Becoming a better scientist with open and reproducible research and... ...supporting the scientist on this journey.

Prof Laurent Gatto
Computational Biology and Bioinformatics Unit,
de Duve Institute, UCLouvain, Belgium

Dr Marta Teperek
Head of Research Data Services at TU Delft Library
Director of 4TU.ResearchData, The Netherlands

LG

Or… the importance of collaboration between researchers and support staff for sustainable open and reproducible research practices.

Prof Laurent Gatto
Computational Biology and Bioinformatics Unit,
de Duve Institute, UCLouvain, Belgium

Dr Marta Teperek
Head of Research Data Services at TU Delft Library
Director of 4TU.ResearchData, The Netherlands

LG

Useful links

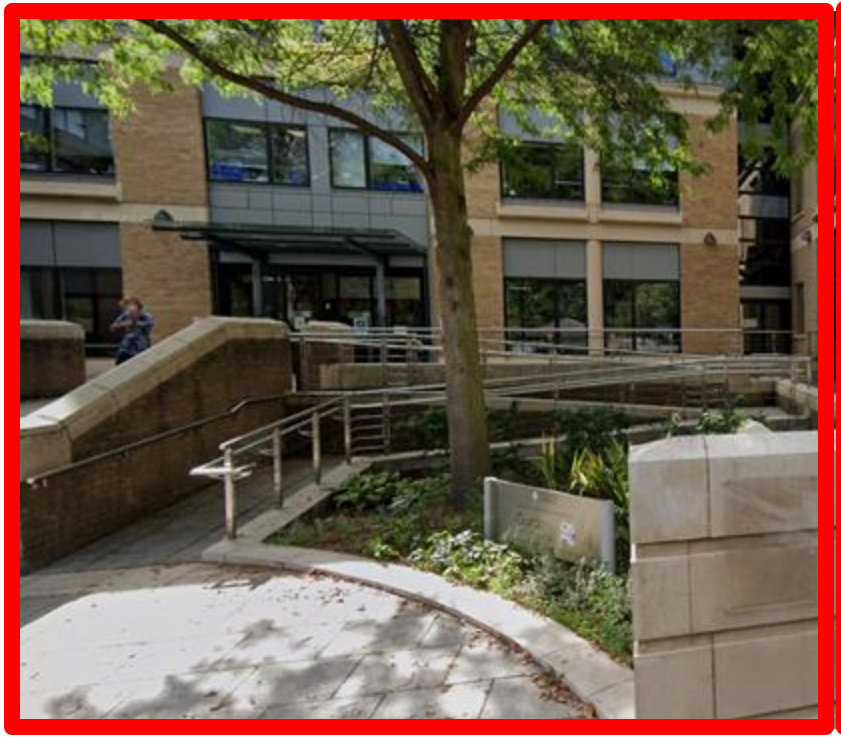
- **Slides**: https://bit.ly/202110RDM
- **Blog post**: https://lgatto.github.io/rdmhk/
- **Interactive questions** at https://www.wooclap.com/NAXHAK

Why a joint seminar?

Collaboration between researchers and support staff is essential for successful implementation of more and reproducible research practices.

By doing this talk jointly we would like to give you some concrete examples of these synergies and we hope to appeal to both researchers and support staff in the audience.

How did it all begin? From 2010 in Cambridge to 2021 in Hong Kong (virtually)

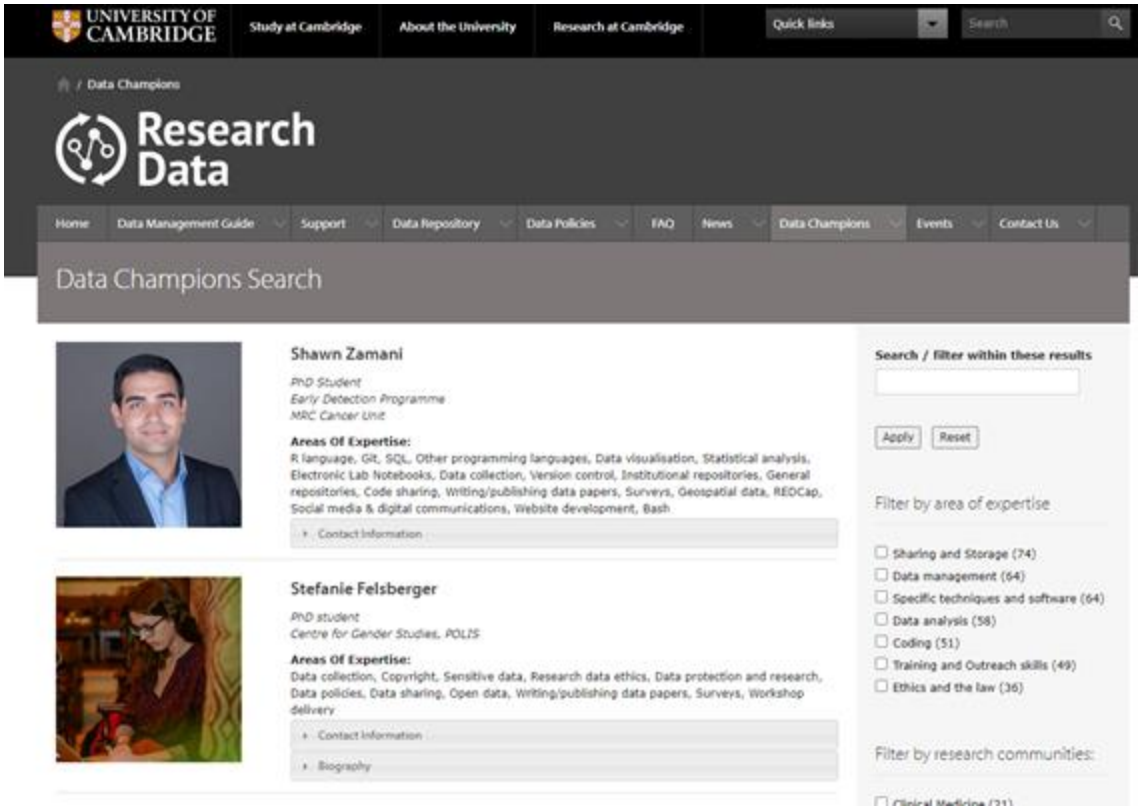


Marta, PhD at Gurdon Institute

Laurent, postdoc at Biochemistry Dept

MT

Data Champions initiative



2016: Laurent became a Data Champion and Marta was part of the Research Data Support team

https://www.data.cam.ac.uk/data-champions-search

MT

On the menu

- What is ‘Open’, ‘Reproducible’ and ‘Good’ science?
- People support
- Practice and policy
- Training
- Rewards
- Community
- Credit

What is ‘Open’, ‘Reproducible’ and ‘Good’ science?

