

# 4TU.ResearchData Strategy 2020-2023

*September 2019*

<b>Past: The Research Data Landscape</b>	<b>2</b>
<b>Current: 4TU.ResearchData - Strengths and Weaknesses</b>	<b>4</b>
Issue 1) National vs Regional, 4TU vs non-4TU, University vs non-University	5
Issue 2) Subject-Focussed vs Generic, Niche vs Broad, Superficial Engagement vs Regular Engagement	5
<b>Future: The 4TU.ResearchData Mission for 2020 to 2023</b>	<b>7</b>
<b>Vision</b>	<b>7</b>
Mission Statement 1) Provide a trusted repository for long-term access to data	8
Mission Statement 2) Build discipline-specific networks to stimulate the creation and re-use of FAIR research data	8
Mission Statement 3) Provide services for scientists using the netCDF format	9
Mission Statement 4) Provide tools for sharing and documenting data	9
Mission Statement 5) To ensure the service is clearly understood by researchers and seen as an attractive, efficient business model by consortium partners	9
<b>Appendix 1. Budget and Staffing</b>	<b>9</b>
<b>Appendix 2. Statistics</b>	<b>11</b>

## Past: The Research Data Landscape

From being a specialist concern a few years ago, research data has become a hot topic. There are now regular funder mandates on the topic (for example, the requirements for Data Management Plans for H2020<sup>1</sup> and NWO<sup>2</sup>); the promise of the European Open Science Cloud, “a one-stop-shop to find, access, and use research data and services from multiple disciplines and platforms”<sup>3</sup>; discussions on integrity are closely allied to how data is handled<sup>4</sup>; and the fundamental role of data is related to the larger reproducibility crisis in science.<sup>5</sup>

There has been a flurry of activity at national and international level. Internationally, the Research Data Alliance has now grown to over 7,000 members and 95 working groups<sup>6</sup>. The International Digital Curation Conference (IDCC) is hosted all over the world, and regularly attracts more than 300 research data and data curation experts.<sup>7</sup> National infrastructures are taking shape, such as the German National Research Data Infrastructure<sup>8</sup> or those suggested by the UK’s Research Data Task Force report.<sup>9</sup>

Within the Netherlands, the research data landscape is littered with engaged groups and networks. The NPOS (National Platform for Open Science) has written a FAIR Data Action Plan<sup>10</sup>, the KNAW (Koninklijke Nederlandse Akademie van Wetenschappen) has established an Advice Group to coordinate research responses to the topic, while the LCRDM (Landelijk Coördinatiepunt Research Data Management) is coordinating a wide network of staff engaged in the topic. These exist next to more-established organisations such as RDNL (Research Data Netherlands, which includes 4TU.ResearchData, SURF and DANS) and various SURF committees (eg the *regiegroep* RDM established by Surf and its university contact points) and services focussing on different aspects of the topic.

The new Dutch Code of Conduct<sup>11</sup> (published in 2018), with a specific section dedicated to the Data Management and the word “data” mentioned over 40 times, has added another layer of interest. The Code of Conduct has also provided an extra spur to universities. Institutions must ensure their staff have the support and infrastructure necessary for good data management.

---

<sup>1</sup> <http://bit.ly/2BW1BB6>

<sup>2</sup> <https://www.nwo.nl/en/policies/open+science/data+management>

<sup>3</sup> [https://ec.europa.eu/research/openscience/pdf/eosc\\_declaration.pdf](https://ec.europa.eu/research/openscience/pdf/eosc_declaration.pdf), p4

<sup>4</sup> <https://crln.acrl.org/index.php/crlnews/article/view/9224/10236>

<sup>5</sup> <https://www.nature.com/news/1-500-scientists-lift-the-lid-on-reproducibility-1.19970>

<sup>6</sup> <https://www.rd-alliance.org/> (as of 29 Nov 2018)

<sup>7</sup> <http://www.dcc.ac.uk/events/international-digital-curation-conference-idcc>

<sup>8</sup> <http://www.rfii.de/en/topics/>

<sup>9</sup> <https://www.gov.uk/government/publications/open-research-data-task-force-final-report>

