Research Data Management & Open Science

Trends, Funding Requirements and Services at LMU

Prof. Dr. Felix Schönbrodt
Ludwig-Maximilians-Universität München

2021-04-22

www.nicebread.de
www.researchtransparency.org
@nicebread303

Today's program

- *Prof. Dr. Felix Schönbrodt* (LMU Open Science Center):
  Open Data - What & Why

- *Dr. Florian Schreck, Dr. Veit Schwab* (LMU Unit for Research Funding):
  Funder's policies: How to presenting data management and open data in the application

- *Laura Meier, Dr. Martin Spenger* (LMU University Library):
  Effective data management workflows: Tools and resources at LMU

Open Science in the research process

LMU Open Science Center

Preregistration ***Power Analysis***

Replication study

Replicate results

Formulate hypotheses & analysis plan

Registered Report ***(1st phase)***

Open Access

Publish & distribute research output

Registered Report

(2nd phase)

Collect data

Open Lab Notebook

Open Data

Interpret & report results

Analyze data

Open Materials

Open Analysis Code

get all material here: https://osf.io/zjrhu/

OSC
LMU Open Science Center

Why open data?

Why open data?

OSC
LMU Open Science Center

- 1. *Nullius in verba* -
  take nobody's word for it
  - Motto of the oldest scientific society (Royal Society, founded 1660)
  - Science is not built upon blind trust, but on verifiability.
  - „Organized skepticism“ (Merton, 1947)

Only when raw data (and other research material) are shared organized skepticism can be enacted, and science can really be self-correcting.
Open data is one part of good scientific practice.

Why open data?

- 2. *Efficiency and Inclusiveness*
  - *Speedy responses in outbreaks; share rare and hard-to-collect data*

OECD Policy Responses to Coronavirus (COVID-19)

Why open science is critical to combatting COVID-19

Updated 12 May 2020

Matt Might
@mattmight

Let's remember this after we cure COVID19. Let's remember how open science shortened the time to find a treatment. And, let's never go back to the old ways again. The moral necessity of open access to all taxpayer-funded research is unassailable now.

Tweet übersetzen

The Economist @TheEconomist · 17. März 2020

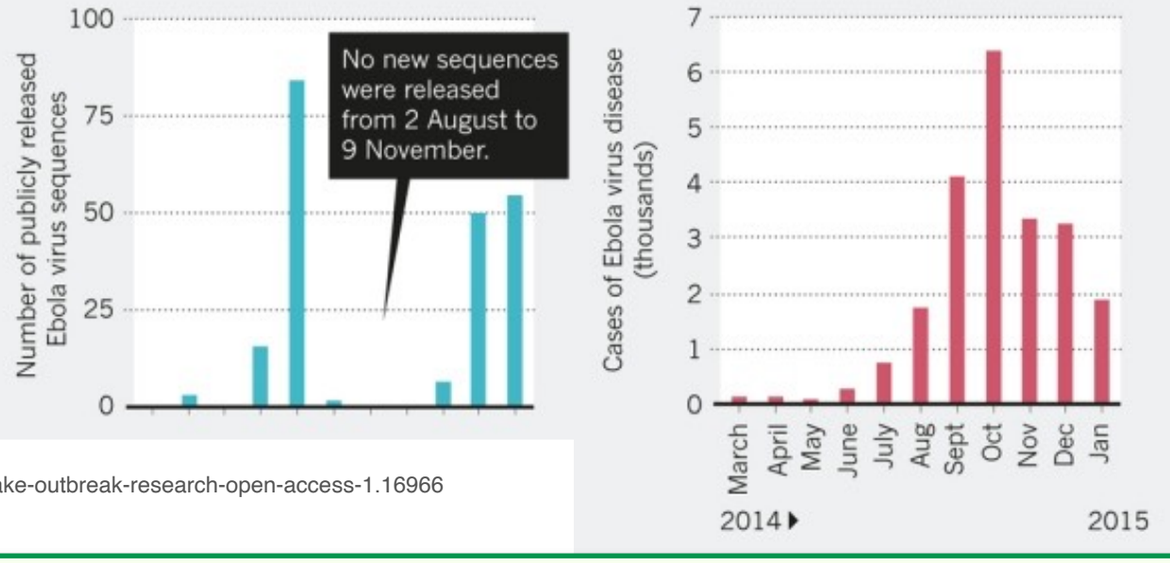
There is no treatment for 90% of rare diseases. Personalised medicines can target such illnesses econ.st/3cZVmN6