- **Poll**: Do you practice open and/or reproducible research?
- **Wordcloud**: Ask for a couple of words describing how participants see open/reproducible research and create a word cloud?

Participate at https://www.wooclap.com/NAXHAK

Open science/research is the process of transparent dissemination and access to knowledge, that can be applied to various scientific practices: open data, open source, open access, …

Reproducible research: there exist several levels, of increasing difficulty, that describe the action of using external data/software/material/informations to attempt to observe the same or comparable results. Formally, we **repeat**, **reproduce**, **replicate** or **re-use** depending on how much of the original material we have access to.

None of these are binary

They are continuous, multidisciplinary, multidimensional.

LG

Open != reproducible

Open != good (by default)

Reproducible != good (by default)

However, open and/or reproducible research relies on/is supported by **good data management**!

Open and reproducible research, **supported by good data management**, lead to **trust, verification and guarantees**:

- Trust in Reporting - result is accurately reported
- Trust in Implementation - analysis code successfully implements chosen methods
- Statistical Trust - data and methods are (still) appropriate
- Scientific Trust - result convincingly supports claim(s) about underlying systems or truths

Which are all hallmarks of *good research*.

(see Gabriel Becker An Imperfect Guide to Imperfect Reproducibility, May Institute for Computational Proteomics, 2019.)

People support

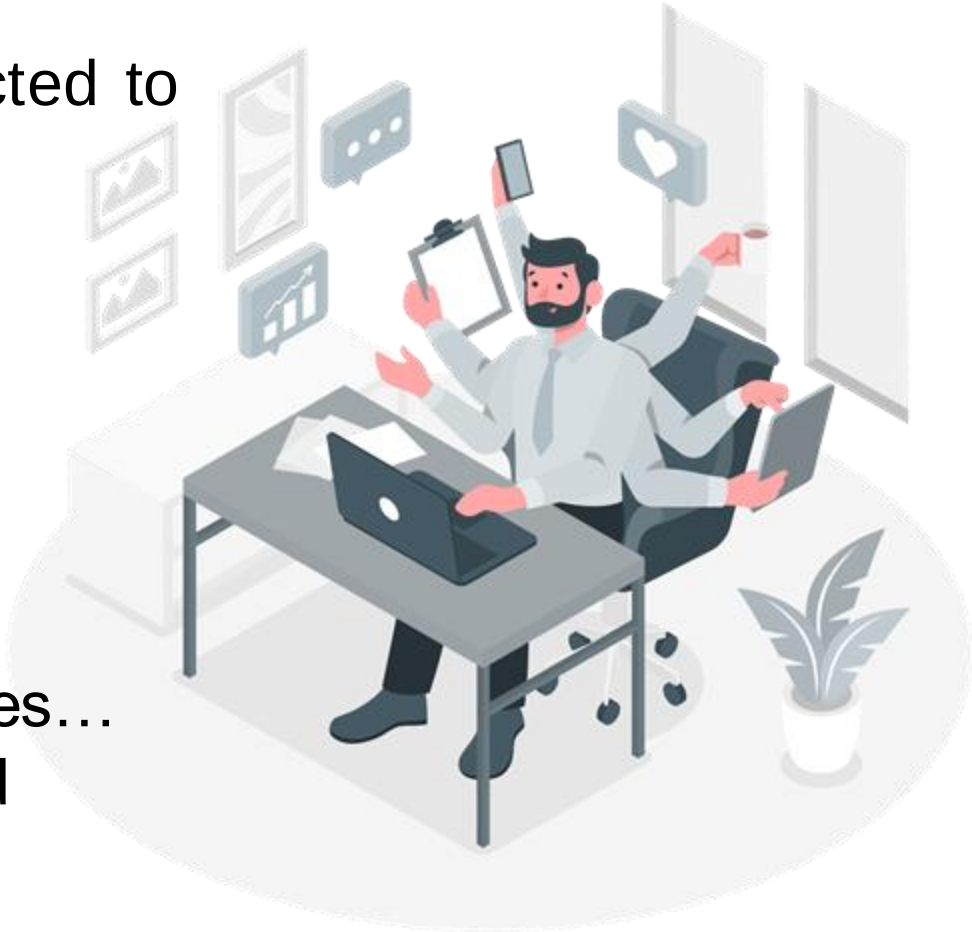
CBIO lab statement:

Open Science and Reproducible Research We are committed to the open, transparent and rigorous practice of scientific enquiry. In particular, we make every possible effort to make our research repeatable, reproducible and replicable, in the hope that it can be re-used and improved upon by as many as possible. Concomitantly, we release all our software and data under open permissible licences. Finally, we will also ensure that our research (such as, but not limited to journals articles, presentations, and book chapters) is published under open access licences to allow everybody to freely read, re-use and mine it.

Centering the lab around the **principles of good data management** to enable open and reproducible research.

Researchers are expected to excel at a lot of tasks:

- Research
- Teaching
- Supervising
- Getting funding
- Writing papers
- Reading papers
- Going to conferences…
- Managing code and data...



MT

Good data management takes time and effort, requires knowledge and skills.

Researchers should not be doing this alone.

MT

Data Stewards

- Trusted, first contact people for any data questions
- Advocacy and training
- Research is central
- Consultants, not police!

https://www.tudelft.nl/en/library/research-data-management/r/support/data-stewardship/contact

MT

Digital Competence Centre Support Team

Virtual team of 4 Research Software Engineers, 2 Data Managers, 1 HPC Specialist & 1 Coordinator:

- Hands-on support with data and software management
- Team members join research groups for up to 2 days per week for up to 6 months
- Teaching by co-creation and working together with research teams

https://www.tudelft.nl/en/library/library-for-researchers/library-for-researchers/setting-up-research/dcc

MT

Get in touch with your colleagues and ask for support https://library.ust.hk/about-us/contact-us/staff-directory/

