

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

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

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)

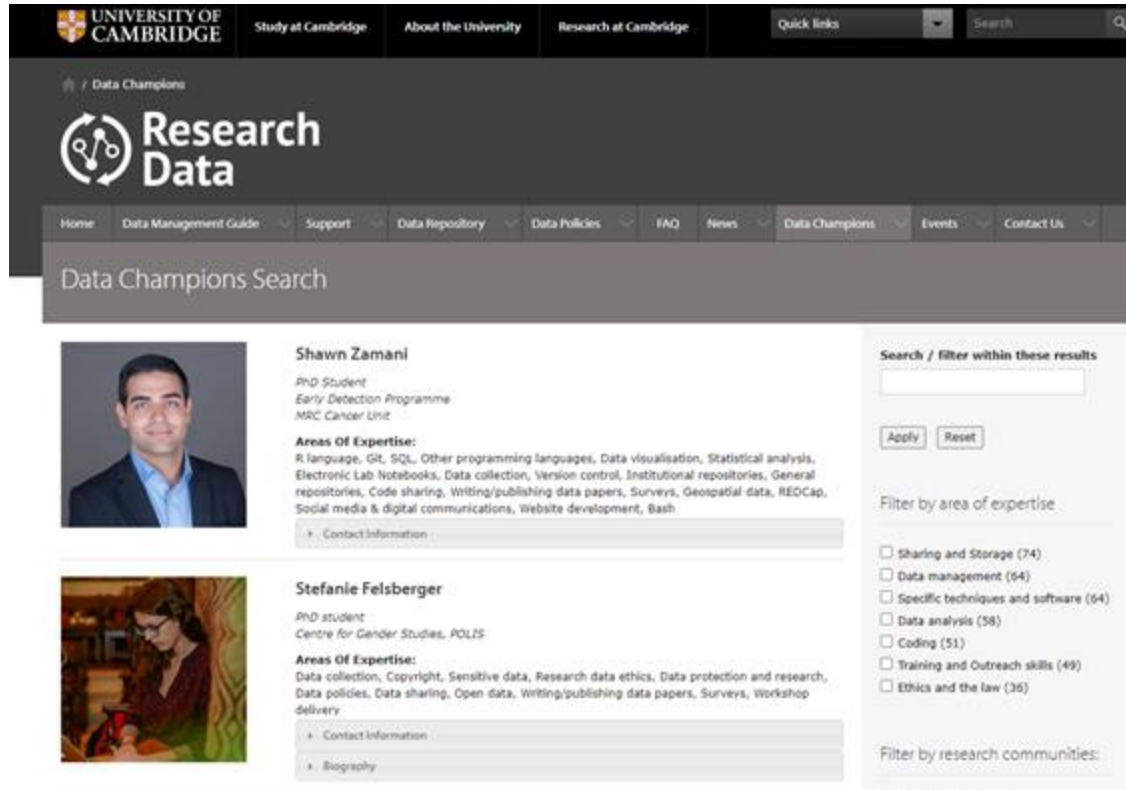


Marta, PhD at Gurdon Institute



Laurent, postdoc at Biochemistry Dept

Data Champions initiative



The screenshot shows the University of Cambridge Research Data website. The header includes the University of Cambridge logo and navigation links: "Study at Cambridge", "About the University", and "Research at Cambridge". There is a "Quick links" dropdown and a search bar. Below the header, the "Data Champions" section is highlighted. A navigation menu includes "Home", "Data Management Guide", "Support", "Data Repository", "Data Policies", "FAQ", "News", "Data Champions", "Events", and "Contact Us". The main content area is titled "Data Champions Search" and displays two profiles:

- Shawn Zamani**: PHD Student, Early Detection Programme, MRC Cancer Unit. Areas of expertise include R language, Git, SQL, Data visualisation, Statistical analysis, Electronic Lab Notebooks, Data collection, Version control, Institutional repositories, General repositories, Code sharing, Writing/publishing data papers, Surveys, Geospatial data, REDCap, Social media & digital communications, Website development, Bash. Contact information is available.
- Stefanie Felsberger**: PHD student, Centre for Gender Studies, POLIS. Areas of expertise include Data collection, Copyright, Sensitive data, Research data ethics, Data protection and research, Data policies, Data sharing, Open data, Writing/publishing data papers, Surveys, Workshop delivery. Biography is available.

