

# On a (cultural) journey towards FAIR data



Slides are available

<https://doi.org/10.5281/zenodo.1485701>

The Zenodo logo consists of the word "zenodo" in a white, lowercase, sans-serif font. The letters are bold and have a slightly rounded, friendly appearance. The logo is centered within a rectangular area that has a blue gradient, transitioning from a darker blue on the left to a lighter blue on the right.

# Some questions to you:

1. Who has never heard of journals requiring that data supporting publications is shared?
2. Who has never heard of funders requiring data management plans?
3. Who has never heard that institutions have policies for data management?

# That's a LOT of new requirements!



# How to best support researchers in good data management and sharing practices?

*Case study from TU Delft*

# Where is TU Delft?

## (Delft University of Technology)



# TU Delft - statistics

- Research intensive:
  - 11th top beneficiary of H2020 funding in 2017
- 8 Faculties
- 5,000 employees
  - including PhD students
- ~23,500 students (Bachelor & Master)

# How to best support researchers in good data management and sharing practices?

- The usual beginning:
  - Let's build some tools
  - Let's improve infrastructure
- Let's instead try to be a bit scientific and do some research....



# What are the main barriers to data sharing?

Invited Forum: Challenges in Making Data Available  
*Empirical Article*



## Data Sharing in Psychology: A Survey on Barriers and Preconditions



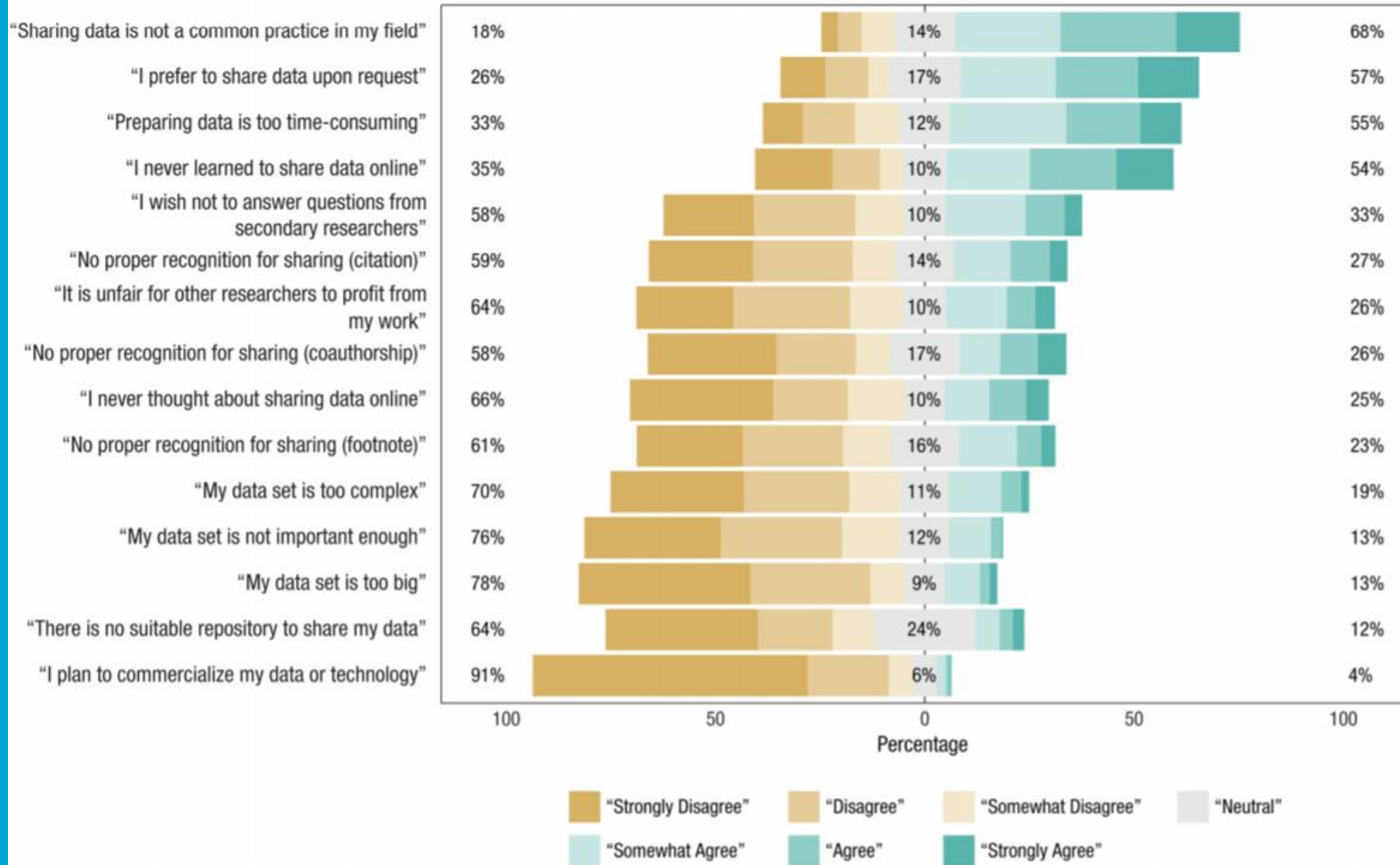
**Bobby Lee Houtkoop<sup>1</sup>, Chris Chambers<sup>2</sup>, Malcolm Macleod<sup>3</sup>,  
Dorothy V. M. Bishop<sup>4</sup>, Thomas E. Nichols<sup>5,6,7</sup>, and  
Eric-Jan Wagenmakers<sup>1</sup>**