Data sharing: Make outbreak research open access

Nathan L. Yozwiak, Stephen F. Schaffner & Pardis C. Sabeti

GAPS IN THE DATA

Genome sequences from the West Africa outbreak of Ebola virus were first made publicly available in April 2014. Since 99 genomes were released in July, data sets have been shared sporadically, even though more are known to have been generated.



https://www.nature.com/news/data-sharing-make-outbreak-research-open-access-1.16966

The covid-19 pandemic has shown how fast scientific progress can be when we share our data and knowledge freely, and that free knowledge is a moral imperative.

Why open data?

- *3. Public money = public good*

https://publiccode.eu

Publicly funded research data does not belong to the researcher who collected it.
S/he has the right of primary usage, but after that the data should be considered a public good (of course respecting privacy rights and applicable copyrights).

Why open data?

- *4. Data persistence*
  - *never lose data due to a crashed hard disk rive*

A publicly funded repository is the right place for long term storage of research data – not your private USB stick, your personal university website (that vanishes after you change affiliation), or the journal's online supplemental material that hides the data behind a paywall.

Why open data?

LMU Open Science Center

- *5. More and more funders and journals demand it.*

Journal	Total	Data Citation	Data Transparency	Analysis Code Transparency	Materials Transparency
The Journal of Politics University of Chicago Press	9	0	3	3	0
Journal of Economic Perspectives American Economics Association	9	2	3		
Journal of Economic Literature American Economics Association	7	0	3		
American Journal of Political Science Wiley	11	0	3		
Nature Methods Springer Nature	11	1	3	1	1
AEA Papers & Proceedings American Economics Association	7	0	3	3	0
Journal of Legal Studies University of Chicago Press	10	1	3	3	3
Journal of Experimental Political Science Cambridge University Press	16	1	3	3	3
Journal of Peace Research SAGE	10	1	3	3	3
Biometrical Journal Wiley	7	0	3	3	0

TOP GUIDELINES
TRANSPARENCY AND OPENNESS PROMOTION
https://www.topfactor.org

TOP level 3: Data must be posted to a trusted repository, and reported analyses will be reproduced independently prior to publication.

The publishing and funding landscape is changing towards open science. Do you want to get a lot of money and publish in top journals?

LMU Open Science Center

What is open data?

What is Data?

OSC
LMU Open Science Center

"The recorded factual material commonly retained by and accepted in the scientific community as necessary to validate research findings."

(EPSRC, 2018)

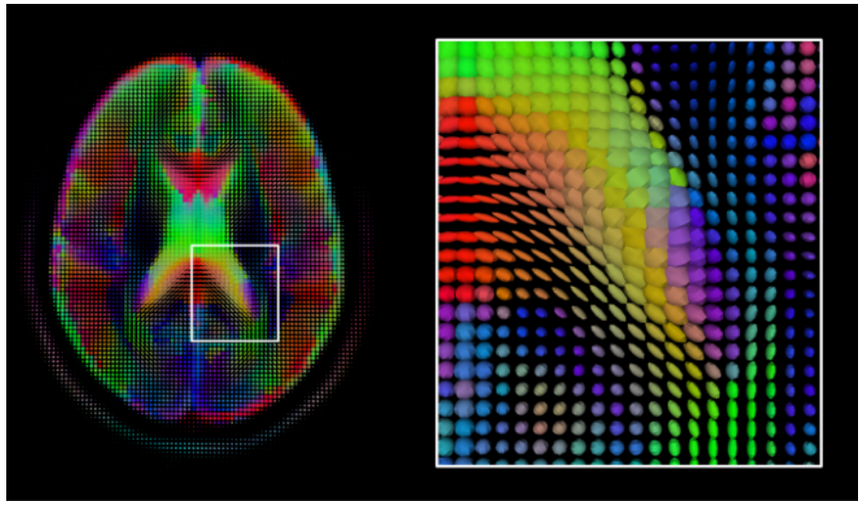
Anything and everything produced in the course of research.

