



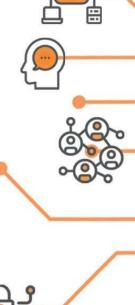
Demystifying Open Science and Research Data Management: A practical workshop for researchers

Dr.ir. Shalini Kurapati

nttps://orcid.org/0000-0003-1492-4100

In collaboration with Dr. Federica Cappelluti, Sara Rollino, Maria Girard, Monica Margara and Enrico Venuto

Turin, 24-10-2019



Today's content

- Busting some myths on Open Science and RDM
- Why Open Science and RDM
- Open Science: What's in for me as a researcher
- RDM in day-to-day research: Tips and tools
- Workshop followed by PoliTo info session





Today's Slides

Can be found at:

Special thanks and attribution to the data stewardship team at TU Delft: https://zenodo.org/communities/datastewards/?page=1&size=20

All the images unless attributed, are from pixabay.com and unsplash.com with a CCO license



Who am I?

Dr.ir. Shalini Kurapati



- Brand new Open Science and RDM fellow at PoliTo (Adjunct role), very excited to be here!
- Past: Data steward at TU Delft, advised researchers on data matters
- PhD and Postdoc at TU Delft, 7 years of research experience
- Co-founder clearbox.ai at I3P, PoliTo
- Certified Information Privacy Professional/Europe



Let talk about Open Science

A little intro exercise

Audience profile

What comes to your mind when you hear the term 'Open Science'?



Busting Myth No.1

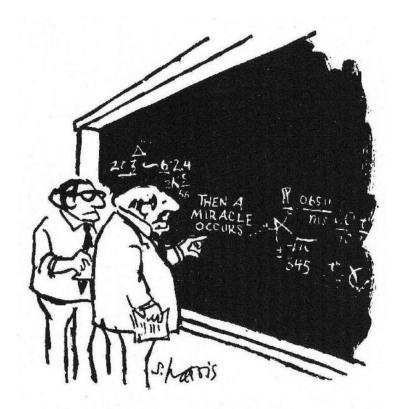
Myth: Open Science means open data

Fact: Open Science ≠ Open Data





Science ≠ Miracles



"I think you should be more explicit here in step two."

10.5281/zenodo.3472587

Open Science means:

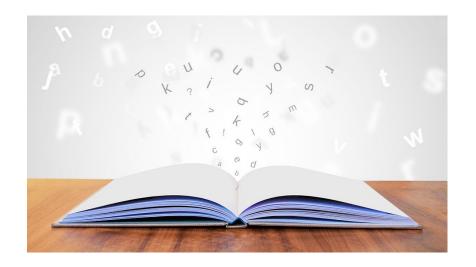
- Evidence based results,
- Transparency, reproducibility, research rigour,
- Validation and verification
- And all other things that basically define science.



OPEN SCIENCE: JUST SCIENCE SCIENCE DONE RIGHT

https://zenodo.org/record/1285575



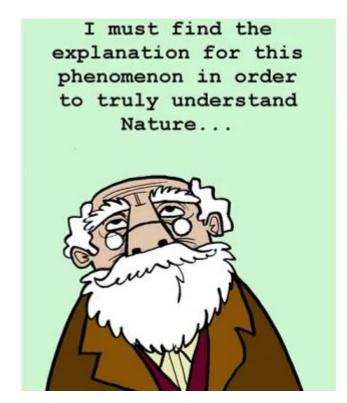


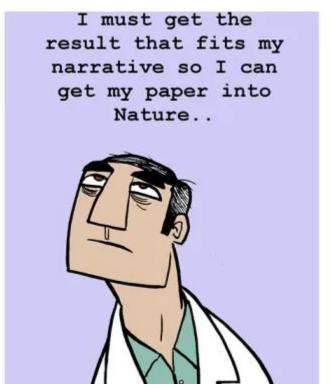
Why Open Science and Research Data Management?



You see the problem?

