archaeological discoveries
- **BBMRI-omics**, a collaboration of all academic medical research centres in the Netherlands, focused on molecular big data for the discovery of disease mechanisms and biomarkers
- **OpenINTEL Active DNS Measurements**, a dataset that captures the state of the Domain Names System on a daily basis, thus creating an almost real-time picture of the internet

The HANZE (Historical Analysis of Natural Hazards in Europe) dataset, available in 4TU. Centre for Research Data, was one of the three nominated datasets in the Exact and Technical Sciences.
Facts & figures

Source of datasets uploaded in 2018

TU Delft remained the biggest contributor, but nearly a quarter of datasets were from outside 4TU

Number of Users Per Month (data.4tu.nl)

Greatest usage of the archive in 2018 was at the start of the academic year

Number of Users Per Year (data.4tu.nl)

In 2018 the numbers of users increased by over 50% showing the continues growth of interest in the archive.
4. Reports

4.1 NetCDF at the 4TU.Centre for Research Data

Written by Maria Cruz and Egbert Gramsbergen, this report provides an overview of the current data and services and explores options for 4TU.ResearchData to expand its services related to netCDF data.

The analysis is broad in scope, assessing opportunities for creating not just technical services related to storing and archiving netCDF data, but also for advice and guidance, and the advantages that could accrue from building a community of data depositors and users.

The main conclusions are that the creators and users of the netCDF data stored in 4TU. ResearchData represent heterogeneous research communities within the Earth sciences. They have different views and attitudes to data archiving and data publishing, and store netCDF datasets with very different spatio-temporal characteristics in the archive.

Ensuring that any new and current netCDF services continue to be relevant to these communities will require taking this diversity into account.

A need for training and guidance – particularly on data management aspects related to documentation, metadata standards and conventions – is the common thread uniting these communities.

This will provide the way forward for 4TU. ResearchData to build a community of data depositors and users.

Expanding technical services beyond what is already provided might be more difficult, given the diversity of the data and the depositors and users, but there are a few services that could help support community building efforts, with the desired goal being higher-quality data and increased rates of data reuse.

4.2 Research Data Management within the 4TU Research Centres

Written by Alastair Dunning and Maria Cruz, this report was based on interviews on the attitudes and practices of the 4TU. Federation Centres towards research data management. The key findings were:

1. Community building and networking are the main focus of activity for those who work for the Centres on a day-to-day basis.

Research activities are on the whole the responsibility of individual research groups connected to the Research Centres. Research data management is not addressed at a strategic level, but left to individual research groups or to individual
researchers.

2. The Centres include a broad range of disciplines and research topics, even within one Centre.

There is, therefore, a broad range of attitudes towards data and a broad range of data types and characteristics, including large datasets; commercially sensitive datasets; privacy and ethical concerns regarding data; software and its sustainability.

3. Software sustainability is an important and much discussed topic, particularly in the computer science and applied mathematics communities, but also in materials science.

Deciding which software should be sustained and how it should be maintained are important questions that are still being considered.

4. Research on human subjects and datasets including personally identifiable information or sensitive personal information are more prominent than might be expected.

Lack of transparency and reproducibility of scientific results can be an issue in these areas because the underlying datasets are often not available.

There is no easy solution to these problems, but more sophisticated sharing of output data could play a role.
5. Workshops and Events

Training Essentials for data support
Essentials 4 Data Support was given three times in 2018, and again more than 30 data supporters were trained in supporting researchers in managing their valuable research data.

The coaches of E4DS are collaborating with the Digital Curation Centre (DCC, UK) on developing a MOOC, that will be published in 2019. The team also contributed to the writing of the Open Science Training Handbook.

Electronic Lab Notebooks
On 15 and 16 of March 2018, two events dedicated to Electronic Lab Notebooks (ELNs) took place at TU Delft Library: “Digital Notebooks – productivity tools for researchers” and “Digital Notebooks – how to provide solutions for researchers?”. The workshops provided more information and knowledge on Electronic Lab Notebooks and the advantages of using them instead of the paper-based traditional notebooks.

GDPR Workshop
On Thursday 30 August and on Friday 31 August 2018 TU Delft Library hosted two events dedicated to the new European General Data Protection Regulation (GDPR) and its implications for research data. Both events were organised by the Research Data Netherlands: collaboration between the 4TU.Center for Research Data, DANS and SURF (represented by the National Research Data Management Coordination Point).

During these days experts from The Netherlands and abroad presented case studies as well as various institutional support services. Plans were made to re-convene in a year’s time to evaluate the different approaches and to share lessons learned.

Data Skills
On September 26th workshop “It’s time for open science skills to count in academic careers” was organized by 4TU.Centre for Research Data and the EOSCpilot project. During this one day workshop researchers discussed the skills they need to take the actions they are expected to make towards more open research. This taking into account that these skills may change at different stages in their academic careers.

Software Carpentries
Given that software management skills, similarly to good data management practices, are needed for Open Science, on 29 November 2018 4TU.Center for Research Data organised the first Software Carpentry workshop at TU Delft.

30 researchers from TU Delft attended the workshop and the feedback was received was overwhelmingly positive. The event was also hugely oversubscribed: the tickets sold out within the first two days and there were 45 more researchers on the waiting list.

To meet that demand, in December 2018 4TU.Centre for Research Data made a strategic decision to join The Carpentries and to become a Silver Member of the organisation. This entitles 4TU.Centre for Research Data to have four more carpentry workshops in 2019 and also to train 6 colleagues as carpentry instructors, which will increase our in-house capacity to deliver carpentry workshops.
Picture: Fred van der Zwan - Electronic designer at the Faculty of Electrical Engineering, Mathematics and Computer Science TU Delft