<sup>10</sup> Awaiting url

<sup>11</sup> <http://www.vsnu.nl/files/documents/Netherlands%20Code%20of%20Conduct%20for%20Research%20Integrity%202018.pdf>

Publishers too, are developing guidelines related to the submission of data, and, in some cases, devising their own services to help researchers with the structuring of their data.<sup>12</sup>

The growing significance of data in the research lifecycle is demonstrated also at 4TU.ResearchData. 2018 saw the largest number of users, the largest number of viewers and the largest number of individual depositors since the archive started in 2010.<sup>13</sup>

But it is readily apparent that 4TU.ResearchData is now just one player in a much larger landscape - a far change from when the service launched in 2010. There now exists a litany of tools and services to assist researchers during the research lifecycle. To name just a few: services for creating FAIR data (such as Go-FAIR, FAIRdom), a panoply of electronic lab notebooks (eg, RSpace, Hivebench, eLABJournal, LabFolder), large distributed clusters of data (Elixir being one of the best examples), and all kinds of subject-focussed repositories (eg, Dryad, Pangea). To this may be added the many thousands of tools that researchers develop themselves for managing data, and now crowd the corridors of GitHub.

The broader research context also deserves consideration. In the research community, the wider concept of Open Science is often seen as an external driver (ie a top-down thing forced on researchers by funders and other stakeholders). It is occasionally framed as a hindrance to the goal of high-quality research. Researchers can be unfamiliar with concepts such as data management, or reluctant to share their data, motivated either by a fear of 'scooping' or the perceived time needed to fully document data.

Academics who work closely with industry are also resistant to invocations to publish research data, suggesting this will undermine the commercial basis they are working in.<sup>14</sup> And despite some clear examples, the practice of re-using others' data is still far from being embedded in typical research practice.<sup>15</sup> Any archive developing its future plans must acknowledge this resistance and find out ways to alleviate it.

---

<sup>12</sup> <https://www.springernature.com/gp/authors/research-data-policy>

<sup>13</sup> Statistics are shown in Appendix 2

<sup>14</sup> <https://sciencebusiness.net/news/industry-universities-and-national-funding-agencies-unite-oppose-eu-plan-make-more-research>

<sup>15</sup> <https://datasupport.researchdata.nl/en/start-the-course/iv-user-phase/data-reuse/>

## Current: 4TU.ResearchData - Strengths and Weaknesses

4TU.ResearchData has been in existence since 2010. In terms of institutionally based services to support data management, this span of eight years is a reasonably long period. It has allowed 4TU to develop strengths in its core tasks as an archive. Over 8,400 datasets have been documented and published; there is a strong collection of data, particularly related to the Dutch environment, land and coast<sup>16</sup>; a simple, unfussy process for uploading data is in place; the 4TU.ResearchData metadata reviewers can check the quality of the metadata and the (technical) integrity of the files being uploaded; the OpenDap services provide specific access to data encoded in the netCDF format; more broadly, 4TU.ResearchData has the Data Seal of Approval for trusted digital repositories.

Nevertheless, the functionality currently provided by 4TU.ResearchData is largely the same as it was when the service started. In the meantime, the demands of researchers in respect to research data management have evolved. The existence of the subject-based repositories and electronic lab notebooks mentioned in the previous section demonstrate that researchers are thinking about the whole data lifecycle, and within the context of their own subject communities.<sup>17</sup> It is clear that the 4TU.ResearchData needs to react to these demands.

In 2018 a procurement process was begun to address many of the functional deficits of the archive.<sup>18</sup> It is planned that during 2020, 4TU.ResearchData will be able to offer an Application Programming Interface (API), restricted access, links to GitHub for the archiving of software, advanced statistics on usage, private spaces for sharing data prior to publication and customised metadata, amongst various other pieces of improved functionality. This will go a long way to ensuring 4TU.ResearchData remains a trusted and reliable location for publishing and archiving scientific data.

But there remain areas where 4TU.ResearchData requires strengthening. Most importantly, there exists external (and indeed internal) confusion over the mission and, in particular the scope, of 4TU.Research Data. In some cases, this is a problem of perception; in other cases, the archive's own direction needs alignment. These problems are unpacked in the two sections below.