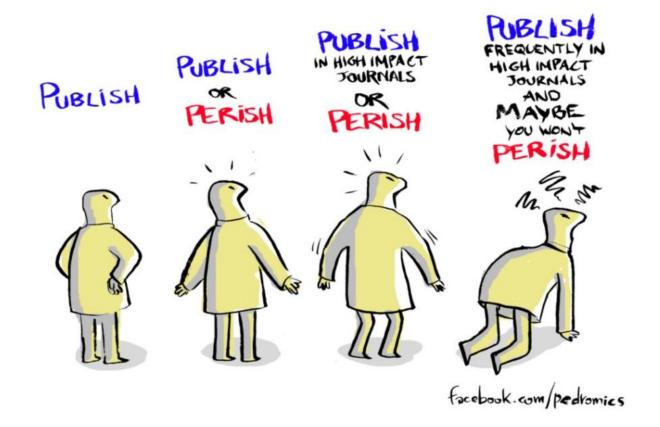
Difference between a 19th century scientist and the 21st century academic





https://www.euroscientist.com/open-scientists-in-the-shoes-of-frustrated-academics-part-i-open-minded-scepticism/

THE EVOLUTION OF ACADEMIA

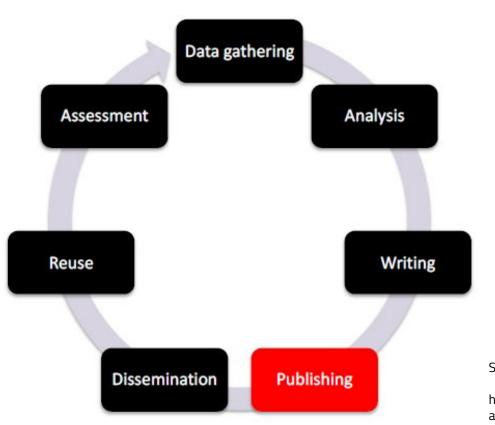


Copyright:

https://velica.deviantart.com/art/Publish-or -Perish-645355248



How did we get here?



The only thing that counts in academia is publication of novel results in high impact journals

Source: Dr. Danny Kingsley

https://www.repository.cam.ac.uk/h andle/1810/276106



Pressure to publish positive results...

- Design the study
- Collect data
- Analyze data as prespecified
- Oops! P> 0.05?
 - Torture data until it confesses
- Then, and only then... write the manuscript







Some consequences

Report finds massive fraud at Dutch universities

Investigation claims dozens of social-psychology papers contain faked data.

Ewen Callaway

When colleagues called the work of Dutch psychologist Diederik Stapel too good to be true, they meant it as a compliment. But a preliminary investigative report

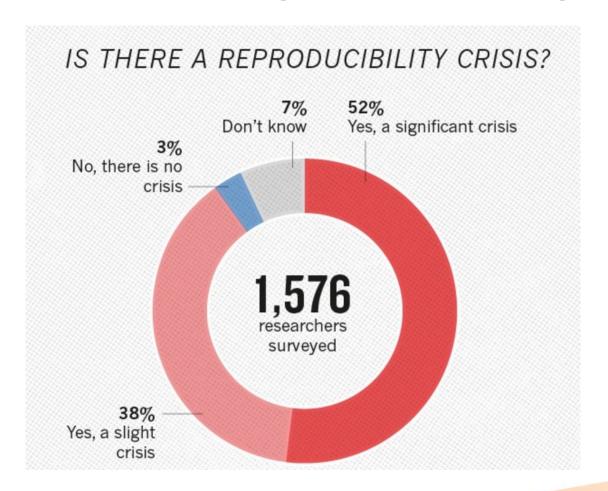
(go.nature.com/tqmp5c) released on 31 October gives literal meaning to the phrase, detailing years of data manipulation and blatant fabrication by the prominent Tilburg University researcher.



It's not all about fraud, but mainly sloppy science



The reproducibility crisis



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



The main reasons for the crisis:



- Selective reporting
- Pressure to publish
- Insufficient supervision and training
- Supporting data / methods/ code not available

Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. Nature, [online] 533(7604), pp.452-454. Available at: https://www.nature.com/news/1-500-scientists-lift-the-lid-on-reproducibi lity-1.19970 [Accessed 26 Apr. 2018].