On the right side, there is a search and filter section titled "Search / filter within these results" with an input field, "Apply", and "Reset" buttons. Below this is a "Filter by area of expertise" section with a list of categories and counts:

- Sharing and Storage (74)
- Data management (64)
- Specific techniques and software (64)
- Data analysis (58)
- Coding (51)
- Training and Outreach skills (49)
- Ethics and the law (36)

At the bottom, there is a "Filter by research communities:" section with a partially visible option: Global Medicine (31).

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

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.

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

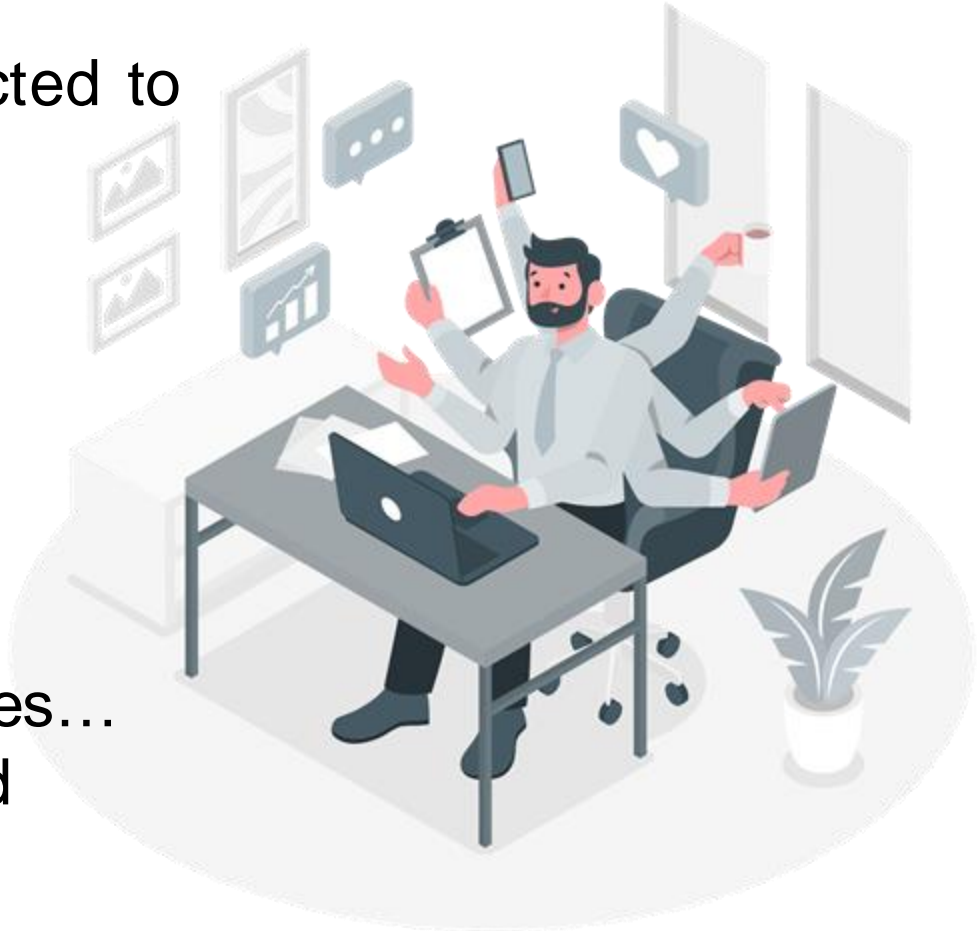
CBIO [lab statement](#):

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...



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

Researchers should not be doing this alone.



Data Stewards



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

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>



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

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

The humans behind infrastructure make a difference!



Nadia Bloemendaal
@Bloemendaal_N

Replying to [@martateperek](#) and [@4TUResearchData](#)

Having gone through the data upload process twice now, I think the librarians are the (secret?) backbone of #FAIR datasets, as they form the frontline by checking if your data complies with the FAIR principles, and suggesting ways to increase the FAIRness of the data!