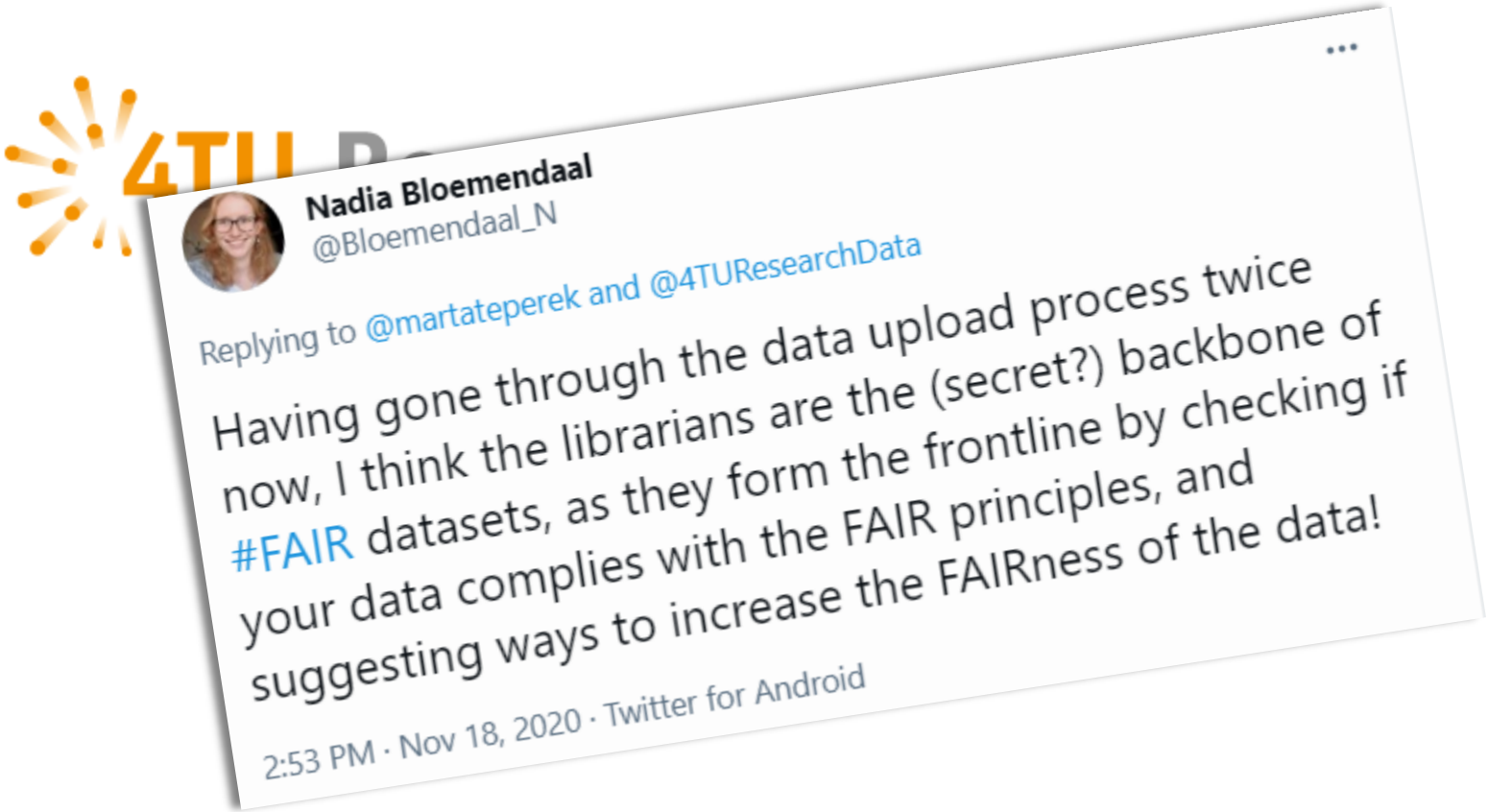
MT

Support for FAIR Data at TU Delft

- Since 2010
- Trusted data and software repository for science, engineering and design
- Open to researchers from all over the world

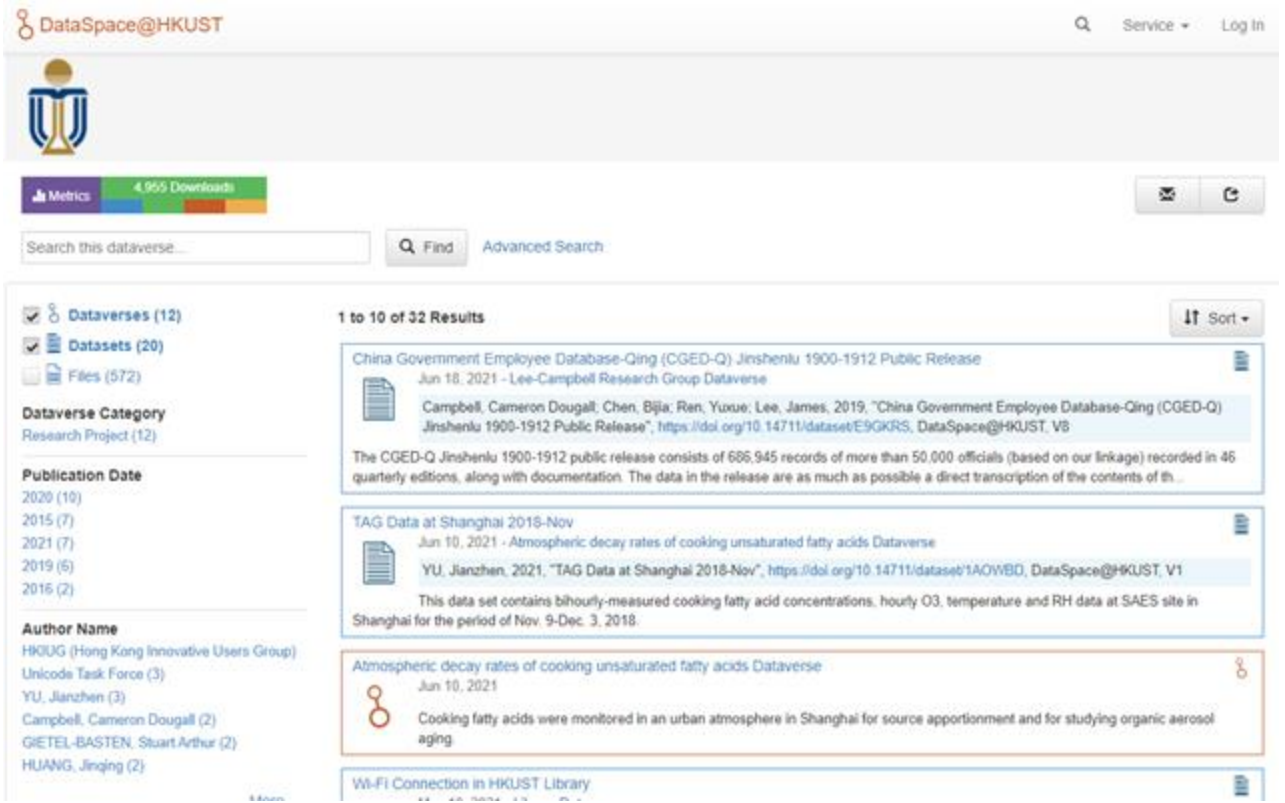
MT

The humans behind infrastructure make a difference!