Funding bodies are pushing for open science, focus on FAIR data

















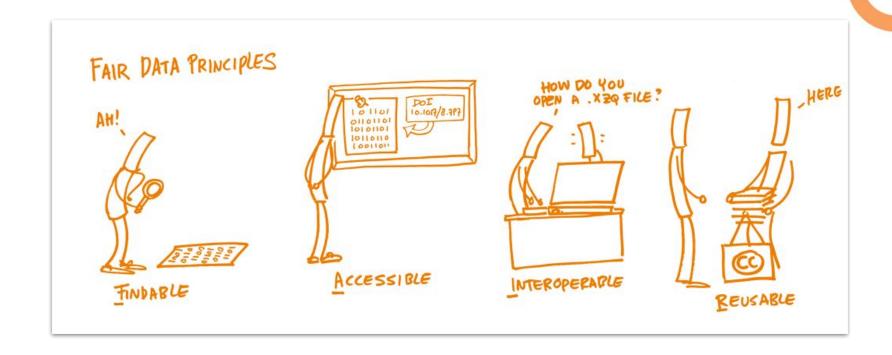


More will follow!

https://www.ukri.org/funding/information-for-award-holders/data-p olicy/common-principles-on-data-policy/



What is FAIR?



Hochstenbach, P. (2018). *Open Research Data Material - FAIR data principles*. [image] Available at: https://hochstenbach.wordpress.com/ [Accessed 26 Apr. 2018].

More on this in future workshops



Research Data Management is key!



We cannot achieve open science or FAIR data without RDM.

It is not simple but crucial!

Solutions at the intersection of technology, culture and awareness and policy



It's not just data!







It's impossible to conduct research without software, say 7 out of 10 UK researchers

By Simon Hettrick, Deputy Director.

No one knows how much software is used in research. Look around any lab and you'll see software – both standard and bespoke – being used by all disciplines and seniorities of researchers. Software is clearly fundamental to research, but we can't prove this without evidence. And this lack of evidence is the reason why we ran a survey of researchers at 15 Russell Group universities to find out about their software use and background.



Headline figures

- · 92% of academics use research software
- . 69% say that their research would not be practical without it
- 56% develop their own software (worryingly, 21% of those have no training in software development)
- 70% of male researchers develop their own software, and only 30% of female researchers do so

- 92% of academics use research software
- 56% develop their own software (21% of those have no training in software development)

https://www.software.ac.uk/blog/2016-09-12-its-imposs ible-conduct-research-without-software-say-7-out-10-u k-researchers



Software reproducibility!



24 May, 2018
17 participants, good balance
between researchers, data
stewards & RSEs

What comes to your mind when we talk about software reproducibility?

■ Mentimeter

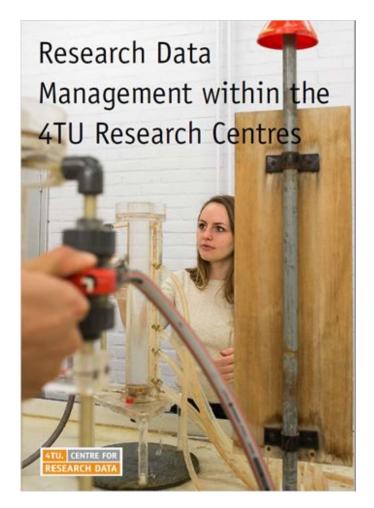


https://openworking.wordpress.com/2018/06/26/workshop-report-software-reproducib ilitv-how-to-put-it-into-practice/



15

Software integral part of RDM



Software sustainability is an important and much discussed topic, particularly in the computer science and applied mathematics communities, but also in materials science.

Cruz, Maria J., and Alastair Dunning. 2018. "Research Data Management Within the 4TU Research Centres." OSF Preprints. March 20. doi:10.31219/osf.io/sgftw.



Myth no.2

Myth 2: Open Science and RDM are only for the benefits of the others

Fact: The biggest beneficiaries of OS and RDM are the researchers themselves



We all know OS and RDM are good for the society



(C)



Selfish benefit: Avoiding disasters





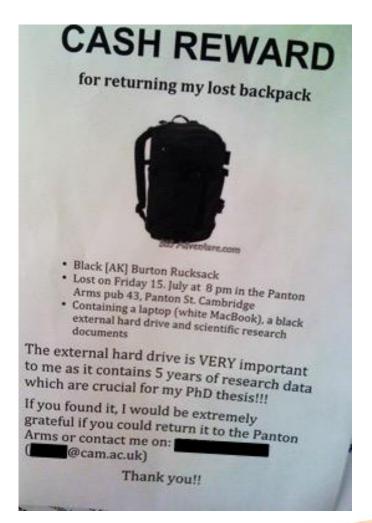
What is your reaction if...