2:53 PM · Nov 18, 2020 · Twitter for Android

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

The screenshot displays the DataSpace@HKUST interface. At the top, the logo and name 'DataSpace@HKUST' are visible, along with a search icon, 'Service' dropdown, and 'Log In' link. Below the logo is a navigation bar with a 'Metrics' section showing '4,955 Downloads' and a search bar with a 'Find' button and 'Advanced Search' link. The main content area shows search results for '1 to 10 of 32 Results'. On the left, there are filters for 'Dataverses (12)', 'Datasets (20)', and 'Files (572)'. The 'Dataverse Category' is 'Research Project (12)'. The 'Publication Date' filter shows results for 2020 (10), 2015 (7), 2021 (7), 2019 (5), and 2016 (2). The 'Author Name' filter lists HKUIG (Hong Kong Innovative Users Group), Unicode Task Force (3), YU, Jianzhen (3), Campbell, Cameron Dougall (2), GIETEL-BASTEN, Stuart Arthur (2), and HUANG, Jingjing (2). The search results list three datasets:

- China Government Employee Database-Qing (CGED-Q) Jinshenu 1900-1912 Public Release**
Jun 18, 2021 - Lee-Campbell Research Group Dataverse
Campbell, Cameron Dougall, Chen, Bijia; Ren, Yuxue; Lee, James. 2019. "China Government Employee Database-Qing (CGED-Q) Jinshenu 1900-1912 Public Release", <https://doi.org/10.14711/dataset/E9GKRS>, DataSpace@HKUST, V8
The CGED-Q Jinshenu 1900-1912 public release consists of 686,945 records of more than 50,000 officials (based on our linkage) recorded in 46 quarterly editions, along with documentation. The data in the release are as much as possible a direct transcription of the contents of th...
- TAG Data at Shanghai 2018-Nov**
Jun 10, 2021 - Atmospheric decay rates of cooking unsaturated fatty acids Dataverse
YU, Jianzhen. 2021. "TAG Data at Shanghai 2018-Nov", <https://doi.org/10.14711/dataset/1AOWBD>, DataSpace@HKUST, V1
This data set contains bihourly-measured cooking fatty acid concentrations, hourly O3, temperature and RH data at SAES site in Shanghai for the period of Nov. 9-Dec. 3, 2018.
- Atmospheric decay rates of cooking unsaturated fatty acids Dataverse**
Jun 10, 2021
Cooking fatty acids were monitored in an urban atmosphere in Shanghai for source apportionment and for studying organic aerosol aging.

Below these results, a partial entry for 'Wi-Fi Connection in HKUST Library' is visible.

<https://dataspace.ust.hk/>

Practice and policy

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

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

igatto / QSep-manuscript

Code Issues Pull requests Insights Settings

Assessing sub-cellular resolution in spatial proteomics experiments

proteomics spatial-proteomics public protocols Manage items

95 commits 1 branch releases

Branch: master - New pull request Create new file Upload files

File Name	Description
data	add marker transfer code/figs
figure	Update for bioRxiv
qfignore	add cover letter 2
stavis.yml	add Travis file
Makalu	addressing more reviewers comments
README.md	Update README.md
cover.pdf	add cover letter
cover.tex	add cover letter
cover2.pdf	add cover letter 2
e14am.rda	qsrep assessment section with rB cluster sums
h4m.rda	qsrep assessment section with rB cluster sums
mkLR	Calculate qsrep distribution medians
mkswitch.pca.pdf	incorporate Kathryn and Lisa's comments
mkswitch-qsrep.pdf	incorporate Kathryn and Lisa's comments
mkswitch.rda	minor updates and change marker transfer paragraph
qsrep.R	fix table
qsrep.Rnw	Update for bioRxiv
qsrep.bib	changes to new part in col
qsrep.pdf	Update for bioRxiv
qsrep.tex	Update for bioRxiv
simn.pdf	incorporate Kathryn and Lisa's comments
simn.R	incorporate Kathryn and Lisa's comments

CSH Cold Spring Harbor Laboratory

bioRxiv THE PREPRINT SERVER FOR BIOLOGY

HOME | ABOUT | SUBMIT | ALERTS | RSS | CHANNELS

Search Advanced Search

New Results [View current version of this article](#)

Assessing sub-cellular resolution in spatial proteomics experiments [Comment on this paper](#)

Laurent Gatto, Lisa M Breckels, Kathryn S Lilley
 doi: <https://doi.org/10.1101/377630>
 Now published in *Current Opinion in Chemical Biology* doi: [10.1016/j.cbpa.2018.11.015](https://doi.org/10.1016/j.cbpa.2018.11.015)

Abstract Info-History Metrics Preview PDF

Abstract

The sub-cellular localisation of a protein is vital in defining its function, and a protein's mis-localisation is known to lead to adverse effect. As a result, numerous experimental techniques and datasets have been published, with the aim of deciphering the localisation of proteins at various scales and resolutions, including high profile mass spectrometry-based efforts. Here, we present a meta-analysis assessing and comparing the sub-cellular resolution of 29 such mass spectrometry-based spatial proteomics experiments using a newly developed tool termed QSep. Our goal is to provide a simple quantitative report of how well spatial proteomics resolve the sub-cellular niches they describe to inform and guide developers and users of such methods.