MT

Make use of the services (and people) who can help you



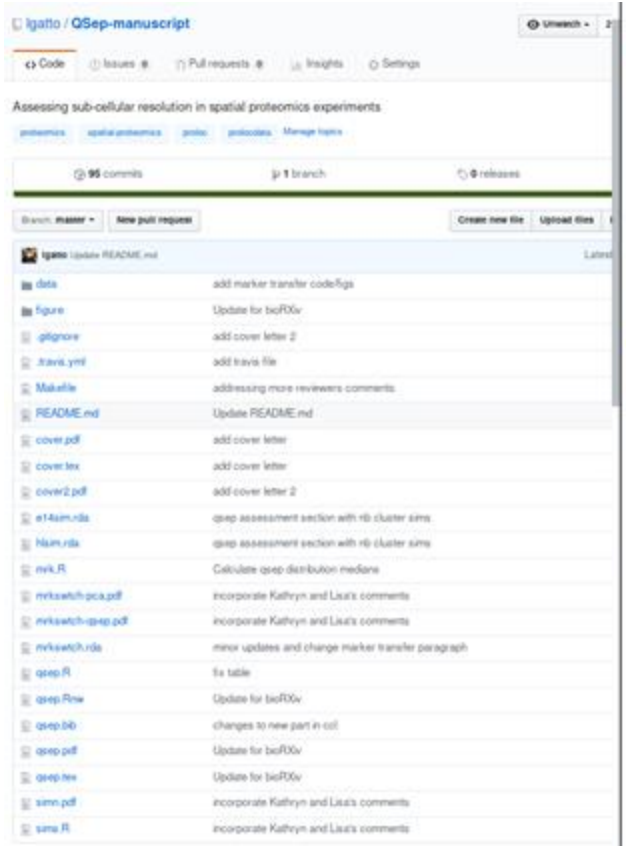
https://dataspace.ust.hk/

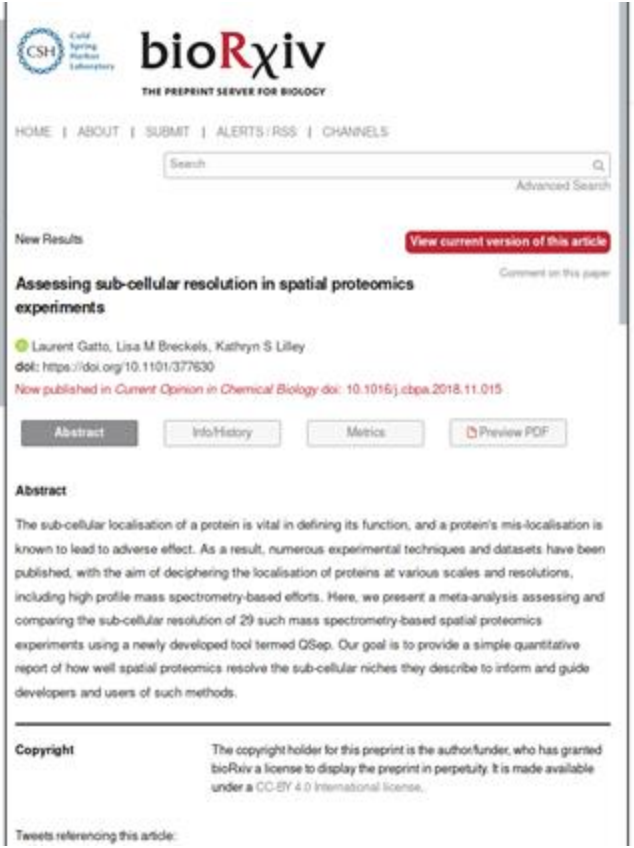
MT

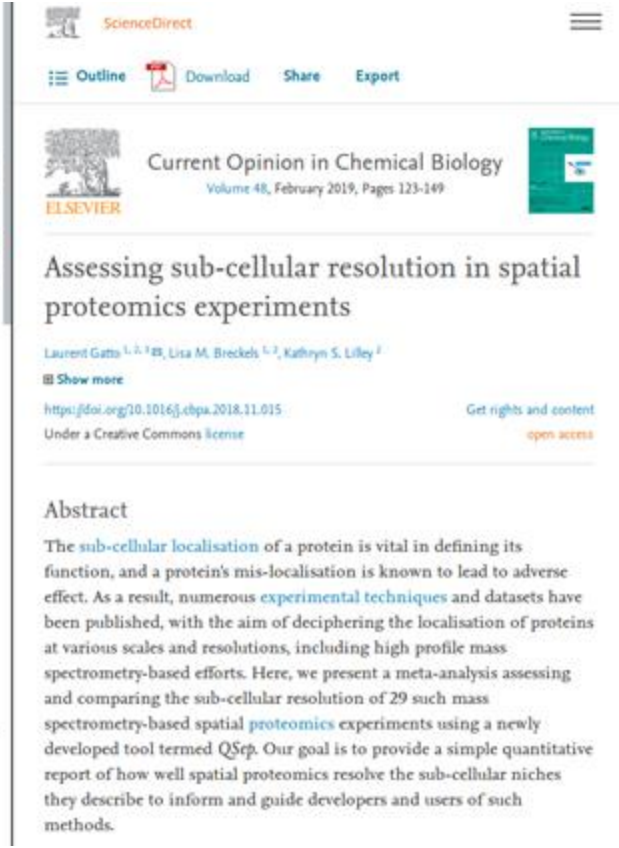
Practice and policy

- **Poll**: Does your institution have policies in place on data management?