- Your laptop got stolen
- You lost your USB stick
- Your portable hard drive got damaged
- Your stuff on Dropbox / Google Drive disappeared





These things happen







These things happen

Visiting Professor Robbed At Gunpoint, Years Of Research Data Stolen -



Visiting Professor Robbed At Gunpoint In Davis, Years Of Research ...

The search is on for the suspects who robbed a UC Davis professor at gunpoint - and now years of valuable scientific research data have been s...

sacramento.cbslocal.com

https://sacramento.cbslocal.c om/2019/09/26/uc-davis-vis iting-professor-robbed-gunpo int-research-stolen/



These things happen



SHARE

LETTERS

Editorial expression of concern

In the 3 June issue, *Science* published the Report "Environmentally relevant concentrations of microplastic particles influence larval fish ecology" by Oona M. Lönnstedt and Peter Eklöv (1). The authors have notified *Science* of the theft of the computer on which the raw data for the paper were stored. These data were not backed up on any other device nor deposited in an appropriate repository. *Science* is publishing this Editorial Expression of Concern to alert our readers to the fact that no further data can be made available, beyond those already presented in the paper and its supplement, to enable readers to understand, assess, reproduce, or extend the conclusions of the paper.

Reference

1. ← O. M. Lonnstedt, P. Eklov, Science 352, 1213 (2016). ejournals@cambridge - find full text Abstract/FREE Full Text



My own experience

3/11/2015 12:00 pm

Game data: Online, no local storage, sometimes data loss due to network issues, servers in NL, sessions in US, DE and NL

Shalini_Kurapati_profile.docx

Name +

Session_10_03 Qualtrics Survey.xlsx

Salomc

Statistics

Supply chain disruptions book

Thesis

Latex format thesis.zip

Scoring game level1 modified.xlsx

15104;1000;110000;110000;110000;110000;110000;110000;110000;110000;110000;110000;110000;110000;11000

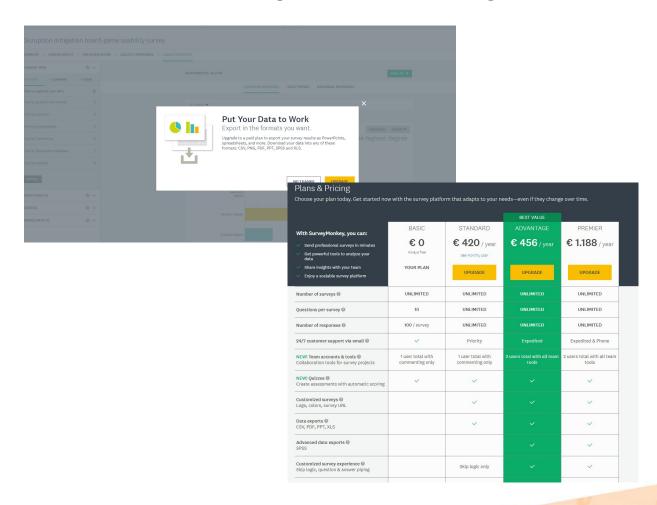
Post-game survey embedded in the game platform, limited access

MATB data local: collected on pendrive, sometimes computer crashed, data loss

A01: Anony.Player id

- COMM 2014 10031139.txt
- MATB 2014 10031139.txt
- RMAN_2014_10031139.csv 🖈
- = RMAN_2014_10031139.txt
- = SYSM_2014_10031139.txt
- TRCK 2014 10031139.csv
- = TRCK_2014_10031139.txt
 - WRS 2014 10031139.csv
 - WRS_2014_10031139.txt

My own experience

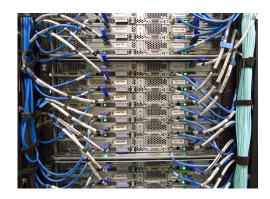


Used survey monkey in 2012. They changed policy for exporting my own data!



Data backup strategy

- Departmental/institutional backup system
- External drives
- Online backups
- At least two backups, at two different locations





Don't leave it to Google!