---

<sup>16</sup> A map of the Netherlands with all entities registered in 4TU.ResearchData is available at <https://data.4tu.nl/repository/resource:location-gn-2759793/geo/html?radius=175&mode=circle>

<sup>17</sup> Borghi, John, Stephen Abrams, Daniella Lowenberg, Stephanie Simms, and John Chodacki. 'Support Your Data: A Research Data Management Guide for Researchers'. Research Ideas and Outcomes 4 (9 May 2018): e26439. <https://doi.org/10.3897/rio.4.e26439>.

<sup>18</sup> The call for tenders is currently available at <https://platform.negometrix.com/PublishedTenders.aspx?tenderId=115127&companyId=135> (though the link may not be durable).

## Issue 1) National vs Regional, 4TU vs non-4TU, University vs non-University

As an archive, 4TU.ResearchData privileges researchers from its three paying members: TU Delft (TUD), TU Eindhoven (TUE) and the University of Twente (UT). Researchers from those institutions can archive 1TB of data per year for free. Non 4TU-researchers can also deposit data but only up to 10GB for free (and at a one-off cost of 4.50 Euros per GB after that). Nevertheless, this does not deter many researchers from uploading - 21% of the uploads in 2018 came from outside Delft, Eindhoven and Twente. More broadly, the message that 4TU.ResearchData gives is that it is open to researchers internationally. Via RDNL, 4TU.ResearchData also promotes itself as a national data service.

Yet this role as a national, or even an international data service, is passive rather than active. There is no specific outreach done to attract researchers datasets from non-4TU universities, (or indeed other institutions) particularly in the Netherlands. There are numerous scientific faculties outside the 4TU members that have no local archiving services; the same is true of other technical universities around Europe.

Therefore this strategy proposes to deepen 4TU.ResearchData's collaboration with other universities, both in the Netherlands and internationally. In the Netherlands in particular, this is correct moment to make such a statement - the landscape is very much evolving with few fixed landmarks in place. Such collaboration will help build the quantity of science, engineering and design data published by the archive. Given that the core data publication infrastructure is in place, there are no significant extra costs for archiving the type of long-tail datasets that 4TU.ResearchData already concentrates on. And by widening its network, this allows 4TU.ResearchData to locate and work with institutions who may wish to become paying members, even if not part of the 4TU Federation.

Related institutions and organisations with similar issues should also become part of the network. In particular, non-profit organisations and private companies that create and use large amounts of data related to the Dutch environment, could benefit from having a trusted partner to archive and share their data.

As a consequence of the extension of its network, 4TU.ResearchData should refresh its branding, helping provide clearer messages of what services it provides, and to whom.

## Issue 2) Subject-Focussed vs Generic, Niche vs Broad, Superficial Engagement vs Regular Engagement

Anecdotal feedback from staff and users suggests that the current ‘technical sciences’ focus of 4TU.ResearchData is both an advantage and a disadvantage. An advantage because it is a catch-all term that many researchers can relate to.<sup>19</sup> But a disadvantage because it currently provides none of the advanced functionality (tailored metadata, data analysis, subject-specific peer reviews) that archives focussed on specific disciplines can.

This lack of advanced functionality makes it difficult for 4TU.ResearchData to have strong connections with specific research communities - the archive is viewed as a tool to be occasionally used but not for helping with the day to day business of research.

Therefore the strategy deploys some specific methods for deepening engagement with researchers. 4TU.ResearchData does not have the capability of creating new complex infrastructure to assist complex analytic or management processes. But it can help in the discussion around these processes and assisting in the development of specific communities. This can be achieved by continuing to offer training in specific areas (eg in using the netCDF format), contributing to the burgeoning discussions about implementing FAIR data in different disciplines and exploring the possibilities for peer review of data.

---

<sup>19</sup> There is ongoing work with RDNL to have a coordinated collections policy between 4TU.ResearchData, DANS and Surf, based on the high-level distinctions between the domains the archives are engaged in.