Participate at https://www.wooclap.com/NAXHAK

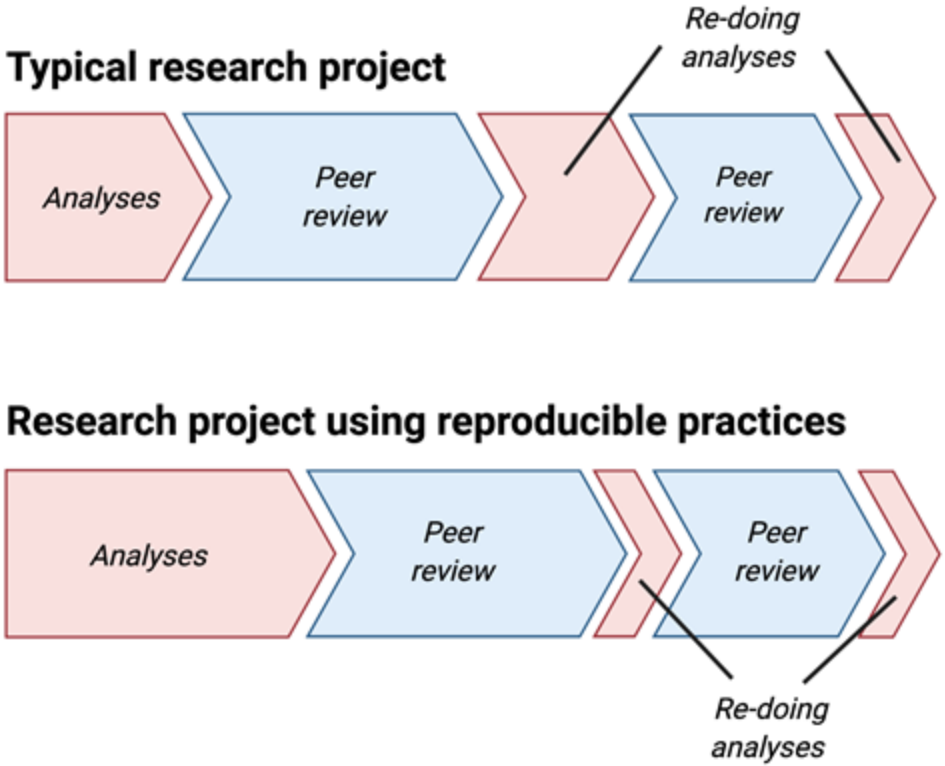






Practice. From left to right: reproducible document (produced with open source software and curated data), preprint and peer reviewer paper.

Does it take more time to work reproducibly?



From: [Five things about open and reproducible science that every early career researcher should know].

- Data management is driven by **intrinsic motivation** to produce open/reproducible/trustful research.
- Easy at a **local scale.**
- How can we act **globally**? … policies.

Policies disconnected from practice can turn into box-ticking exercises

Policies disconnected from practice can turn into box-ticking exercises

MT

Policies can be very useful when developed hand-in-hand with practice

Teamwork between researchers and support staff essential!

TU Delft experience with policy development:
https://datascience.codata.org/articles/10.5334/dsj-2019-045/

MT

TU Delft Research Data Framework Policy

"When PhD students leave, they usually write up papers, but their research data is not archived… **what if someone ask me questions about the reproducibility of their data**? I won't be able to check if they were right if the data is not there"

Researcher, PhD student supervisor

"I would love to share my code and data openly, but my **supervisor thinks it is a waste of time**…"

PhD student

TU Delft Research Data Framework Policy: https://doi.org/10.5281/zenodo.4088123

MT

TU Delft Research Data Framework Policy

1. PhD students need to write a Data Management Plan as part of their Go/No Go
2. PhD students need to upload data into data repository before they leave
3. Supervisors responsible for approving both

"PhD supervisors are coming to us to learn more about data management because their students are asking questions about a data management plan"

Data Steward

"I found documentation important and also the DMP, since my supervisor had never mentioned anything about the importance of it."

PhD student

MT

TU Delft Research Software Policy

Anton Akhmerov

If I was to follow TU Delft regulations on copyright, I would need to file an Invention Disclosure Form with every Git pull request!

=> Working group with researchers and staff

MT

Results (3 years later!)

If researchers want to publish their software openly, TU Delft transfers copyright to them

TU Delft Research Software Policy: https://doi.org/10.5281/zenodo.4629662

MT

Training

- **Poll**: have you had any dedicated training on data management and reproducible research?.

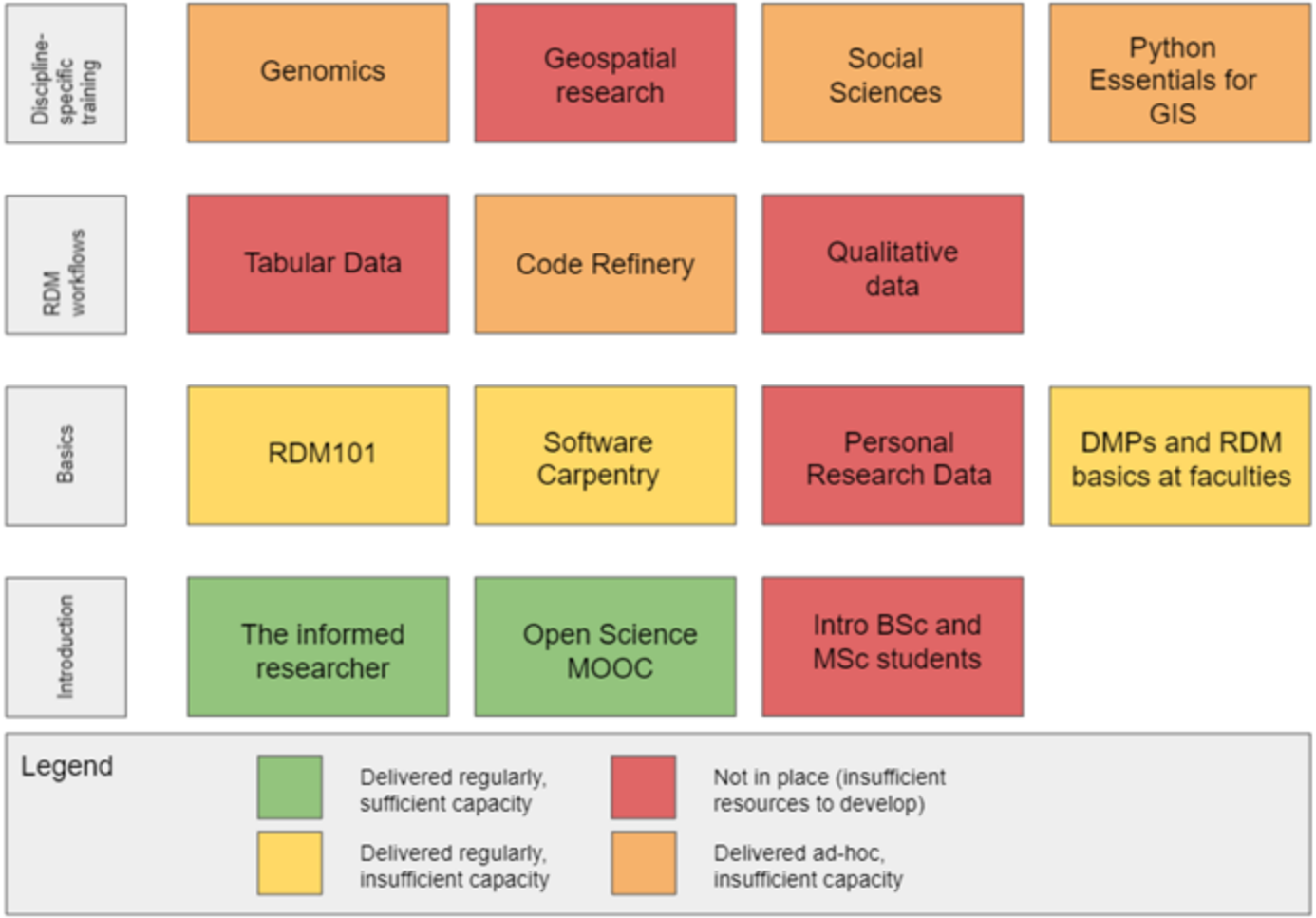
Participate at https://www.wooclap.com/NAXHAK

