



DATA MANAGEMENT LANDSCAPE

from the viewpoint of climate sciences and libraries

Open Data and Data Management – Issues and Challenges

29th October 2018, SCNAT, Bern

Dr. Ana Sesartic Petrus, Research Data Management and Digital Curation



Content

Reproducibility
Crisis

Data Explosion

Demands for
Open Data
and Data
Management

Issues in
Climate
Research

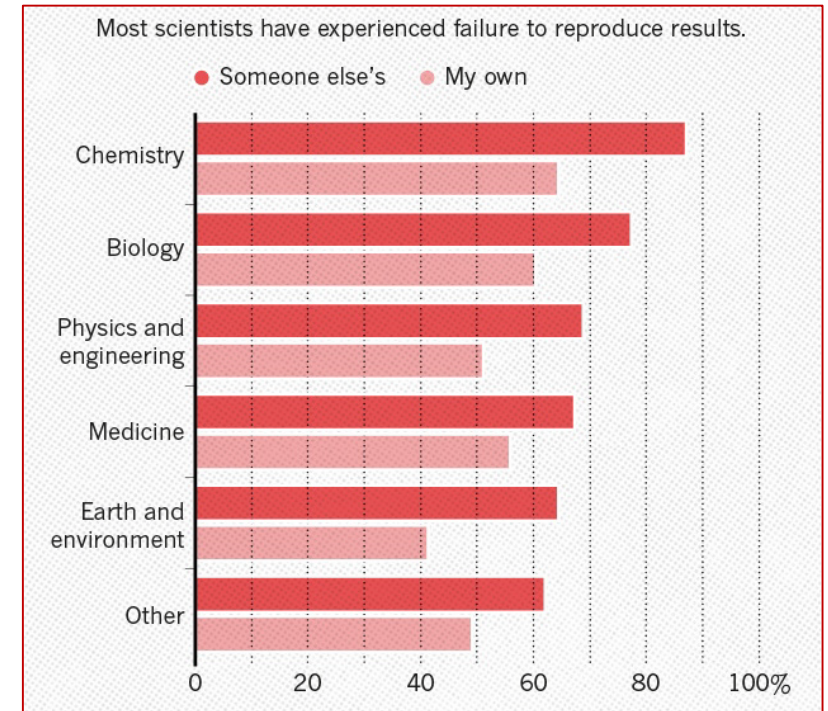
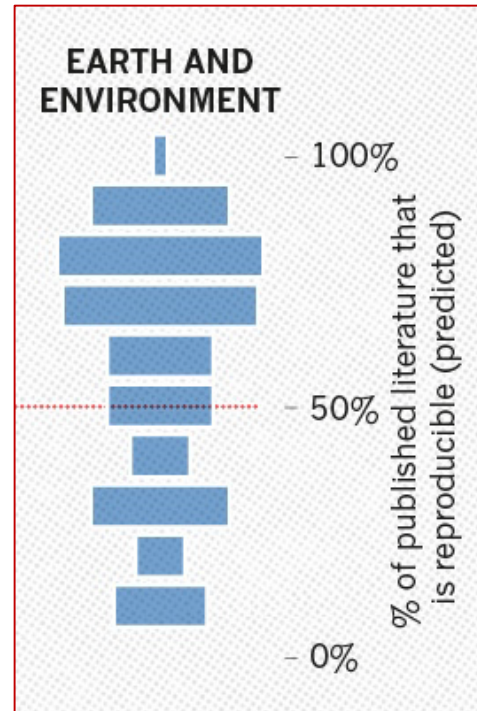
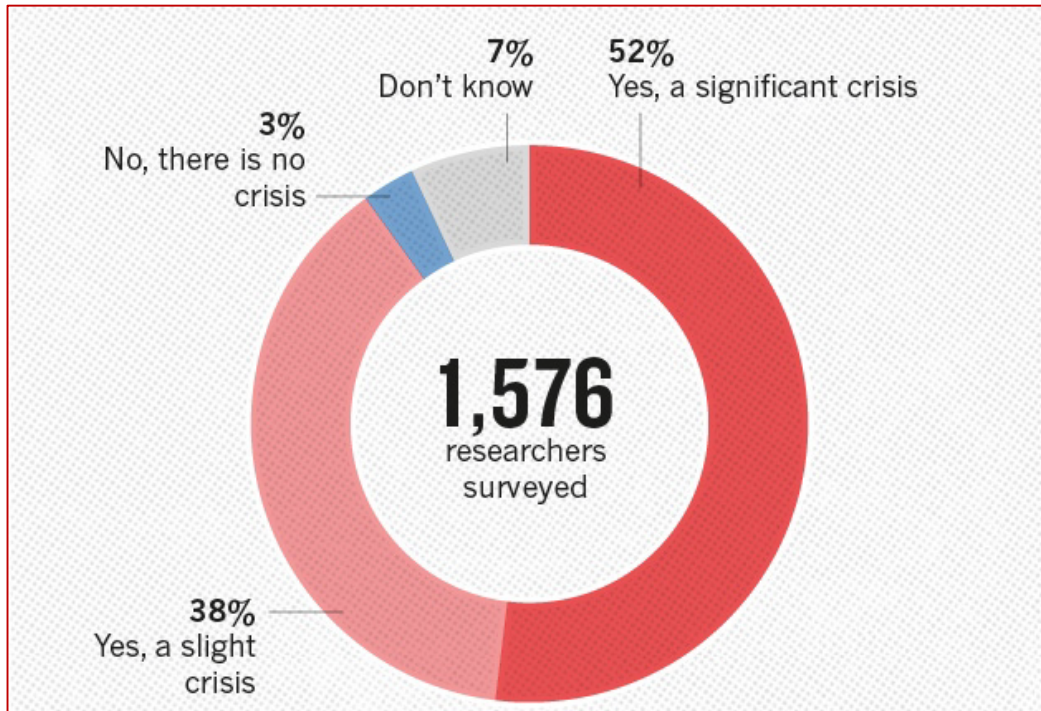
Service Zoo