Copyright

The copyright holder for this preprint is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a [CC-BY 4.0 International license](#).

Tweets referencing this article:

ScienceDirect

Outline Download Share Export

ELSEVIER

Current Opinion in Chemical Biology
 Volume 48, February 2019, Pages 123-149

Assessing sub-cellular resolution in spatial proteomics experiments

Laurent Gatto^{1,2,*}, Lisa M. Breckels^{1,2}, Kathryn S. Lilley²

Show more

<https://doi.org/10.1016/j.cbpa.2018.11.015> [Get rights and content](#)
 Under a Creative Commons license [open access](#)

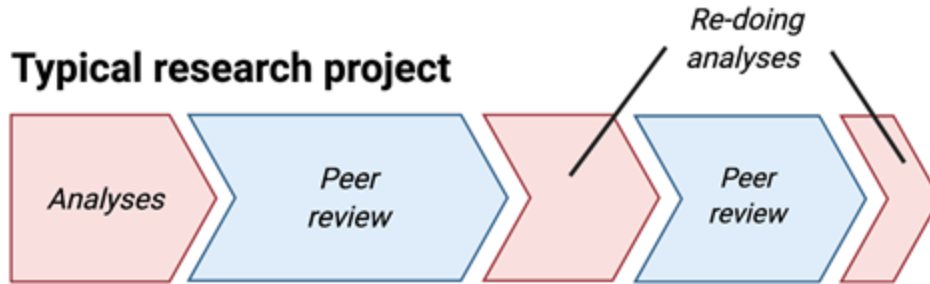
Abstract

The **sub-cellular localisation** of a protein is vital in defining its function, and a protein's mis-localisation is known to lead to adverse effect. As a result, numerous **experimental techniques** and datasets have been published, with the aim of deciphering the localisation of proteins at various scales and resolutions, including high profile mass spectrometry-based efforts. Here, we present a meta-analysis assessing and comparing the sub-cellular resolution of 29 such mass spectrometry-based spatial **proteomics** experiments using a newly developed tool termed QSep. Our goal is to provide a simple quantitative report of how well spatial proteomics resolve the sub-cellular niches they describe to inform and guide developers and users of such methods.

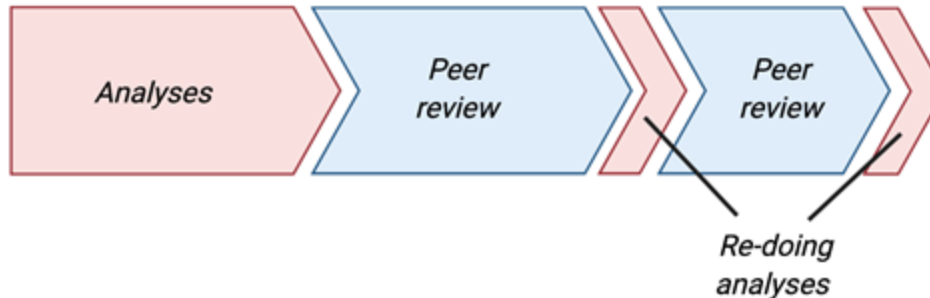
Practice. From left to right: [reproducible document](#) (produced with [open source software](#) and [curated data](#)), [preprint](#) and [peer reviewer paper](#).

Does it take more time to work reproducibly?