Laurent’s experience/path with learning and teaching the tools to appreciate the importance of good data analysis and the importance of good data management:

- Carpentries instructor (as a early career researcher).
- Workshops for graduate students and ECR.
- Re-use/incorporate this material (and experience) in my university teaching (bachelor and masters in biomedical sciences).

Teaching data management, open and reproducible research principles when running a lab: students that are well trained in RDM and RR are readily up to speed.

LG



TU Delft Library's Vision for data and software management training

MT

Carpentries - another example where collaboration between staff and researchers is essential

Genomic data carpentry at TU Delft

Victor Koppejan is a PhD student in computational Bioprocess Engineering. His doctoral research involves the purification of proteins from biological matrices for industrial use using expanded bed adsorption. Koppejan employs open source computer simulation tools to model the dynamics of fluid and particle flow in a fluidized bed using the Dutch National Super Computer.

Raúl Ortiz Merino is a Postdoctoral researcher in Industrial Microbiology. He works within the field of comparative genomics to characterise microbial gene sequences of commercial relevance. After years spent training as a wet lab biochemist, Ortiz Merino made his transition to dry lab computational science and is now an experienced bioinformatician.

Both Data Champions have made a significant contribution to their local research community by sharing their knowledge and expertise during Software and Data Carpentry workshops.

https://openworking.wordpress.com/2019/08/01/carve-your-niche-with-the-carpentries/

MT

Rewards

- **Wordcloud**: What would motivate you towards being more open/RR? More citations, better chances of getting hired, ….

Participate at https://www.wooclap.com/NAXHAK

LG

Rewards for researchers

Benefits for your academic career: How open science helps researchers succeed and more examples from the Open as a career boost paragraph:

- Open access articles get more citations.
- Open publications get more media coverage.
- Data availability is associated with citation benefit.
- Openly available software is more likely to be used. (I don't have any reference for this, and there are of course many counterexamples).
- Benefit from institutional support of open research practices.

Networking opportunities (this talk here today)

See also Why Open Research

- Increase your visibility: Build a name for yourself. Share your work and make it more visible.
- Take back control: Know your rights. Keep your rights. Decide how your work is used
- Publish where you want: Publish in the journal of your choice and archive an open copy. (See The cost of knowledge boycott of Elsevier).
- Get more funding: Meet funder requirements, and qualify for special funds such as the Wellcome Trust Open Research Fund.
- Get that promotion: Open research is increasingly recognised in promotion and tenure. See also Reproducibility and open science are starting to matter in tenure and promotion July 14th, 2017, Brian Nosek) and the EU's Evaluation of Research Careers fully acknowledging Open Science Practice defines an Open Science Career Assessment Matrix (OS-CAM).

And of course the **Five selfish reasons to work reproducibly**!

Change is here - San Francisco Declaration on Research Assessment (DORA)

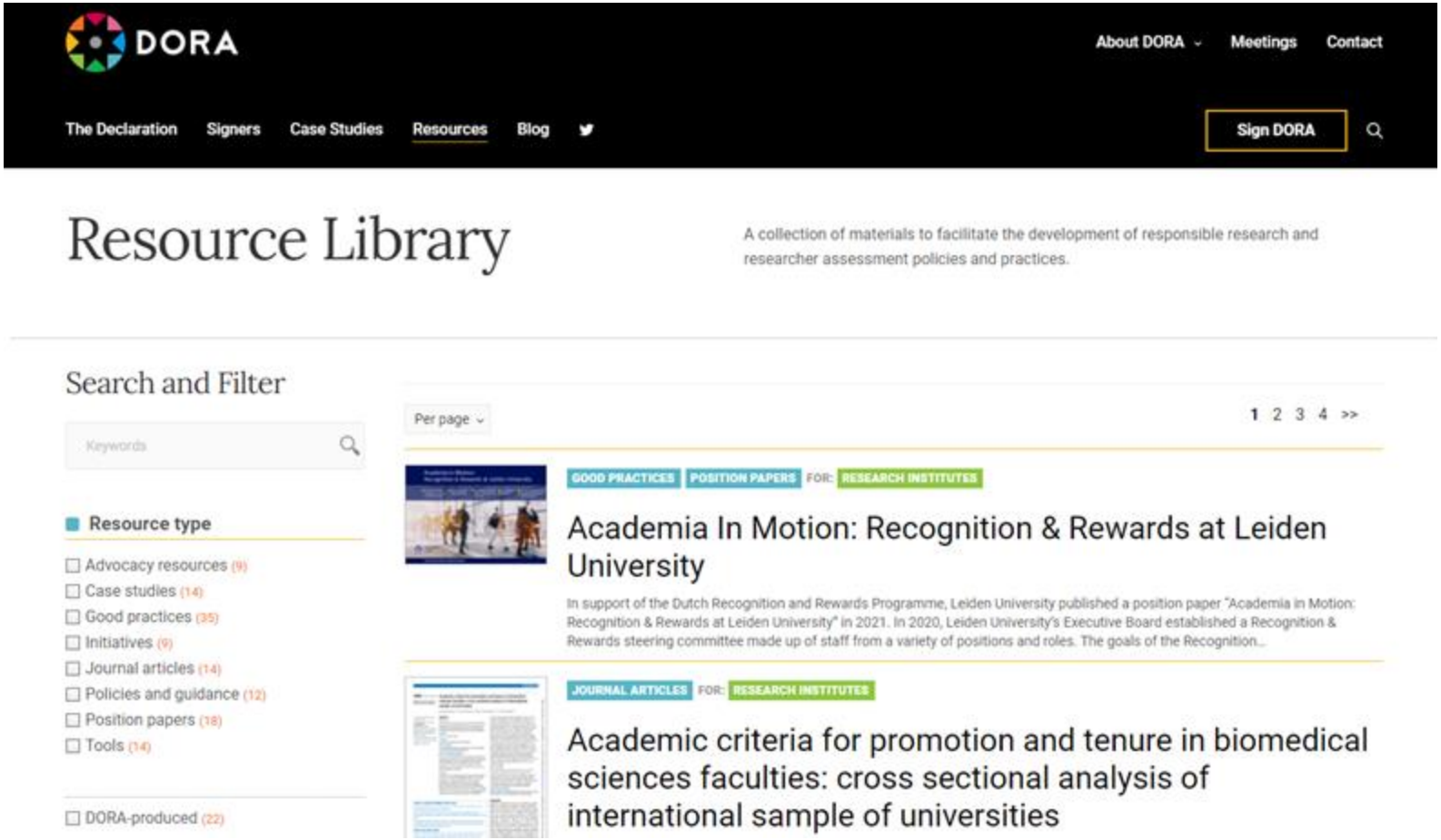
"A number of themes run through these recommendations:

- the need to **eliminate the use of journal-based metrics**, such as Journal Impact Factors, in funding, appointment, and promotion considerations;

- the need to **assess research on its own merits** rather than on the basis of the journal in which the research is published"

https://sfdora.org/

MT



https://sfdora.org/

MT

nature

Explore content ⌄ About the journal ⌄ Publish with us ⌄ Subscribe

nature › career news › article

CAREER NEWS | 04 August 2021

Dashboard will track hiring and promotion criteria

International coalition aims to identify how universities use impact factors and related metrics.

Chris Woolston

https://doi.org/10.1038/d41586-021-02145-x

MT

Funders are promoting change

Open Research Fund (Closed)

This funding supports researchers to develop and test incentives for making health research more open, accessible and reusable.

Career stage: Leading a research programme, Postdoctoral research, Returning to research

Level of funding: Up to £100,000

Duration of funding: Up to two years

https://wellcome.org/grant-funding/schemes/open-research-fund

MT

Funders are promoting change

Budget: 1M EURO

https://www.nwo.nl/en/calls/open-science-os-fund-2020/2021

MT

Funders are promoting change

https://sfdora.org/2019/11/14/quality-over-quantity-how-the-dutch-research-council-is-giving-researchers-the-opportunity-to-showcase-diverse-types-of-talent/

MT

But more importantly… it is all in your hands!

YOU can make a difference by being kind to each other and acknowledging each other's contributions

Most important changes are the ones done by individuals.

Everyone matters, everyone can contribute to making a difference.

MT

Community

There is

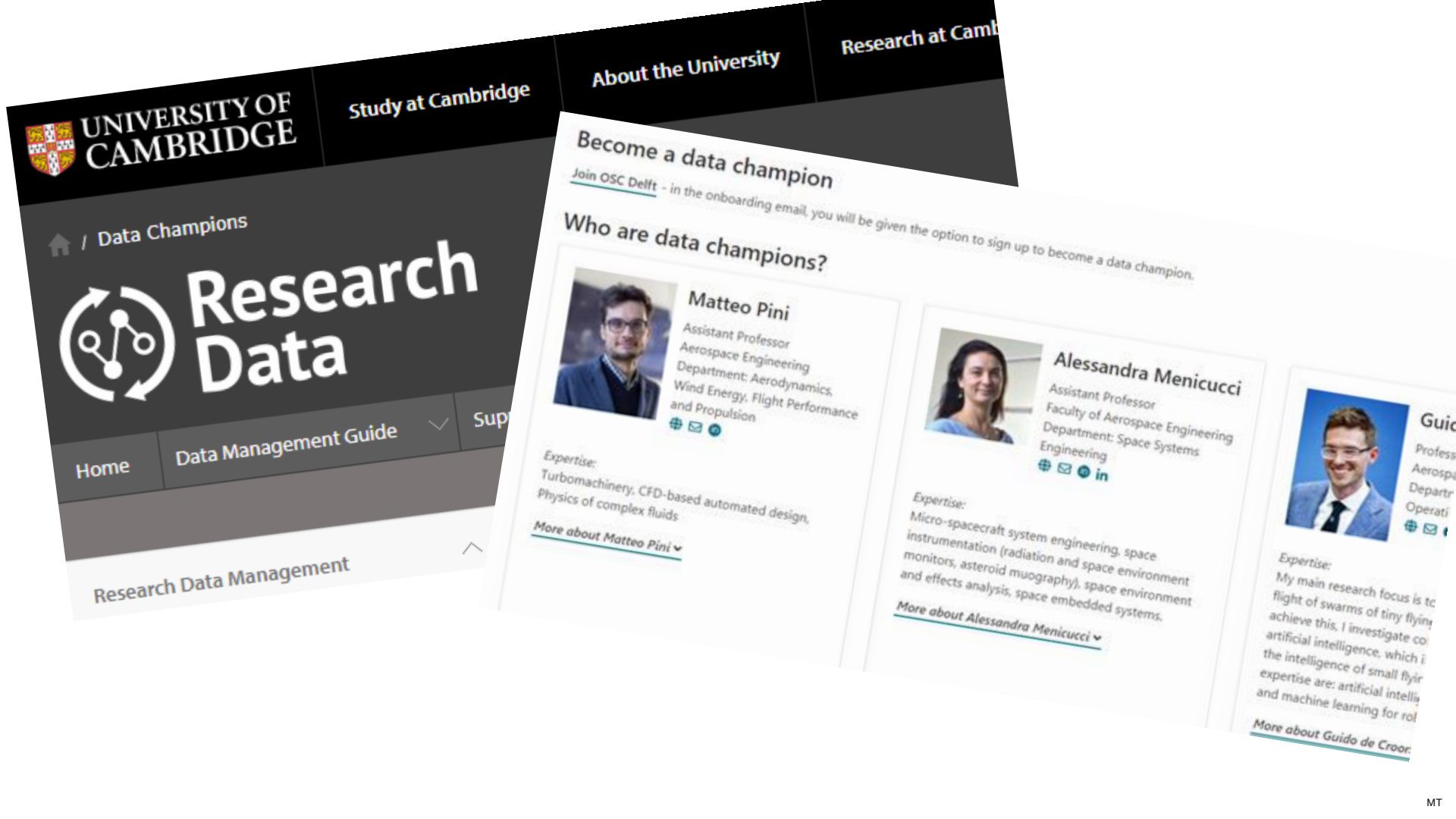
Open Science as in widely disseminated and openly accessible