<sup>1</sup>Psychological Methods Programme Group, University of Amsterdam; <sup>2</sup>Brain Research Imaging Centre (CUBRIC), School of Psychology, Cardiff University; <sup>3</sup>Centre for Clinical Brain Sciences, University of Edinburgh; <sup>4</sup>Department of Experimental Psychology, University of Oxford; <sup>5</sup>Oxford Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, Nuffield Department of Population Health, University of Oxford; <sup>6</sup>Wellcome Centre for Integrative Neuroimaging, FMRIB, Nuffield Department of Clinical Neurosciences, University of Oxford; and <sup>7</sup>Department of Statistics, University of Warwick

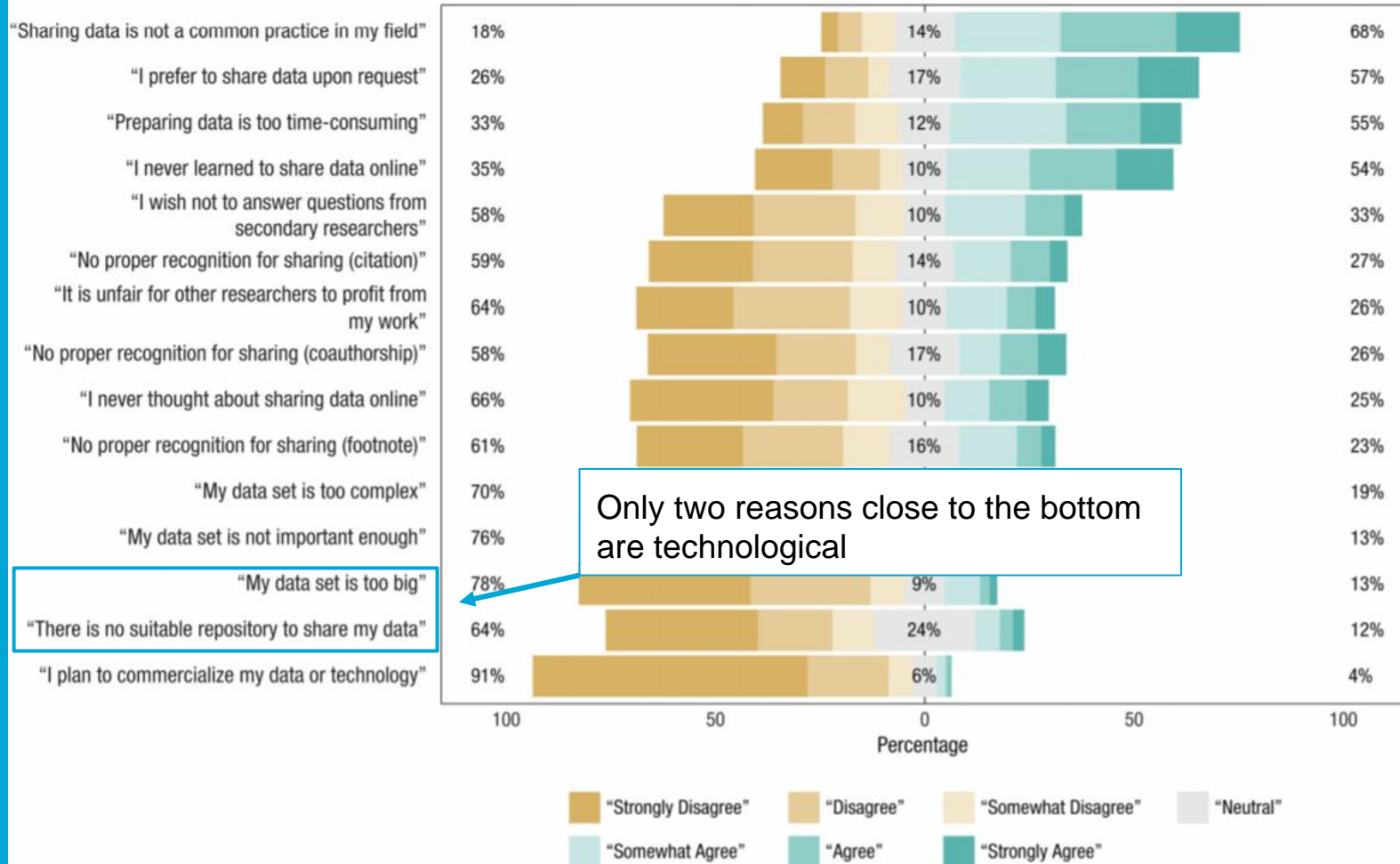
Advances in Methods and  
Practices in Psychological Science  
2018, Vol. 1(1) 70–85  
© The Author(s) 2018  
Reprints and permissions:  
sagepub.com/journalsPermissions.nav  
DOI: 10.1177/2515245917751886  
www.psychologicalscience.org/AMPPS



