

# Building research data repository for the future

Marta Teperek (and others!)

16 January 2023

Institutional RDM Services Workshop

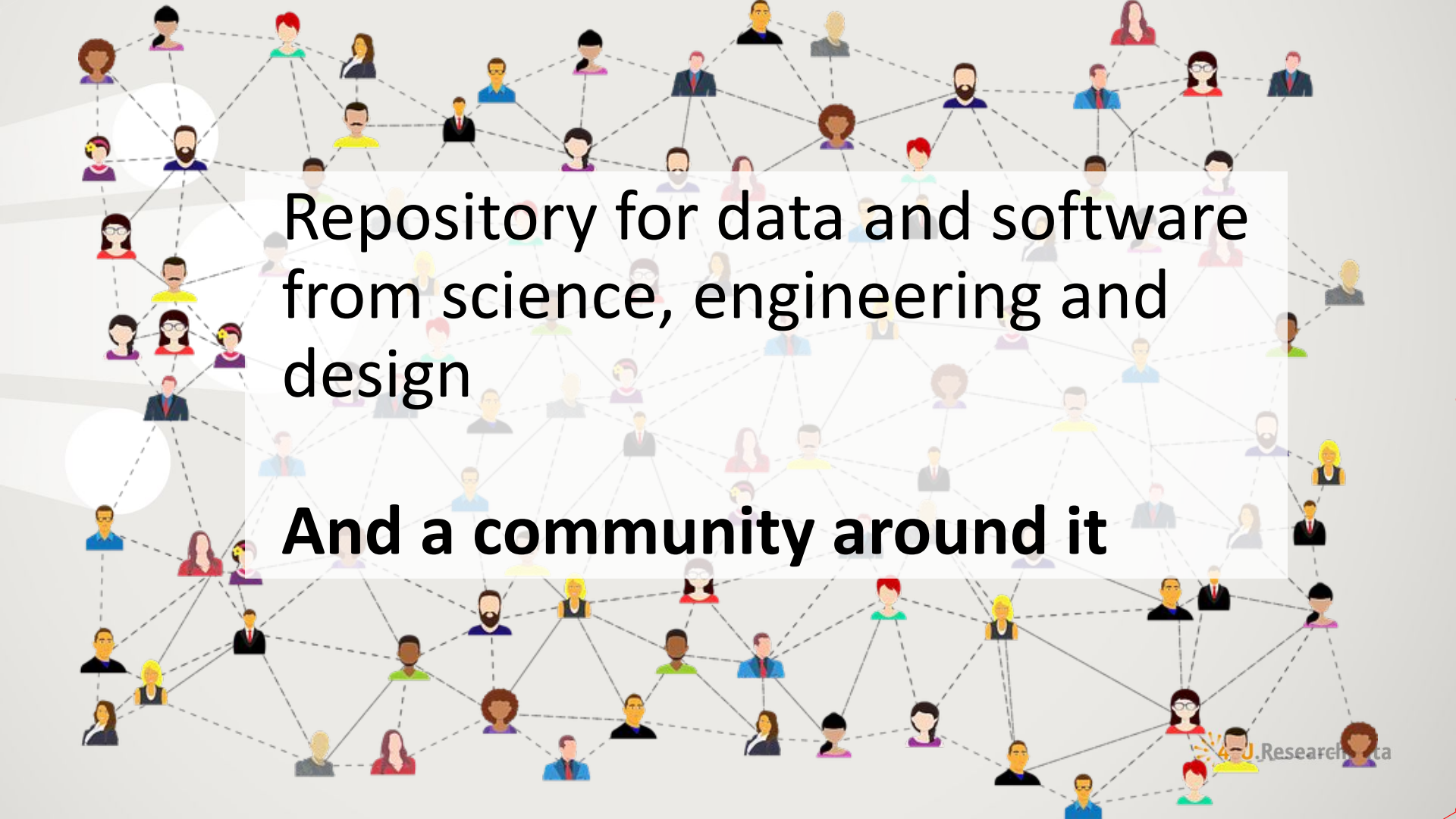
---

# Today

1. What is 4TU.ResearchData
2. Technical infrastructure: towards building the next generation repository for data and software
3. Lessons learnt
4. Discussion