Google services Terms of Use:

When you upload, submit, store, send or receive content to or through our Services, you give Google (and those we work with) a worldwide license to use, host, store, reproduce, modify, create derivative works (such as those resulting from translations, adaptations or other changes we make so that your content works better with our Services), communicate, publish, publicly perform, publicly display and distribute such content. The rights you grant in this license are for the limited purpose of operating, promoting, and improving our Services, and to develop new ones. This license continues even if you stop using our Services (for example, for a business listing you have added to

https://policies.google.com/terms?hl=en



PoliTo has a lot to offer

- The Information Technology area provides services to research: computing and data backup needs.
- Researchers can request secure backup (for regular data) in two different locations for upto 6 months of retrieval
- Applications: Web server, database, computing power, software, and virtual infrastructure (table)

Virtual N	/lachines	Virtual CPUs	RAM [GB]	Storage [TB]	Backup [TB]
3	0	98	250	11	40



PoliTo has a lot to offer

HPC, Private cloud and Big Data services in collaboration with SmartData@PoliTO e HPC@POLITO.































Selfish benefit: Being smart and efficient and professional

- What if someone asks you for data supporting your publication?
- What if someone asks you for data supporting your publication, 5 year after publication?
- What if the request comes 10 years later?

Forget others, what if YOU need to revisit that data?

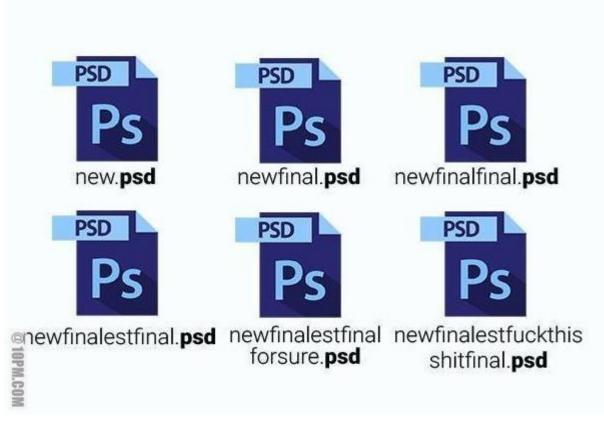


Recognise this?





And this?



Copyright: http://10pm.com/



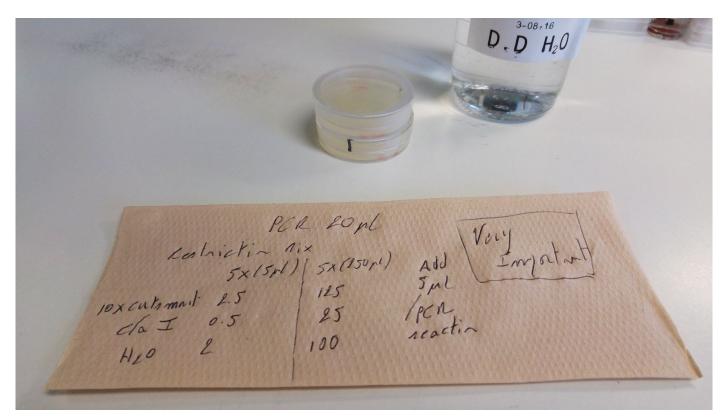
And this...



In three years would you recognise these?



Or Maybe..





Vincent Gaggioli

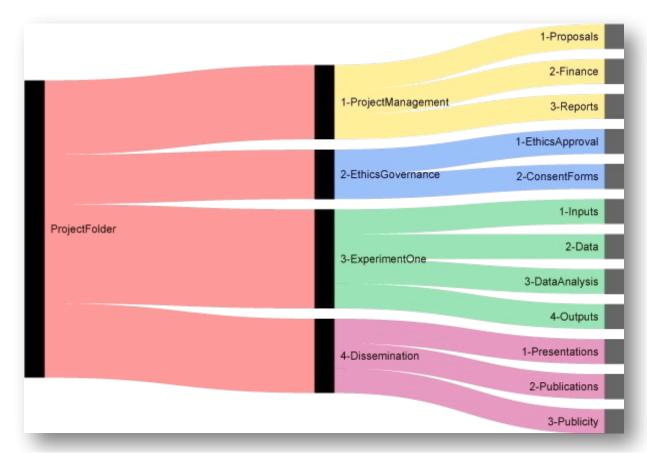


Data organisation: Tips

- Consistent
- Meaningful to you and your colleagues
 - Think about the continuity of your research
- Would you be able to easily find your own data files?
- Would your colleague be able to easily find and understand your data files?
- Critically look at your files (and clean them if needed!) once a week