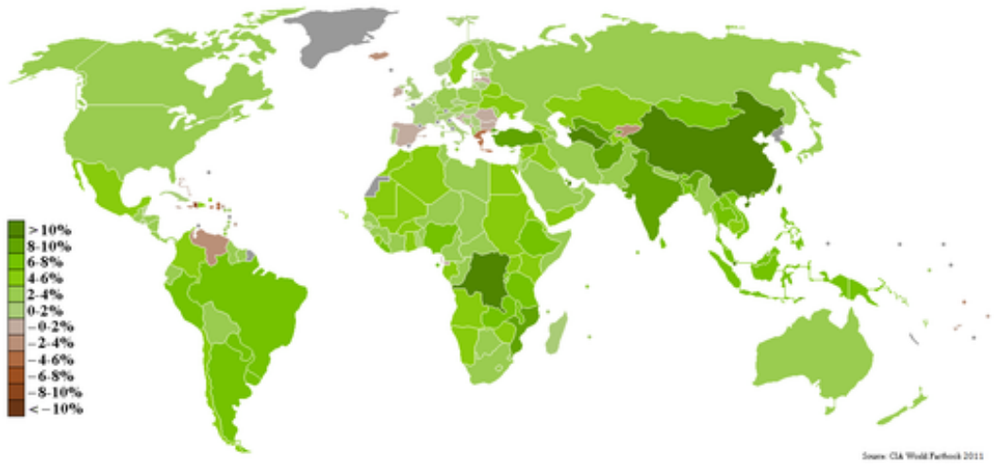
Lynch (2014)

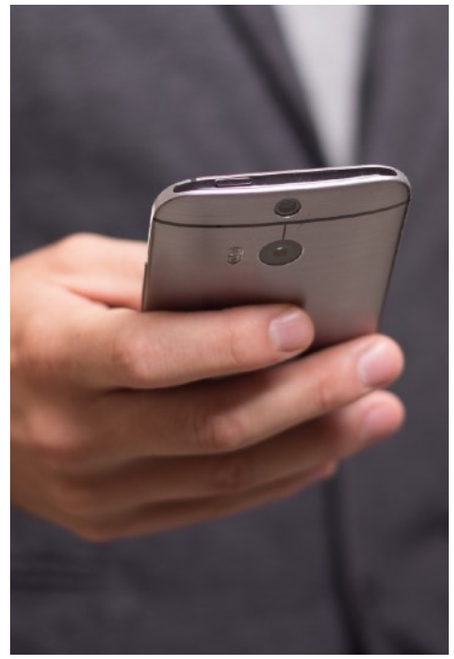
Icon from flaticon.com by freepic

What is Data?

OSC
LMU Open Science Center







➙ we need field-specific definitions: What constitutes „research data“?

Pictures from en.wikipedia.org/w/index.php?curid=36808161; pixabay.com; pxhere.com/en/photo/1192476; pxhere.com/en/photo/914805; pxhere.com/en/photo/595225; commons.wikimedia.org/wiki/File:Gdp_real_growth_rate_2007_CIA_Factbook.PNG

Example: Psychology

Thus, primary data in psychology are unaltered (i.e., untransformed, not aggregated, etc.) quantitative or qualitative data available in digital form, e.g.:

- Each manipulated and measured variable of every experimental session of every study participant in an experiment;
- Each response of every person to each item in a survey;
- Original wording of inputs in free text fields;
- Digitized video recordings;
- […].

Primary data also include the data of cases that were excluded from the analyses (with the exception of those cases in which participants withdrew their consent during or after data collection).
To summarize, we define primary data as the set of all data points collected during the course of a study or project, as initially digitized, but otherwise in a completely unaltered form.