## Future: The 4TU.ResearchData Mission for 2020 to 2023

The mission responds to the fact that for many time-poor researchers, familiar with the usability of commercial digital services, the process of uploading data needs to be as smooth and painless as possible. But they also acknowledge that change is occurring; **there are more researchers engaging with data as part of their research and adapting their workflows** and pipelines accordingly. It is crucial for the success of 4TU.ResearchData that it provides a service that can connect to these new types of workflows, facilitating the flow of data from data creation all the way through to publication.

The strategy also recognises that **researchers' affiliations are not just to their university, faculty and department but the global community that their own subject constitutes**. In many cases, 4TU.ResearchData will not be in a position to offer the specific tooling required for particular disciplines. However, 4TU.ResearchData can contribute to the discussion about the development of FAIR standards for data within specific disciplines, particularly those where a mass of data has already been archived (Atmospheric Sciences, Climate and Climate Change, Physical Geography and Environmental Geoscience, Civil Engineering).<sup>20</sup> And following that, it needs to be able to enable access to the data via the domain-specific tools and services managed by other parties. In this respect, connecting to European research infrastructures, particularly those allied to the European Open Science Cloud, will be a crucial piece of work in the future.

The strategy also points to 4TU.ResearchData **engaging with stakeholders in science, engineering and design throughout the Netherlands** and further afield. Given that much of the infrastructure is in place identifying both researchers that wish to deposit data, and institutions that wish to join as members. Therefore a plan of engagement and execution will be developed during the lifetime of this strategy.

### Vision:

**4TU.ResearchData is part of the data-driven evolution of research. It provides trusted services for publishing and sharing FAIR data in science, engineering and design.**

### Mission:

---

<sup>20</sup> For the subject areas most commonly featured in the archive, see Appendix 2

1. To **provide a trusted repository for long-term access to data**, for use by all researchers in science, engineering and design
2. To help shape the re-use of research data by building **networks for training and collaboration in relevant disciplines** in science, engineering and design
3. To provide specific **services for scientists using the netCDF format**
4. (For researchers at consortium partners) To provide **tools for sharing and documenting data** prior to publication
5. To ensure the **service is clearly understood by researchers** and seen as an **attractive, efficient business model** by consortium partners

The 4TU.ResearchData Mission for 2020 to 2023 contains five areas that build on current strengths and address the issues described above.

### Mission Statement 1) **Provide a trusted repository for long-term access to data**

4TU.ResearchData has provided a trusted data service for nearly ten years. It will remain an open service for all researchers in science, engineering and design who wish to upload or download research data. The current cost model for depositing data will be reviewed to ensure the costs for researchers are realistic while still privileging the position for researchers at consortium members.

The collection policy will focus on small and medium size datasets in the disciplines commonly found at the technical universities - science, engineering and design. Within the Dutch context, data from other domains can be archived at DANS, and large data sets at SURF.

### Mission Statement 2) **Build discipline-specific networks to stimulate the creation and re-use of FAIR research data**

Many disciplines are still at very early stages in being able to create FAIR data, that can be embedded and reused in research. Much work needs to be done in creating awareness, building standards and tools, and developing interoperable datasets

4TU.ResearchData will create networks of data stewards, researchers and other relevant contacts to help stimulate these issues. It will help push forward the appropriate training for researchers and data stewards, and assist with the development and promotion of relevant standards, and provide access to tools (eg FAIR Data points)

### **Mission Statement 3) Provide services for scientists using the netCDF format**

NetCDF is the most popular format that the data archive publishes. 4TU.ResearchData will continue to make the netCDF data available for online querying via its OpenDap servers and will continue to develop tools and work with the netCDF community.

### **Mission Statement 4) Provide tools for sharing and documenting data**

While many researchers simply need to archive their data, there are also those who need tools to help with data management during research. 4TU.ResearchData will provide services to connect to these new types of pipelines and workflows for storing, documenting and managing data, facilitating the flow of data from creation to publication.

These tools will be available only to researchers from consortium partners.