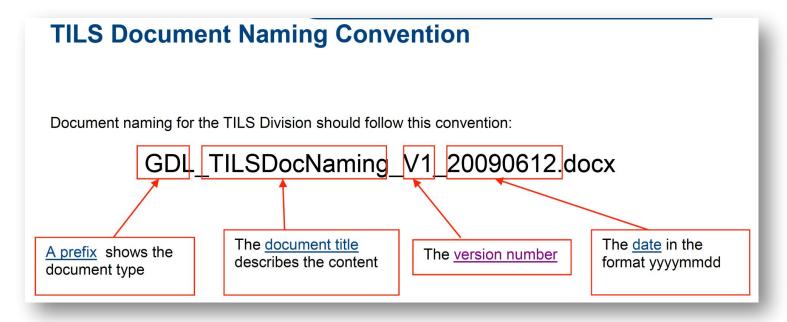
Folder organisation



Copyright: http://nikola.me/folder structure.html



File naming: Human and machine friendly



http://www.data.cam.ac.uk/files/gdl tilsdocnaming v1 20090612.pdf



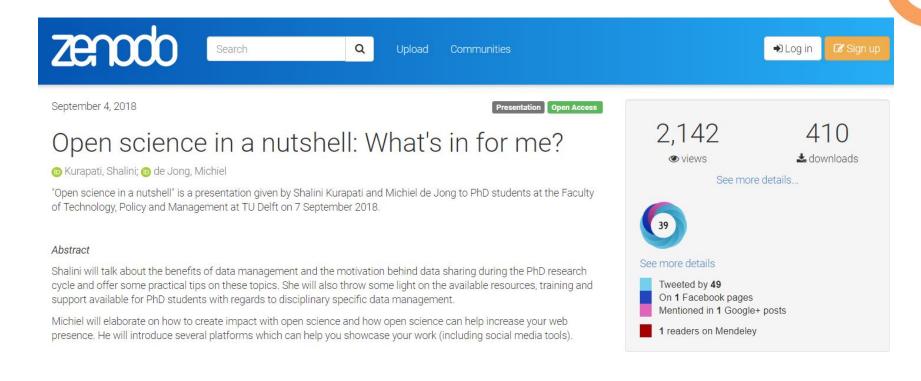
Selfish benefit: Increasing impact

Who looks into your data set? Google does!





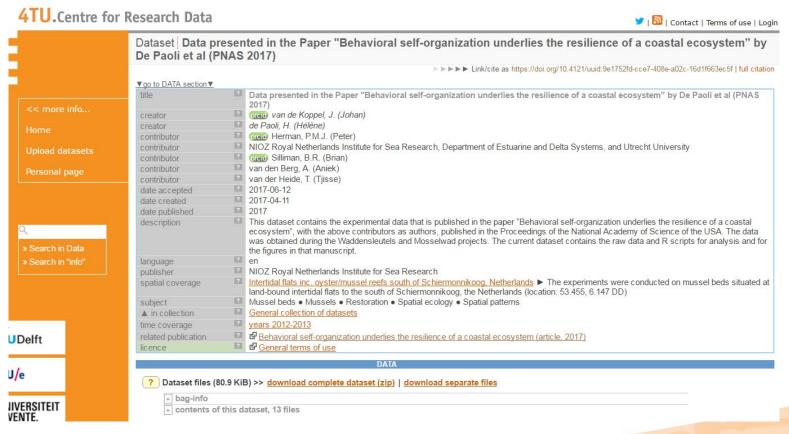
Publish your presentations



https://doi.org/10.5281/zenodo.1409027



Publish your data in a certified repository



https://doi.org/10.4121/uuid:9e1752fd-cce7-408e-a02c-16d1f663ec5f



Track the impact

Behavioral self-organization underlies the resilience of a coastal ecosystem

H de Paoli, T van der Heide... - Proceedings of the ..., 2017 - National Acad Sciences Abstract Self-organized spatial patterns occur in many terrestrial, aquatic, and marine ecosystems. Theoretical models and observational studies suggest self-organization, the formation of patterns due to ecological interactions, is critical for enhanced ecosystem resilience. However, experimental tests of this cross-ecosystem theory are lacking. In this study, we experimentally test the hypothesis that self-organized pattern formation ... Cited by 2 All 6 versions Web of Science: 1 Cite Save