Recommendations of the German Psychological Association, https://psyarxiv.com/24ncs/

DFG

OSC
LMU Open Science Center

Fachspezifische Empfehlungen zum Umgang mit Forschungsdaten

- Stellungnahme der Akademie für Soziologie zum Umgang mit Forschungsdaten in der Soziologie (PDF | 219 KB)
- Stellungnahme des Fachkollegiums 111 „Sozialwissenschaften" zum Forschungsdatenmanagement in der Soziologie, der Politikwissenschaft und der Kommunikationswissenschaft (PDF | 100 KB)
- Stellungnahme zum Umgang mit Forschungsdaten in der Erziehungswissenschaft, Bildungsforschung und Fachdidaktik (PDF | 803 KB)
- Handreichung des Fachkollegiums 101 zum Umgang mit Forschungsdaten (PDF | 45 KB)
- Handreichung des Fachkollegiums 106 zum Umgang mit Forschungsdaten (PDF | 124 KB)
- Bereitstellung und Nachnutzung von Forschungsdaten in der Soziologie: Stellungnahme des Vorstands und Konzils der DGS (PDF | 761 KB)
- Digitaler Wandel in der Wissenschaft: Herausforderungen und Chancen für das Fachgebiet Materialwissenschaft und Werkstofftechnik (PDF | 92 KB)
- Information für die Wissenschaft Nr. 66/2015: DFG verabschiedet Leitlinien zum Umgang mit Forschungsdaten ›
- Information für die Wissenschaft Nr. 36/2015: Richtlinien zum Umgang mit Forschungsdaten in der Biodiversitätsforschung ›
- Richtlinien zum Umgang mit Forschungsdaten in der Biodiversitätsforschung (PDF | 192 KB)

https://www.dfg.de/foerderung/antrag_gutachter_gremien/antragstellende/nachnutzung_forschungsdaten/index.html

Not only open, but FAIR

OSC
LMU Open Science Center

The FAIR principles




Findable **A**ccessible **I**nteroperable **R**eusable

Good research data management

Icon from flaticon.com by freepic, mynamepong, monkik

FORCE11(2011)

Balancing values:
Three fields of tension with (human subject) data

Open Data, public interest/ entitlement to publicly funded data

Privacy rights of research subjects

Right of first usage, incentives to collect data in the first place

Optimal and efficient gain of knowledge by data reuse

Reproducibility and verifiability of published analyses

Protect original authors against inadequate burden and potential attacks

Balancing values: Three fields of tension with human subject data

OSC
LMU Open Science Center

Open Data, public interest/ entitlement to publicly funded data

Privacy rights of research subjects

- Privacy rights > openness; but also: „legitimate interest“ of research
- Ask participants for a broad consent of open reuse
- Restrict access with „scientific use files“; publish aggregated data (e.g., ratings of videos) without the primary data (videos)
- Sharing something > sharing nothing
- As open as possible, as restricted as necessary

Balancing values: Three fields of tension with human subject data

LMU Open Science Center

Right of first usage, incentives to collect data in the first place

Optimal and efficient gain of knowledge by data reuse

- Right of first usage, possibility of embargo
- At the end of the day (resp., the embargo), all data are as open as possible
- Incentivize data sharing

Balancing values: Three fields of tension with human subject data

LMU Open Science Center

Reproducibility and verifiability of published analyses

Protect original authors against inadequate burden and potential attacks

- Primary focus: openness and transparency. Correcting errors is painful, but a necessary condition for doing science
- Data providers should be informed if their data are going to be reused or reanalyzed → allows to prepare a reaction