# To what extent do you agree with the following statements about barriers related to data sharing?



# To what extent do you agree with the following statements about barriers related to data sharing?



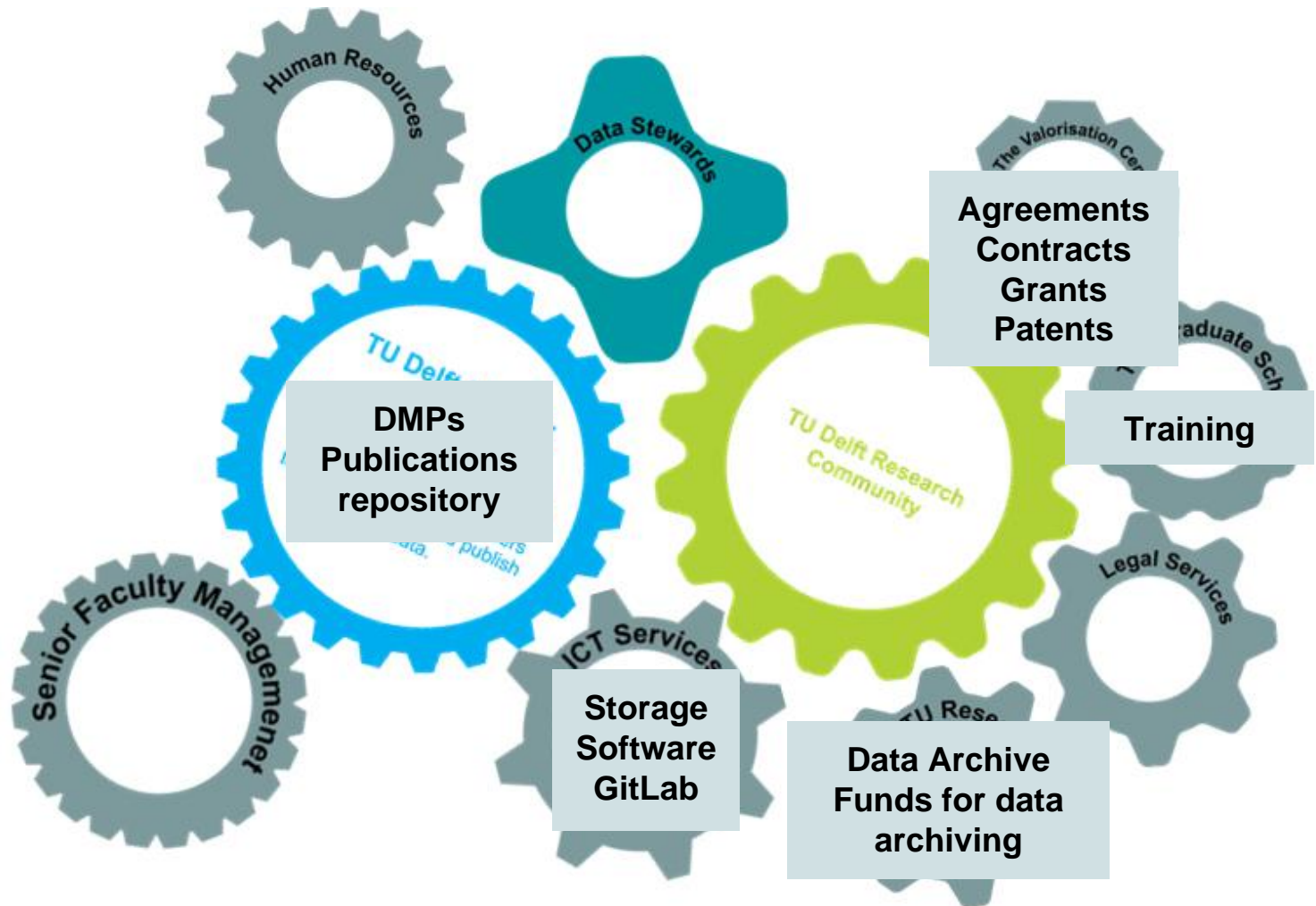
# Similar findings in other reports (from other disciplines), e.g.

- Life sciences, social sciences and humanities:
  - Van den Eynden et al. (2016)
  - <https://doi.org/10.6084/m9.figshare.4055448.v1>
- All disciplines:
  - Johnson et al. (2016)
  - <http://doi.org/10.5281/zenodo.177856>

What do we already have?  
A lot of services (and infrastructure support) already available



What do we already have?  
A lot of services (and infrastructure support) already available





# However... Some anonymous feedback from researchers

*“People don't tell us anything, we don't know the options, we just do it ourselves.”*

*“I think data management support, if it exists, is not well-known among the researchers.”*

*“I think I miss out on a lot of possibilities within the university that I have not heard of. There is too much sparsely distributed information available and one needs to search for highly specific terminology to find manuals.”*

<https://openworking.wordpress.com/2018/02/07/do-as-you-preach-results-of-2017-data-management-survey-now-published/>

# Summary

- The main obstacles to data management and sharing are **cultural, not technological**