# What is 4TU.ResearchData?



Repository for data and software  
from science, engineering and  
design

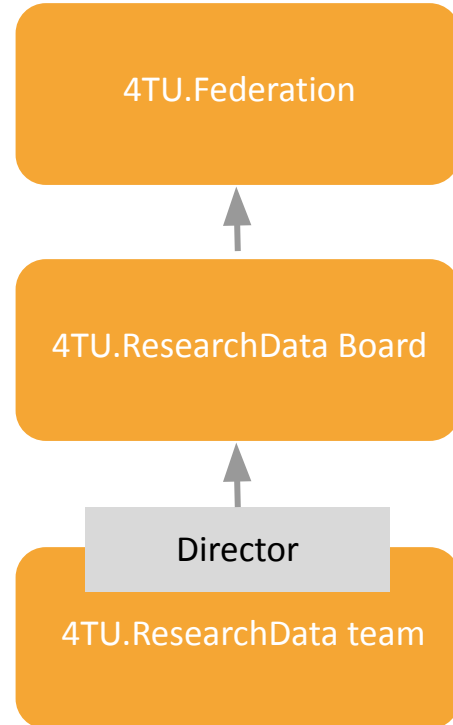
**And a community around it**

## Partnership of four Dutch technical universities



*Open to any researcher in the world!*

## Governance



# 4TU.ResearchData services

- ◇ **Repository - to publish research data and software**
  - DOIs for every datasets, long-term preservation and integrity check, versioning, metadata review, statistics, restricted access, software publication, specialist services for disciplinary formats and workflows
  
- ◇ **Training - to teach skills needed for reproducible working with data**
  - Essentials for Data Support - key training for aspiring data stewards - as part of RDNL
  - Hands-on workshop for computational skills for reproducible working with data (data and software carpentry)
  
- ◇ **Consultancy**
  
- ◇ **4TU.ResearchData Community - network of FAIR Data and Software leaders exchanging knowledge and practice**
  - Working groups
  - Researcher Stories
  - Events and workshops
  - FAIR Data Fund
  - Fellowship programme
  - *Recognising and rewarding FAIR Data*



# **Technical infrastructure: building the next generation repository**

# In 2020, 4TU.ResearchData started using Figshare as its institutional repository

AUGUST 18, 2020

## Why figshare? Choosing a new technical infrastructure for 4TU.ResearchData

<https://openworking.wordpress.com/2020/08/18/why-figshare-choosing-a-new-technical-infrastructure-for-4tu-researchdata/>



## Evaluation of Figshare

- ◇ Figshare is a black box for us
- ◇ Research community cannot contribute to development
- ◇ We have no insight in the black box and its development
  - Figshare determines the roadmap
- ◇ We maintain the wrapping paper
  - House style
  - Tweaking/adding functionality etc
- ◇ Maintenance is by definition reactive





**But what do we want?**

---

## Strategic principles for the 4TU.ResearchData repository

1. Open Source
2. Community engagement and community-driven innovation
3. Discipline-specific
4. Interoperable and aligned with national and international developments