M Create alert



Publish your code



https://doi.org/10.5281/zenodo.1065538

You can event make your github code citable



Increase your funding chances

Winning Horizon 2020 with Open Science

Grigorov, Ivo; Davidson, Joy; Donnelly, Martin; Jones, Sarah; Elbaek, Mikael

Investing in research towards solutions for Societal Challenges is a key priority of the EU Innovation Union[1], and the EC has placed focus on "Open Innovation, Open Science and Open to the World"[2] as a strategy to make the EU leader in Research & Innovation.

Open Science is about removing all barriers to full sphere basic research knowledge and outputs (research data, researc code, publications, policy briefs), and thus feeding Open Innovation and the knowledge-based economy. The concept is central to EU's Responsible Research and Innovation philosophy[3], and removing barriers to basic research measurably contributes to the EU's Blue/Green Growth Agenda[4]. Despite the potential of the internet age to deliver on that promise only 50% of today's basic research is freely available[5].

The workshop will demonstrate how and why Open Science can be a key tool for research managers formulating and managing grant proposals.

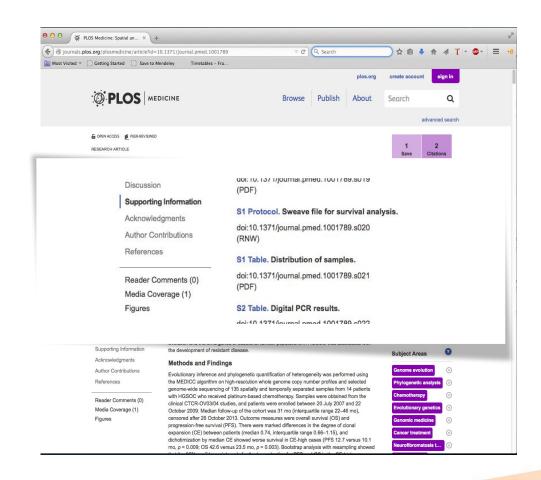
Learning Outcomes follow FP7 FOSTER Open Science Learning Objectives (http://dx.doi.org/10.5281/zenodo.15603):

- Understand EC's Open Science Agenda priorities;
- Gain insight into what is, and will be, mandated on access to research, and data management;
- Be capable to integrate Open Science in the research concept, and its impact;
- Understand Open Science's impact on EC Evaluation process based on examples from 2014-2015.

https://zenodo.org/record/46602#.XbCnKeqzZPY



Helps you while writing papers



Capture all the complexity of your research beyond just text and pdf



Being world class

Every major university in the US and Europe has embraced Open Science and RDM!

MIT: https://libraries.mit.edu/data-management/

Cambridge: https://www.data.cam.ac.uk/

Caltech: https://libraries.mit.edu/data-management/

ETH: https://ethz.ch/services/en/service/a-to-z/research-data.html

if you permit me,

TU Delft:

https://www.tudelft.nl/en/library/current-topics/research-data-management/research-data-management/data-stewardship/



But.. what about my career



Things are changing.. slowly but surely!



Top journals need it already!

nature > nature research > editorial policies > reporting standards and availability of data, materials, code and protocols

natureresearch



Editorial policies

Authorship

Competing interests

Confidentiality

Plagiarism and duplicate publication

Image integrity and standards

Preprints & Conference Proceedings

Peer-review policy

Reporting standards and availability

Reporting standards and availability of data, materials, code and protocols

An inherent principle of publication is that others should be able to replicate and build upon the authors' published claims. A condition of publication in a Nature Research journal is that authors are required to make materials, data, code, and associated protocols promptly available to readers without undue qualifications. Any restrictions on the availability of materials or information must be disclosed to the editors at the time of submission. Any restrictions must alsobe disclosed in the submitted manuscript.