# How to achieve cultural change?



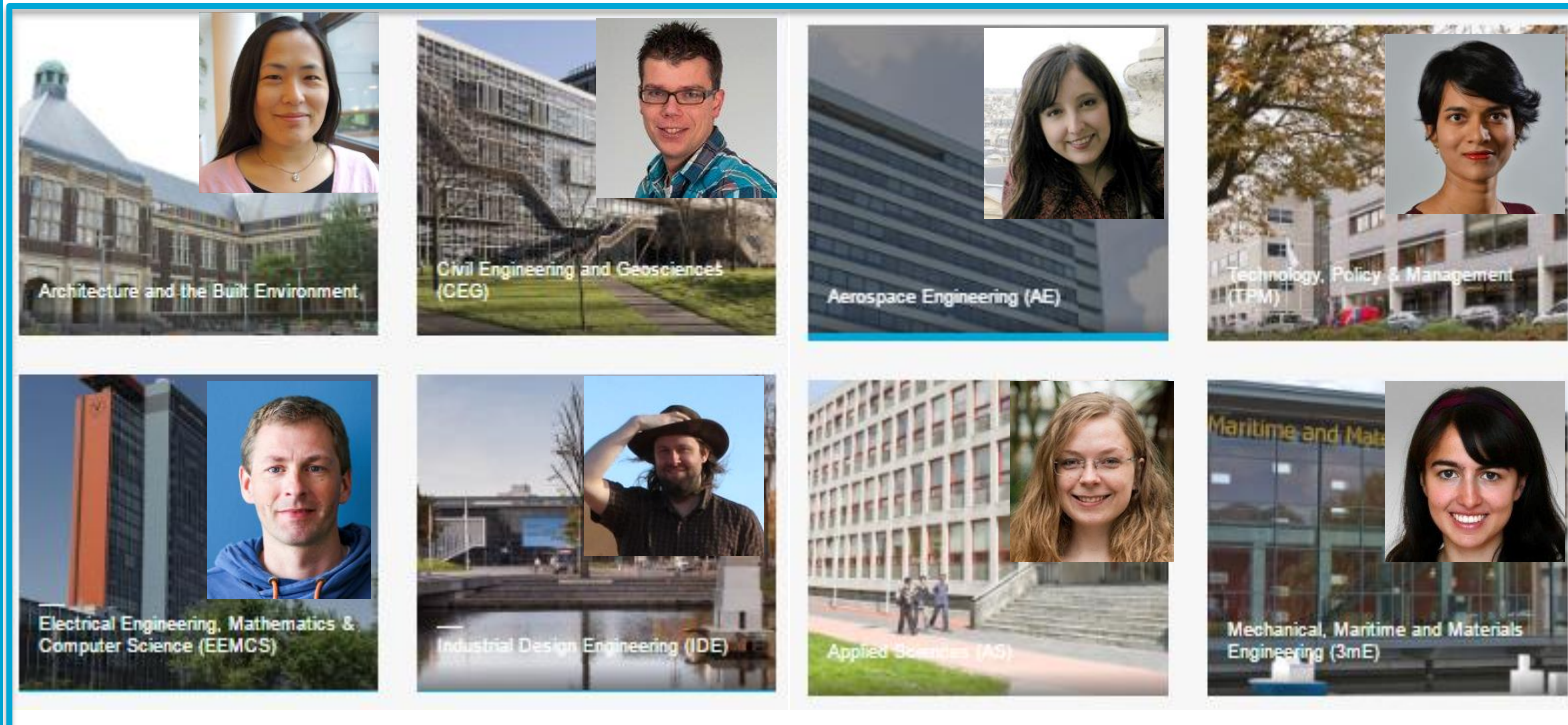


Researchers need support - someone to give them a helpful hand



# For support to be relevant, it needs to be discipline-specific

We appointed a dedicated Data Steward at every faculty: all have a PhD (or equivalent)



Training in data management provided:

<https://openworking.wordpress.com/2017/09/18/training-for-data-stewards/>

# Key principles behind the initiative

- Research is central
- Data stewards are consultants, not police!
  - Objective: improve culture (not compliance)
  - Focus on incremental changes, be realistic