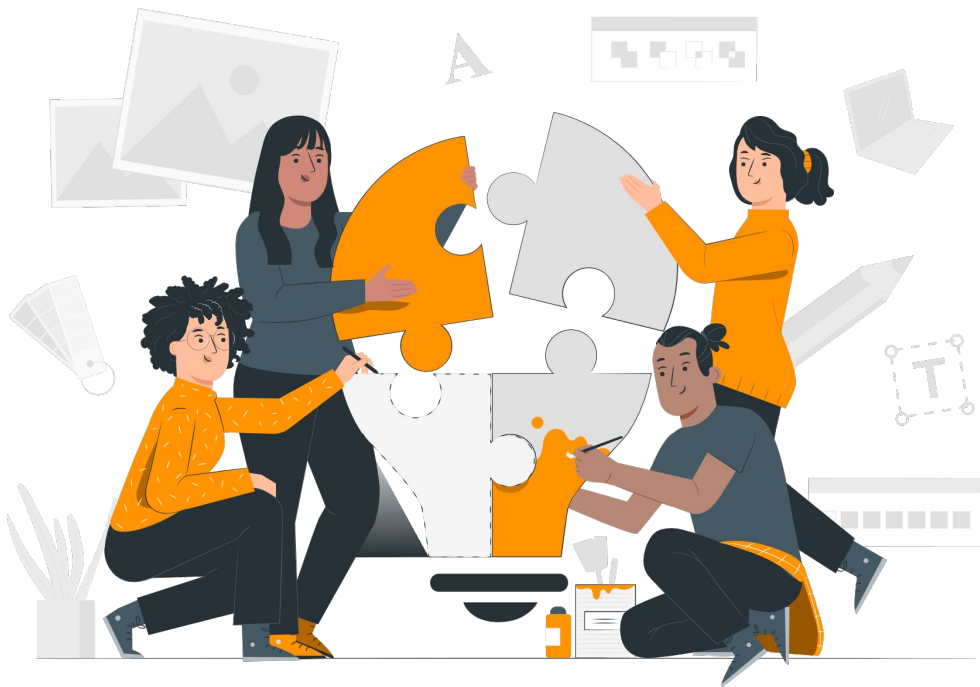
## Community is centre-stage



Partnership with the research community facilitates **sustainability** and an infrastructure which is **agile** and able to respond to **changing needs**.

**Together we develop what's needed**

## We need discipline-specific solutions which integrate with researchers' workflows



- Unique selling point: 50% of all data files are NetCDF
- Services based on *specific* file formats
- Discipline-specific metadata standards

# Community engagement and community-driven innovation



## JupyterFAIR project

Budget: 50k EURO

Goal:

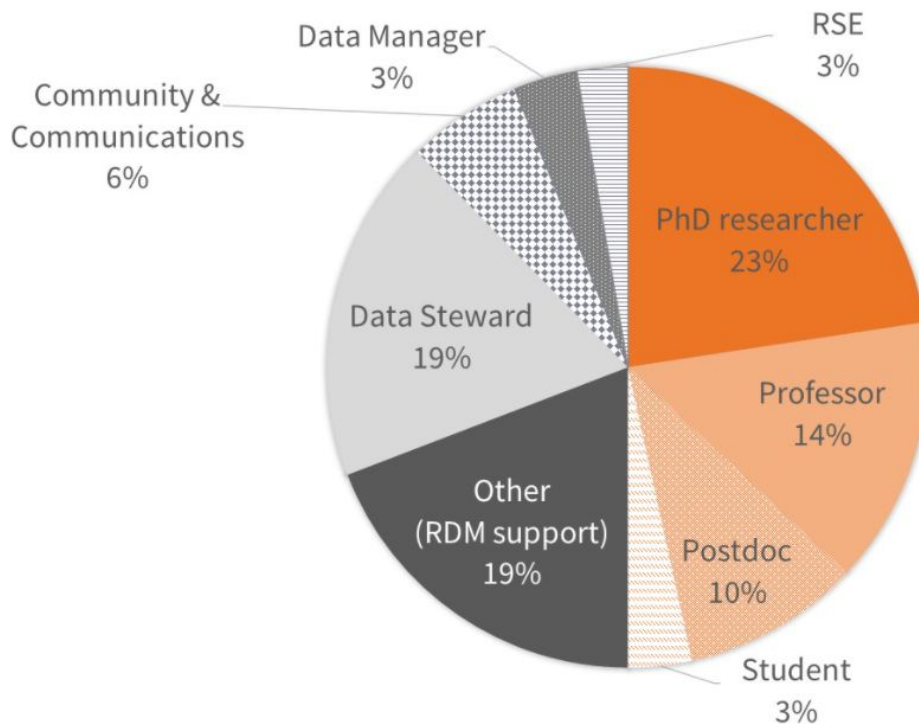
Connect research environments with 4TU.ResearchData to allow researchers easily re-use and deposit their data and metadata



Professor Serkan Girgin,  
University of Twente

<https://community.data.4tu.nl/2021/11/04/jupyterfair-integrating-research-environments-with-data-repositories-for-fair-research-data-management/>

# Community engagement and community-driven innovation



*Type of members of  
4TU.ResearchData  
community*

Image credit:  
Connie Clare

## The importance of open source in the community's own words



Emmy Tsang, ex-Open  
Science Community  
Manager at TU Delft


*'...our researchers and libraries are left in thrall to the power wielded by commercial companies. These **companies are profit-driven**, meaning not only do they not prioritise serving the marginalised, **we, as users, have little oversight, control or negotiation power over how their products develop.**'*

[Rethinking the Scaffolds of Open](#), Open Working Blog



## Open source facilitates digital sovereignty: right and ability to control own digital data





**First, we thought of re-using  
an existing open source  
software**

## Evaluation of open source solutions

- ◇ 11 existing open source repository solutions were identified and investigated
- ◇ Based on evaluation from various perspectives, 4 were shortlisted
- ◇ Based on further in depth investigation, 2 were selected for functional testing
- ◇ No solution passed the test

None of the 11 investigated solutions are fit for purpose because they miss essential features and require much development work and maintenance.