Typical research project



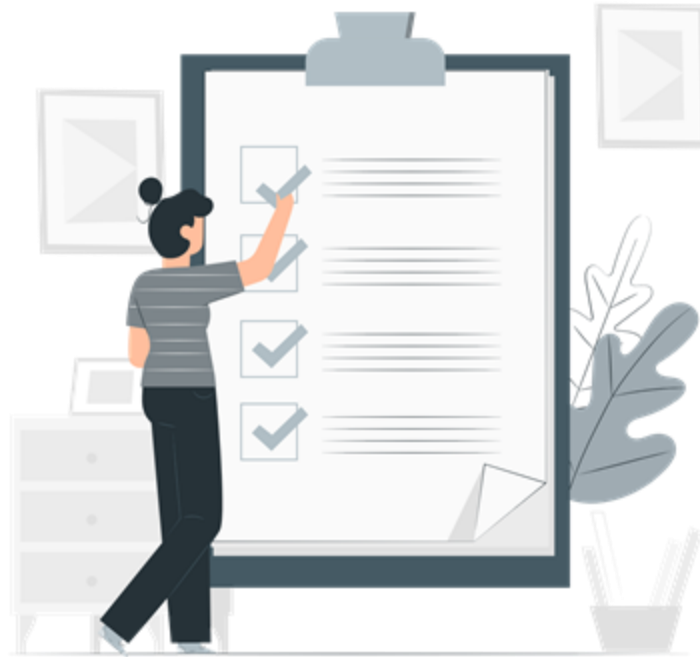
Research project using reproducible practices



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



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/>

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

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

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

Results (3 years later!)

A banner with a dark blue background and light blue abstract shapes. The text "TU Delft Research Software Policy" is written in white, bold, sans-serif font.

TU Delft Research Software Policy

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>

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.

Discipline-specific training	Genomics	Geospatial research	Social Sciences	Python Essentials for GIS
RDM workflows	Tabular Data	Code Refinery	Qualitative data	
Basics	RDM101	Software Carpentry	Personal Research Data	DMPs and RDM basics at faculties
Introduction	The informed researcher	Open Science MOOC	Intro BSc and MSc students	

Legend



Delivered regularly, sufficient capacity



Not in place (insufficient resources to develop)



Delivered regularly, insufficient capacity



Delivered ad-hoc, insufficient capacity

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

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.

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>

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”

Resource Library

A collection of materials to facilitate the development of responsible research and researcher assessment policies and practices.

Search and Filter



Resource type

- Advocacy resources (9)
 - Case studies (14)
 - Good practices (35)
 - Initiatives (9)
 - Journal articles (14)
 - Policies and guidance (12)
 - Position papers (18)
 - Tools (14)
-
- DORA-produced (22)

Per page ▼

1 2 3 4 >>

**GOOD PRACTICES** **POSITION PAPERS** FOR: **RESEARCH INSTITUTES**

Academia In Motion: Recognition & Rewards at Leiden University

In support of the Dutch Recognition and Rewards Programme, Leiden University published a position paper "Academia In Motion: Recognition & Rewards at Leiden University" in 2021. In 2020, Leiden University's Executive Board established a Recognition & Rewards steering committee made up of staff from a variety of positions and roles. The goals of the Recognition...

**JOURNAL ARTICLES** FOR: **RESEARCH INSTITUTES**

Academic criteria for promotion and tenure in biomedical sciences faculties: cross sectional analysis of international sample of universities

[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



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>

Funders are promoting change

Calls > Open Science (OS) Fund 2020/2021

Open Science (OS) Fund 2020/2021

The Open Science Fund aims to support researchers to develop, test and implement innovative ways of making research open, accessible, transparent and reusable, covering the whole range of Open Science. With this call, NWO wants to stimulate Open Science by incentivizing and rewarding researchers who are or would like to be at the forefront of this movement.

Budget: 1M EURO

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

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/>

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.



Community

There is

Open Science as in widely disseminated and openly accessible

and

Open Science as in inclusive and welcoming

Citing [Cameron Neylon](#):





UNIVERSITY OF CAMBRIDGE

Study at Cambridge

About the University

Research at Cambridge

Home / Data Champions



Research Data

Home

Data Management Guide

Support

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



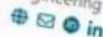
Expertise:
Turbomachinery, CFD-based automated design,
Physics of complex fluids

[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](#)



Guido de Croor

Professor
Aerospace Engineering
Department: Operational Research



Expertise:
My main research focus is the
flight of swarms of tiny flyers
to achieve this, I investigate co-
operative artificial intelligence, which is
the intelligence of small flyers
whose expertise are: artificial intelligence
and machine learning for robotics

[More about Guido de Croor](#)



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



Open Science Community Starter Kit

Set up and foster a local Open Science Community

[Get Started](#)



<http://www.startyouosc.com/>

The

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*

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 Eglén](#), [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 Urrea, 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/>)