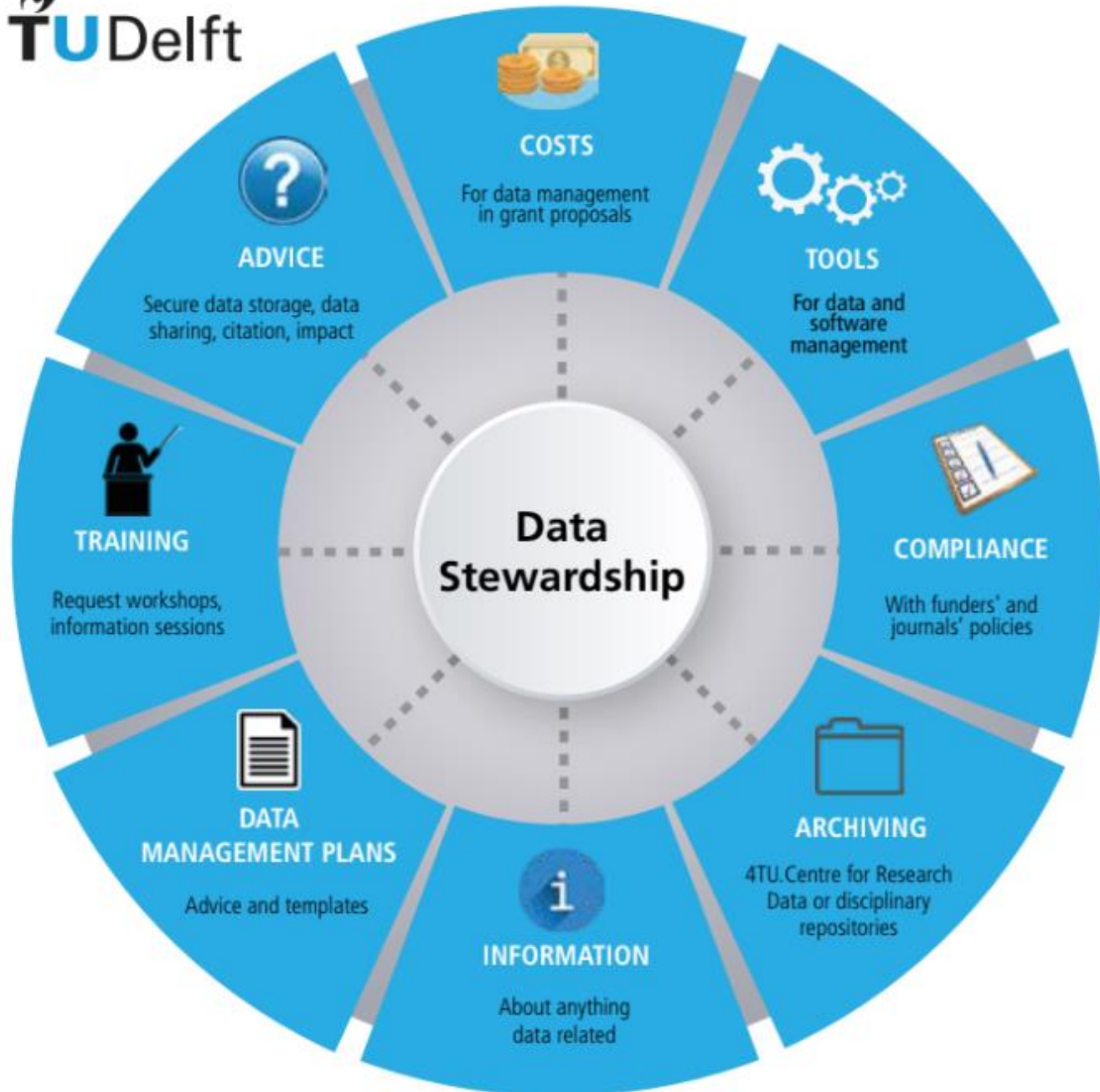




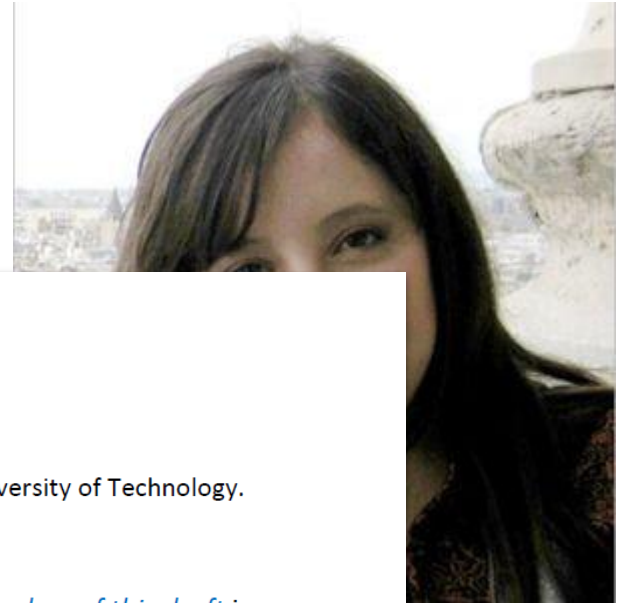
# What do the Data Stewards do?