## Parallel to the exploration of existing of open source solutions

- ◇ We were building our own database to enrich and understand the statistics
- ◇ We were preparing continuation of API services (backward compatible) that is a requirement anyway
- ◇ Figshare recommended to build our own user interface as solution for our problems with reactive maintenance

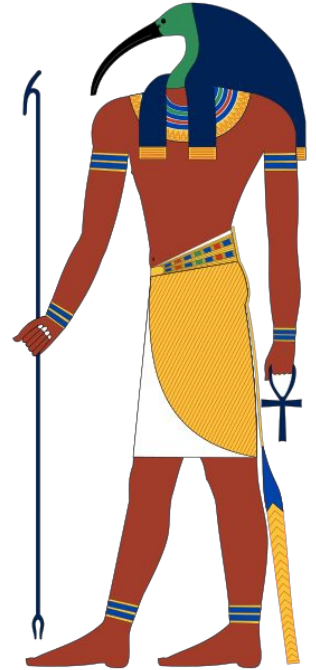
→ The work initiated to deal with these issues quickly evolved to be a possible replacement for Figshare

---

## Djehuty

Thoth (Djehuty) became credited by the ancient Egyptians as the **inventor of writing**

- Littleton, C. Scott (2002).  
*Mythology. The illustrated anthology of world myth & storytelling.*  
London: Duncan Baird Publishers.



## Where we are now...

1. Open source since summer 2022:  
<https://github.com/4TUResearchData/djehuty>
2. Test version available (and tested): <https://next.data.4tu.nl/portal>
3. Already 7 contributors to the code (the technical team counts 4 members)
4. Partnership with a teacher from the design faculty to improve software design
5. All in a little bit over 1 year
6. Going live date: 4th of March 2023



# Lessons learnt

# 1. Don't be afraid of trying out new things



*“Succes  
of een les”*

Irene Haslinger

TU Delft Library Director





## 2. An aligned and motivated team, with the community around it can achieve a lot!



Diversity and inclusion  
is essential

### 3. Involve your local IT colleagues early in the process

- ◇ The principle 'procure first' in conflict with the strategic direction of 4TU.ResearchData
- ◇ Different view on working with the research community
- ◇ Ways of working, communication styles need alignment
  
- ◇ What can we learn from one another?
- ◇ How can we adapt our cultures and ways of working to effectively collaborate?
- ◇ Could SURF/Géant play a role?

## Thank you - the team who led this work



Roel Janssen  
Senior Developer



Madeleine de Smaele  
Repository Manager



Alessandra Soro  
Community Manager



Egbert Gramsbergen  
Data Engineer



Kees den Heijer  
ex-team lead



Connie Clare  
ex-Community  
Manager

## Thank you - all colleagues who supported this work

- ◇ Contributors outside of the technical team: Frederique Belliard, Arie Braat, Catherina Vaendel
- ◇ Community feedback: Giordano Lipari, Serkan Girgin, Ashley Cryan, Heather Andrews, Jose Urria, all data stewards from TU Delft, TU/e, the University of Twente, WUR
- ◇ The Board of 4TU.ResearchData: Hubert Krekels, Irene Haslinger, Alastair Dunning, Wiljan Puttenstein, Wendy Ginkel, Marjolein Dohmen-Janssen, Iza Witkowska (ex-Board member), Marjolein Drent (ex-Board member)
- ◇ The Management Team of the TU Delft Library: Irene Haslinger, Alastair Dunning, Karin Clavel, Ingrid Klok
- ◇ ICT colleagues from all partner institutions: Arie van Nistelrooij, Joost Bengsch, Peter van Boheemen, Willem Oosting, Hyeokjin Kwon, Andjela Tomic, Hendri Hondorp, Ingeborg Ahlers
- ◇ 4TU.ResearchData non-technical team: Jan van der Heul, Eric Rumondor, Iulia Popescu, Paula Martinez Lavanchy, Carlos Utrilla Guerrero, Deirdre Casella (ex-member)
- ◇ DCC team at TU Delft

And many, many others!



# Thank you!

Images: Storyset <https://storyset.com/>