Let's Discuss!

Research Data Management

- **Organise your data** in a way...
 - ...which not only **allows yourself to track back** what you have done some time ago...
 - ...but also **facilitates sharing and publishing** relevant data...
 - ...and ensures its **comprehensibility** over time
- **Data** means...
 - ...text documents, image files, audio files, software...
 - “A reinterpretable representation of information in a formalized manner suitable for communication, interpretation, or processing.”

Is there a reproducibility crisis?



All graphics from: Baker (2016) *Nature*, doi: 10.1038/533452a

A substantial fraction of research in general, here with focus on Earth and Environmental Sciences is perceived as irreproducible.

Some possible reasons?

- Pressure to publish
- Selective reporting, biased towards positive results
- Unavailability of methods and code
- Unawareness of implicit knowledge and hidden dependencies

According to both our own experiences as researchers and consultants, as well as *Baker (2016) Nature*, doi: 10.1038/533452a

Are we facing a data explosion?

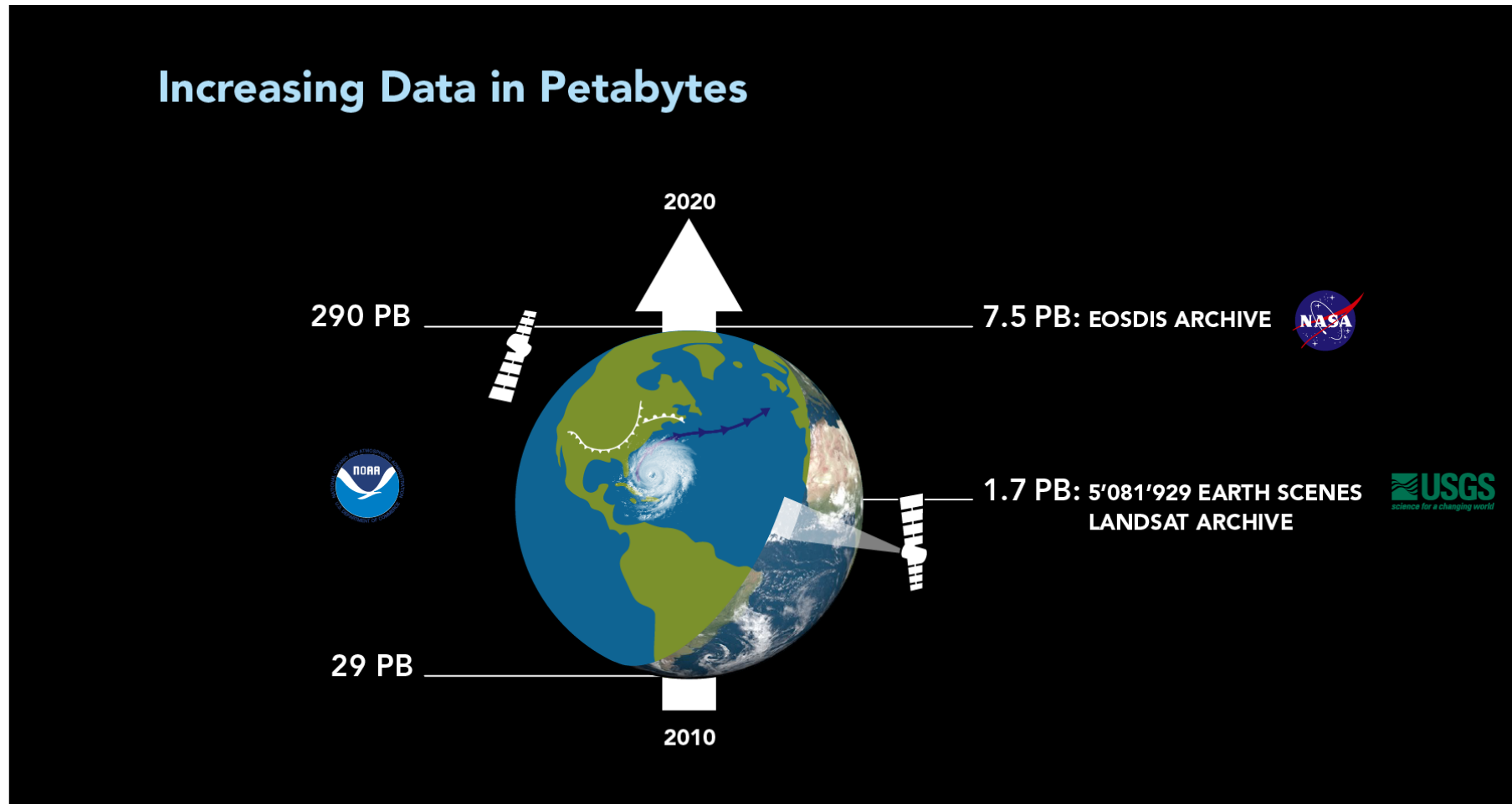


Figure adapted by Andres Bucher based on [scality.com](https://www.scality.com)

Demands for open data and data management



Higher Education Institutions

- Demand scientific integrity
- Good scientific practice calls for reproducibility



FONDS NATIONAL SUISSE
SCHWEIZERISCHER NATIONALFOND
FONDO NAZIONALE SVIZZERO
SWISS NATIONAL SCIENCE FOUNDATION

Funders

- Swiss National Science Foundation (since October 2017)
- Horizon 2020 EU funding programme (since beginning of 2017)



Research Journals

- Science, Nature and AGU journals demand open data / data statements
- Many more journals are doing so, too

Preservation and sharing of (climate) models and data

- **Context** information needed
 - How to run the model
 - Which software is needed (incl. compilers)
 - Which hardware is needed
- Always preserve a **reference output** together with the code, to ensure reproducibility
- When writing a **publication**, preserve the model code, input and output needed to reproduce the results/figures presented in the article
- Save aggregated data if possible (eg. daily or monthly means, depending on time resolution, saving raw output with each timestep not always necessary. But always **refer to community best practices!** No requirements by the ETH.)