- First contact points for any data questions
- Advocacy and training
  - *Reaching out to researchers who “don’t have data” and who “don’t have any problems”*



# Development of disciplinary protocols



## Core Requirements for Astronomical Imaging Data

Version September the 20<sup>th</sup>, 2018

Creator: Dr. Heather Andrews, Data Steward at Faculty of Aerospace Engineering at Delft University of Technology.

*Text to be filled in by the researcher* is depicted in red, while *extra information for readers of this draft* is described in italic blue font. This documents is for one type of data used by astronomers. An astronomical project would in most cases include more than imaging data only.

ANDREWS  
Faculty of  
Engineering

Core Requirement	Guiding questions
1. Data description and collection or re-use of existing data	Data Type: Astronomical imaging data from ground-based and/or space-based telescopes.  Data Format: Flexible Image Transport System format

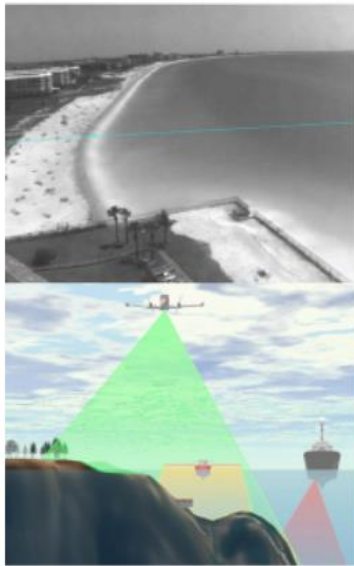
# Development of disciplinary protocols

## Procedure

The data collection procedure consists of four phases.

1. Extract. Collecting data and storing the measured data into files.
2. Transform. Enriching gathered data with metadata and storing in a standard file format.
3. Load. Storing the files in a database.
4. Provide. Giving access to the database.

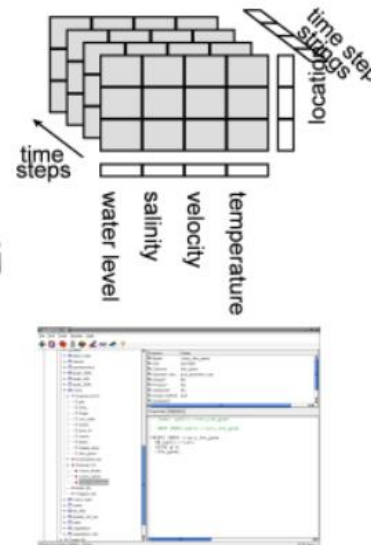
### Extract



### Transform



### Load



### Provide





# So has the cultural change happened?



# Visible change: Data Champions



[Yilin Huang](#)

*Assistant Professor*  
Faculty of Technology, Policy and Management  
Department: Multi-Actor Systems

**Areas of expertise:**  
Data analysis, modelling, simulation

- + Contact information
- + Motivation



[Frans van der Meer](#)

*Assistant Professor*  
Faculty of Civil Engineering and Geosciences  
Department: Materials, Mechanics, Management & Design

**Areas of expertise:**  
Computational mechanics, coding, simulations

- + Contact information
- + Motivation



[Anneke Zuiderwijk-van Eijk](#)

*Postdoctoral Researcher*  
Faculty of Technology, Policy and Management  
Department: Engineering Systems and Services

**Areas of expertise:**  
Open data, research data, data infrastructures, design, user service patterns, online education