### **Mission Statement 5) To ensure the service is well promoted to researchers and seen as an attractive, efficient business model by consortium partners**

To ensure that the core services of 4TU.ResearchData are recognised as being open to all researchers in science, engineering and design, a rebranding is required. More broadly, 4TU.ResearchData will develop clear, consistent messages for all the services it develops.

4TU.ResearchData will develop its network. It will have a particular focus on Dutch institutions with sizeable faculties of science, engineering and design and technical universities internationally.

As part of an updated business model, immediate costs for current consortium members will be reduced. Fees for new partners - perhaps joining as associate partners - will be determined.

## Appendix 1. Budget and Staffing

TUe and UT currently pay membership of 125,000 Euros per year. TU Delft pays 250,000 a year. The rest of the budget is made up from TU Delft's payment to the TU Delft Library as part of the library's work on behalf of the university

### 4TU.ResearchData Budget Estimate 2019-2023



## Staffing

As of April 2019, the staffing consists of the below. It consists of 6.2FTE.

Head: Alastair Dunning (0.5 FTE)

Manager: Madeleine de Smaele (0.8FTE)

Technical Manager: Ardi Nonhebel (0.9FTE)

Tester: Arie Braat (0.4FTE)

Data Engineer: Egbert Gramsbergen (0.8FTE)

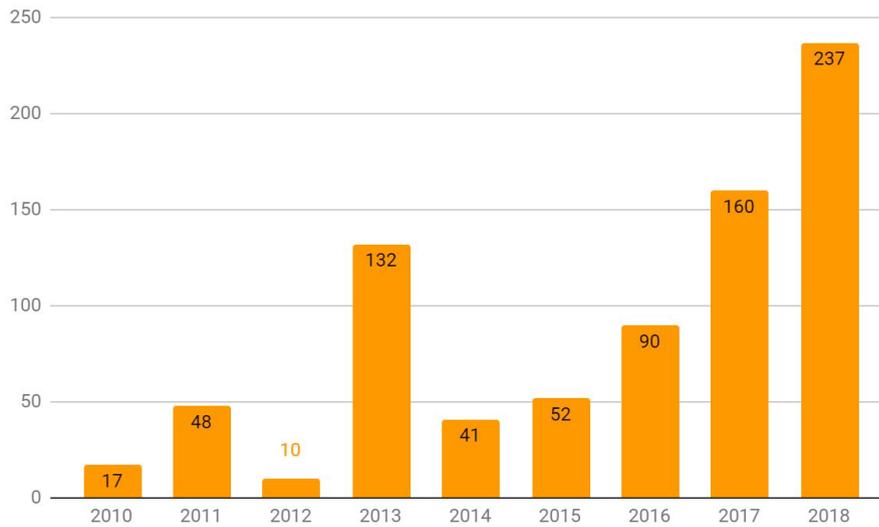
Metadata Reviewers: Jan van der Heul, Eric Rumondor (2FTE)

Trainer: Ellen Verbakel (0.4FTE)

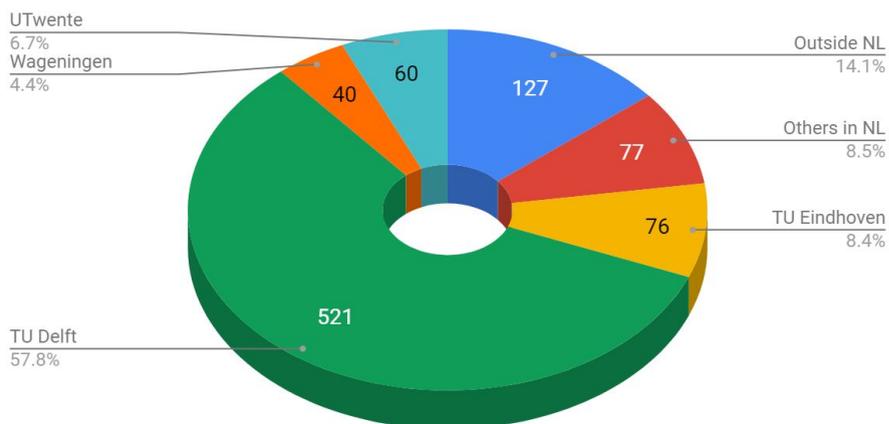
Communications: Deirdre Casella (0.4FTE)

In addition, a 1FTE developer is externally appointed.