[Gulf Stream Sea Surface Currents and Temperatures](#) by NASA/Goddard Space Flight Center Scientific Visualization Studio

There are many services in higher education institutions... ...but they are often unknown to researchers (here some examples)

SNSF

- Collection of **SNSF information** on Open Research Data including **FAQ**:
http://www.snf.ch/en/theSNSF/research-policies/open_research_data/
- SNSF's explanation of the **DMP expected content**:
http://www.snf.ch/SiteCollectionDocuments/DMP_content_mySNF-form_en.pdf

ETH Library

- **Guidance for ETH researchers** on filling out SNSF Data Management Plans:
<https://documentation.library.ethz.ch/display/DD/Guidance+for+ETH+researchers+on+filling+out+SNSF+Data+Management+Plans>
- **Training on data management**, information research, reference management, scientific writing, open access, «**Book a Librarian**»):
www.library.ethz.ch/en/Services/Training-courses-guided-tours
- **Next data management workshop**: 31.10.2018
<http://www.library.ethz.ch/en/Ueber-uns/Veranstaltungen/Datenmanagement-in-der-Forschung-Warum-und-Wie-Zentrum>

Scientific IT Services

- Trainings and courses (e.g. openBIS, Python, bioinformatics, etc.): <https://sis.id.ethz.ch/consulting/>
- Active research data management support, for data management during the course of research
www.sis.id.ethz.ch/researchdatamanagement
- openBIS **Electronic Lab Notebook & Laboratory Information Management System**
<https://labnotebook.ch/>

Many insecurities and questions remain...

- Reproducibility: what does it really mean for your field?
 - Appraisal: what should be preserved and shared?
 - What metadata suits a community best?
 - How much can and/or should we reduce data volume and costs?
 - Preparing data for easy use by third parties takes a lot of time and resources. Is this sufficiently covered by research funding?
 - Who can or should pay for data storage and preservation once a project is finished?
- Each community has different needs, one size does NOT fit all. Do the current guidelines take this sufficiently into account?
- Let's discuss!

Contact

Dr. Ana Sesartic Petrus

ana.petrus@library.ethz.ch

Research Data Management and Digital Curation

ETH Library

ETH Zurich

<http://www.library.ethz.ch/RDM>

data-management@library.ethz.ch

Research Data @ ETH

www.ethz.ch/researchdata
researchdata@ethz.ch