Decision makers are pushing for change



Evaluation of Research Careers fully acknowledging Open Science Practices

Rewards, incentives and/or recognition for researchers practicing Open Science

https://cdn1.euraxess.org/sites/default/files/policy library/os-rewards-wgreport-final integrated 0.pdf

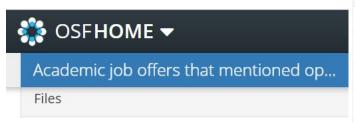


Funders might prefer those who practise Open Science.. soon

Open Science Career Assessment Matrix (OS-CAM)	
Open Science activities	Possible evaluation criteria
RESEARCH OUTPUT	
Research activity	Pushing forward the boundaries of open science as a research topic
Publications	Publishing in open access journals Self-archiving in open access repositories
Datasets and research results	Using the FAIR data principles Adopting quality standards in open data management and open datasets Making use of open data from other researchers
Open source	Using open source software and other open tools Developing new software and tools that are open to other users
Funding	Securing funding for open science activities



There's a whole collection of academic jobs mentioning OS



https://osf.io/7jbnt/

Name AV	Modified A V
$\ensuremath{\mathfrak{D}}$ Academic job offers that mentioned open sci	
 OSF Storage (United States) 	
180911_OTTT_Sensory_Sciences_deuts	2019-03-01 08:24 AM
180911_OTTT_Sensory_Sciences_englis	2019-03-01 08:24 AM
2019_HH-Uni_W2-3_Sportpsychologie.p	2019-05-21 08:26 AM
Aachen_W3_Prof_Angewandte_comput	2019-04-10 11:52 AM
AllgemeinePsychologieUHH.pdf	2018-10-12 08:51 AM
Announcement_W3_Professur_SocialPs	2018-01-18 02:25 PM
Bielefeld_Psychologie_Sportwiss_W2_P	2019-04-23 11:53 AM
Charite_Berlin.png	2018-03-10 09:08 AM
Dresden_WiMi.pdf	2018-07-13 12:37 PM
Keele_Lecturer_2018.pdf	2018-07-11 01:14 PM
Leipzig-W2_Professorship_in_Psycholog	2019-05-17 06:11 AM



There is no looking back really



Open Science is the only way forward in the post truth world

Comic by https://thenorwichradical.com/2017/01/12/post-truth-politics-and -the-war-onintellect/



Recap

- Open Science is just science done right
- Plenty of selfish benefits to researchers
 - Avoiding disaster
 - Being smart and efficient
 - Visibility and impact
 - Trust and reputation
 - Increasing funding and job chances
 - Being world class
 - Being a front runner
 - Preparing for changes to come
- Some practical tips and tricks



My role at PoliTo

Helping Dr. Federica Cappelluti, rector's advisor on open science on:

- Training and Awareness on OS and RDM
- Demonstrating value and driving cultural change towards open science
- Collaborating with university stakeholders
- Design of a policy roadmap
- Advice (as much as I can) on all data related matters of Open Science
- To be fully researcher oriented



Future events in our plan

- Workshop on FAIR data and Data Management Plans
- Personal Data Management
- Industry collaborations and Open Science
- Data and Software Carpentries
- ... and many more
- We'll keep you informed!



Contact

Dr.ir. Shalini Kurapati

shalini.kurapati@polito.it

🏏 @shalini_kr

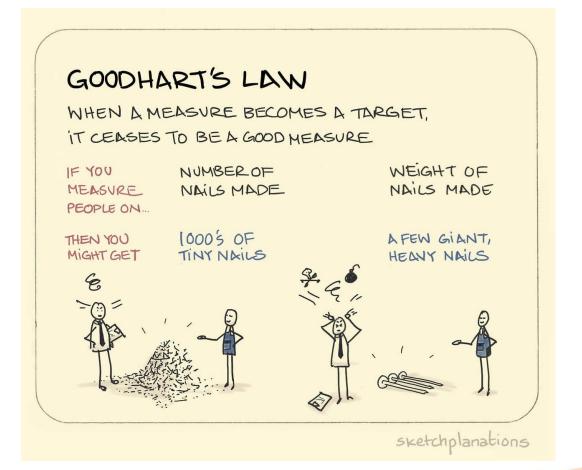
PoliTo working days: Tue, Thu



https://blogs.lse.ac.uk/impactofsocialsciences/20 19/04/30/becoming-a-data-steward/



Thank you!



https://www.sketchplanations.com/post/167369765942/goodharts-law-when-a-measure-becomes-a-target