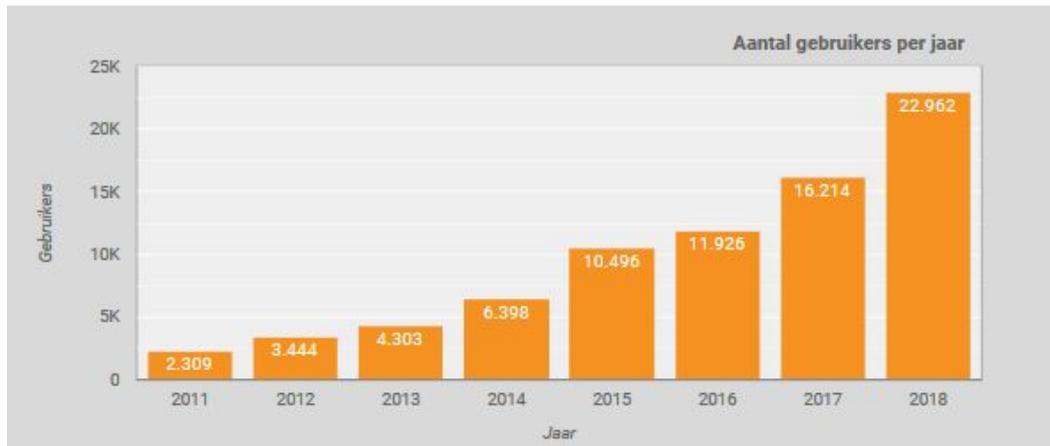
## Appendix 2. Statistics



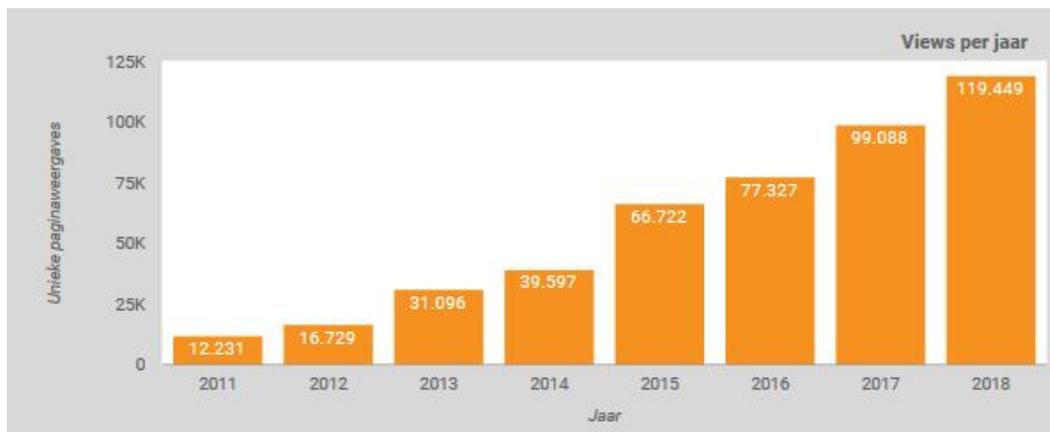
**Chart 1. Number of collections deposited per year. (2010 - 2018)** *Note: a collection can include multiple datasets for time-based data, eg climate data*



**Chart 2. Number of collections deposited per institution (2010 - April 2019).** *Note: a collection can include multiple datasets for time-based data, eg climate data*



**Chart 3. Number of users on data archive pages (<https://data.4tu.nl>) per year from 2011 to 2018**



**Chart 4. Number of page views on data archive pages (<https://data.4tu.nl>) per year from 2011 to 2018**

**5. List showing 15 most popular subject areas per number of datasets. *Note: datasets fall under more than one subject area***

3013 [Atmospheric Sciences](#)

2937 [Geomatic Engineering](#)