and

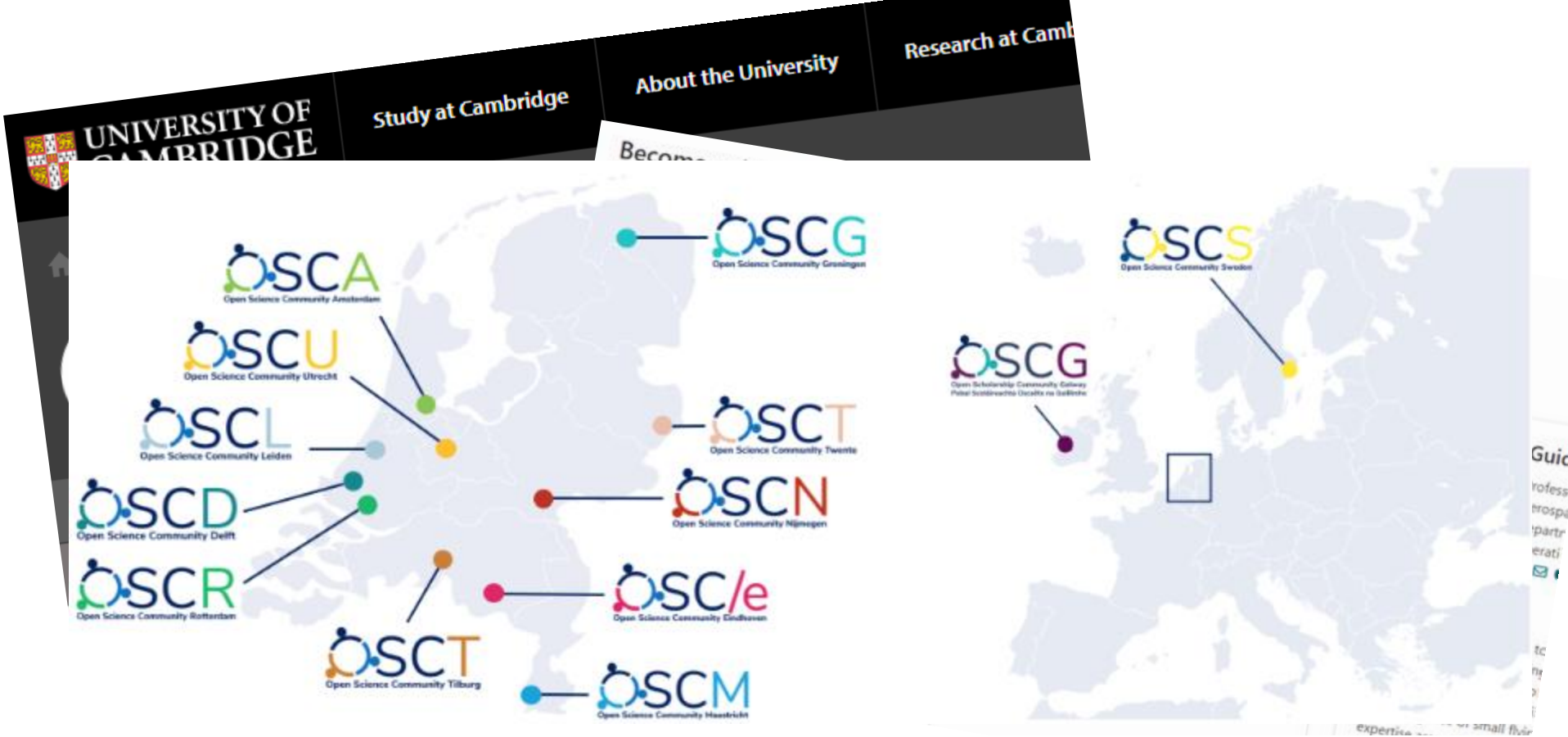
Open Science as in inclusive and welcoming

Citing Cameron Neyon:

LG


UNIVERSITY OF CAMBRIDGE
Study at Cambridge
About the University
Research at Camb
/ Data Champions
Research Data
Home
Data Management Guide
Sup
Research Data Management
Become a data champion
Join OSC Delft - in the onboarding email, you will be given the option to sign up to become a data champion.
Who are data champions?
Matteo Pini
Assistant Professor
Aerospace Engineering
Department: Aerodynamics, Wind Energy, Flight Performance and Propulsion
Expertise:
Turbomachinery, CFD-based automated design, Physics of complex fluids
More about Matteo Pini
Alessandra Menicucci
Assistant Professor
Faculty of Aerospace Engineering
Department: Space Systems Engineering
Expertise:
Micro-spacecraft system engineering, space instrumentation (radiation and space environment monitors, asteroid muography), space environment and effects analysis, space embedded systems.
More about Alessandra Menicucci
Guid
Profess
Aerospa
Departr
Operati
Expertise:
My main research focus is to
flight of swarms of tiny flyin
achieve this, I investigate co
artificial intelligence, which i
the intelligence of small flyir
expertise are: artificial intelli
and machine learning for rol
More about Guido de Croon

MT



The 13 open science communities worldwide. Image by Anita Eerland, licensed under CC BY-ND 3.0.

MT

INOSC Starter Kit

Open Science Community Starter Kit

Set up and foster a local Open Science Community

Get Started

The http://www.startyourosc.com/

By being true to yourself and to your values, you are part of a larger network, meet great people and gain long-lasting friendships

Open not only as in sharing, but as inclusive and welcoming. ♥

Illustration: Mark van Huystee, Digital Competence Centre showcase event, TU Delft, 12 October 2021

MT

Credit

Credits:

Laurent One of my advice when engaging in open and reproducible research is to make allies and friends. It isn't always easy, and support, whether technical or other, is always welcome. I have been lucky to meet wonderful allies and inspiring friends along the path towards open and reproducible research that works for me. Among these, I would like to highlight Corina Logan, Stephen Eglen, Marta Teperek, Kirstie Whitaker, Chris Hartgenink, Naomie Penfold, Yvonne Nobis.

Marta Would like to give credit to numerous colleagues from TU Delft and beyond who were the driving force behind all the work described in this post, and in particular:
Alastair Dunning, Anke Versteeg, Connie Clare, Data Stewards (Diana Popa, Esther Plomp, Heather Andrews, Jasper van Dijk, Jeff Love, Kees den Heijer, Nicolas Dintzner, Robert Eggermont, Santosh Ilamparuthi, Shalini Kurapati, Yan Wang, Yasemin Turkyilmaz-van der Velden), Digital Competence Centre Support Team (Amir Fard, Ashley Cryan, Jose Urra, Julie Beardsell, Manuel Garcia, Mark Schenk, Maurits Kok, Meta Keijzer-de Ruijter, Niket Agrawal, Susan Branchett), Eirini Zormpa, Emmy Tsang, Irene Haslinger, Karel Luyben, Maria Cruz, Paula Martinez Lavanchy.

Images: Pixabay and Storyset (https://storyset.com/)

MT