- + Contact information
- + Motivation



[Thomas Abeel](#)

*Assistant Professor*  
Faculty of Electrical Engineering, Mathematics and Computer Science  
Department: Intelligent Systems

**Areas of expertise:**  
Bioinformatics, genomics, high performance computing, data integration

- + Contact information
- + Motivation



[Anton Akhmerov](#)

*Associate Professor*  
Faculty of Applied Sciences  
Department: Quantum Nanoscience

**Areas of expertise:**  
Quantum nanoscience, numerical simulations



[Scott Cunningham](#)

*Associate Professor*  
Faculty: Technology, Policy and Management  
Department: Multi-Actor Systems

**Areas of expertise:**

# Front runners: Data champions



## Zoltán Perkó

*Assistant Professor*

Faculty of Applied Sciences. Department of  
Science and Technology

### **Areas of expertise:**

Computational physics, coding (Python, Matlab, C),  
version control (Gitlab, smartgit), patient data  
management, project management

For increasing the reproducibility of research, best practices in data management, software development and collaboration are crucial.



## Frans van der Meer

*Assistant Professor*

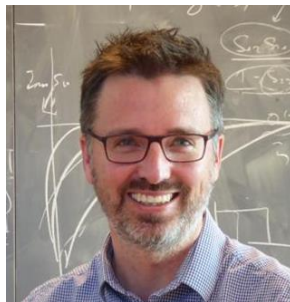
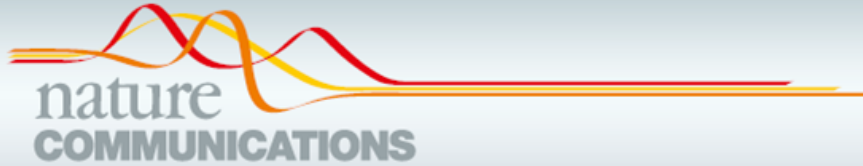
Faculty of Civil Engineering and Geosciences  
Department: Materials, Mechanics, and  
Design

### **Areas of expertise:**

Computational mechanics, coding, simulation

For the code to be re-usable, it is important that it is findable and documented. Data and code sharing outside the group is good for the visibility of our own research and for the scientific progress in general.

# Front Runners: Data Champions



ARTICLE

DOI: [10.1038/s41467-018-06595-2](https://doi.org/10.1038/s41467-018-06595-2)

OPEN

## A ballistic graphene superconducting microwave circuit

Felix E. Schmidt<sup>1</sup>, Mark D. Jenkins<sup>1</sup>, Kenji Watanabe<sup>2</sup>, Takashi Taniguchi<sup>2</sup> & Gary A. Steele<sup>1</sup>

### Data availability

All raw and processed data as well as supporting code for processing and figure generation is available in Zenodo with the identifiers <https://doi.org/10.5281/zenodo.1296129><sup>43</sup> and <https://doi.org/10.5281/zenodo.1408933><sup>44</sup>.



# What do the Data Champions do?



[Anton Akhmerov](#)

## Proposal for QN guidelines on data publication

*Preparing QN for open data*

### The proposal

We identify two possible “levels” of publishing data:

- Level 0: Uploading of the numerical data shown in figures in a format that is readable by others.
- Level 1: Publication of the raw data and scripts for the full data processing chain that produce the final plots shown in the paper.

Here, we propose that all QN PIs commit to publishing the data accompanying every paper at least at level 0, simultaneously with the paper being published in a journal. Additionally, we offer the following step-by-step guide to publishing research data, and offer our personal help in advising how to prepare your data for publication.

### How to prepare data for Level 0 data publication

To publish data compliant with “level 0”, you should first have the data extracted in a format that is readable to others

# Summary - quote

*Perhaps we should focus not so much on “do we have the right infrastructure” but on “do we have the right people?”*



Alastair Dunning  
Head of Research Data Services at  
TU Delft

And by the way...



<http://bit.ly/datastewardjob>

# Time for questions



Project updates: <https://openworking.wordpress.com/data-stewardship/>  
Slides: <https://doi.org/10.5281/zenodo.1485701>