2922	<u>Climate and Climate Change</u>
2917	<u>Atmosphere and Weather</u>
1341	<u>Physical Geography and Environmental Geoscience</u>
177	<u>Civil Engineering</u>
79	<u>Geology</u>
76	<u>Information Systems</u>
67	<u>Business and Management</u>
66	<u>Interdisciplinary Engineering</u>
66	<u>Urban and Regional Planning</u>
50	<u>Environmental and Natural Resource Evaluation</u>
50	<u>Land and Water Management</u>
41	<u>Condensed Matter Physics</u>
36	<u>Environmental Engineering</u>
31	<u>Materials Engineering</u>
30	<u>Quantum Physics</u>

**6. 20 most popular datasets (Page impressions from 2011 to 2018). Note: This excludes access to datasets via the OpenDap server for netCDF**

1. [Real-life event logs – Hospital log](#) – Eindhoven University of Technology (8,832 page impressions)
2. [BPI Challenge 2012](#) – Eindhoven University of Technology (5,598)

3. [Receipt phase of an environmental permit application process \('WABO'\), CoSeLoG project](#) – Eindhoven University of Technology (4,761)
4. [BPI Challenge 2017](#) – Eindhoven University of Technology (4,646)
5. [Road Traffic Fine Management Process](#) – Eindhoven University of Technology (3,438)
6. [BPI Challenge 2015](#) – Eindhoven University of Technology (3,290)
7. [BPI Challenge 2014](#) – Rabobank Nederland (2,963)
8. [Large Bank Transaction Process](#) – Universitat Politècnica de Catalunya (Barcelonatech) (2,802)
9. [BPI Challenge 2016](#) – UWV (2,550)
10. [BPI Challenge 2015 Municipality 1](#) – Eindhoven University of Technology (2,312)
11. [IDRA weather radar measurements – all data](#) – TU Delft, Faculty of Civil Engineering and Geosciences
12. [Production Analysis with Process Mining Technology](#) – NooL – Integrating People & Solutions
13. [BPI Challenge 2013](#) – Volvo IT
14. [Environmental permit application process \('WABO'\), CoSeLoG project](#) – Eindhoven University of Technology
15. [Activities of daily living of several individuals](#) – Universitat Politècnica de Catalunya, Barcelona, Spain
16. [Signatures of Majorana fermions in hybrid superconductor-semiconductor nanowire devices](#) -TU Delft
17. [Sepsis Cases](#) – Event Log – Eindhoven University of Technology, Department of Mathematics and Computer Science
18. [CFD in drinking water treatment](#) – TU Delft, Faculty of Civil Engineering and Geosciences, Department of Water Management
19. [BPI Challenge 2017 – Offer log](#) – Eindhoven University of Technology
20. [Loan application example](#) – Eindhoven University of Technology

## Acknowledgements:

This document has been written by the 4TU.ResearchData team, led by Alastair Dunning. Madeleine de Smaele, Marta Teperek, Leon Osinski, Sjef Ollers, Maarten van Bentum and Paula Martinez Lavanchy provided specific input. Egbert Gramsbergen, Maria Cruz, Ardi Nonhebel, Lotte Melenhorst, Arie Bratt, Eric Rumondor, Deirdre Casella, Jan van der Heul, Jasmin Bohmer and Ingeborg Verheul all provided helpful ideas and feedback at different stages of the writing. Thanks also go to the 4TU.ResearchData Board (Marjolein Drent, Han Heijmans, Paul Hofman, Wilma van Wezenbeek) who have provided regular strategic direction and guidance.

Thanks is also due to the external reviewers of the document

- Christina Elsenga (Research Data Consultant, University of Groningen)
- Henk van den Hoogen (Programme Manager, Research Support, Maastricht University)
- Karel Luyben (Chairman, European Open Science Cloud)
- Jacqueliijn Ringersma (Coordinator Data Management, Wageningen University & Research)
- Melle de Vries (Head, Information Policy Department, KNAW)
- Ingrid Dillo (Deputy Director, DANS)
- Pieter Schipper (Policy Advisor, NWO)
- Jacko Koster (Programmanager Implementation Datamanagement, Leiden University)
- Martine Pronk (Head, Academic Services, Utrecht University)

This document is licenced as CC-BY-SA