Balancing values: Three fields of tension with human subject data

OSC
LMU Open Science Center

Reproducibility and verifiability of published analyses

Protect original authors against inadequate burden and potential attacks

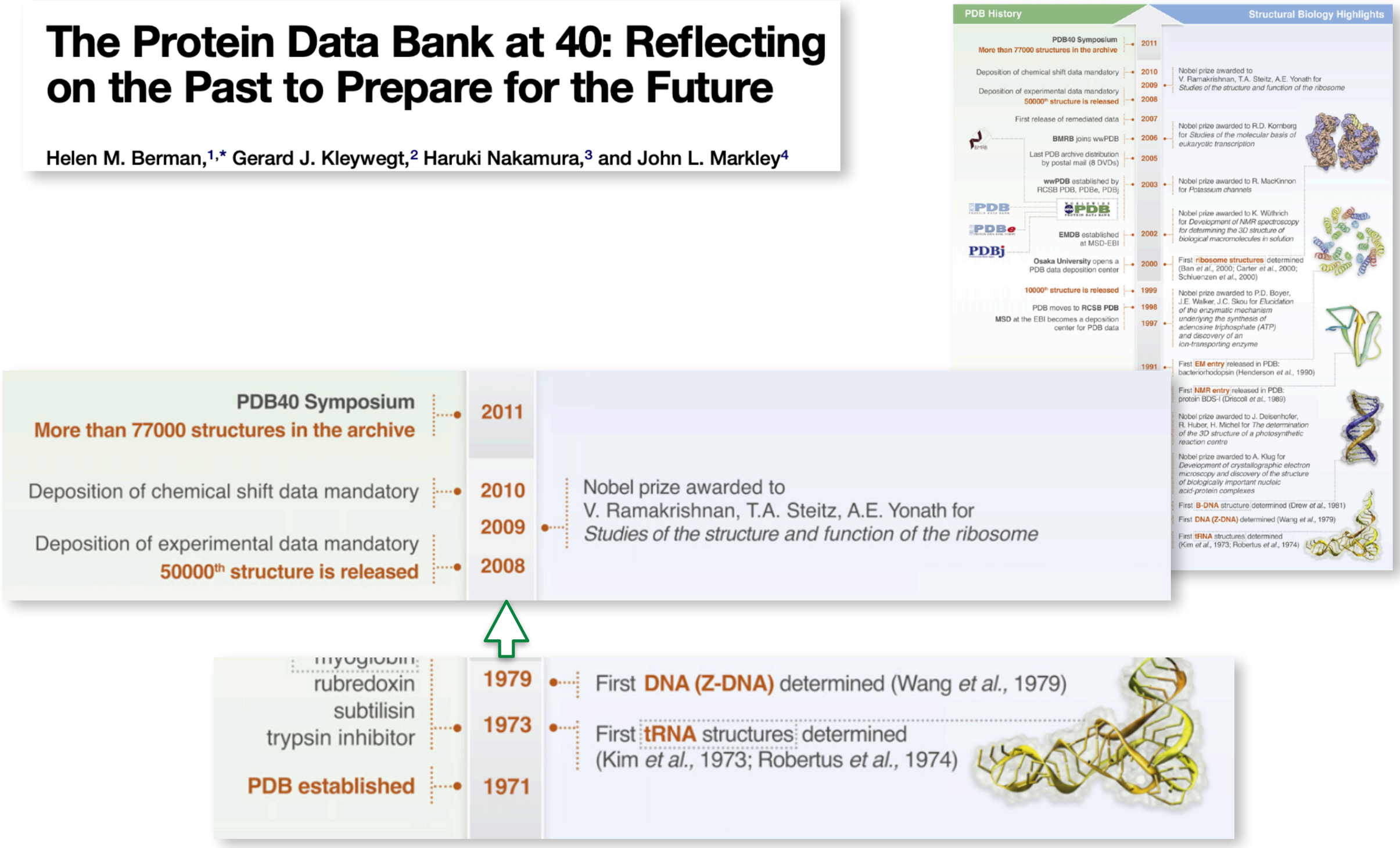
- Problematic asymmetry:
  - Data provided ➞ often errors get detected
  - No data provided ➞ no errors are detected (because not possible). Default assumption: „Everything is OK. Perfect paper, because no errors are spotted!“
- Making oneself vulnerable is good for science, and should also be good for reputation!
- Change default assumption?
  „No data ➞ Probably erroneous analysis.“

Success stories

OSC
LMU Open Science Center

The Protein Data Bank at 40: Reflecting on the Past to Prepare for the Future

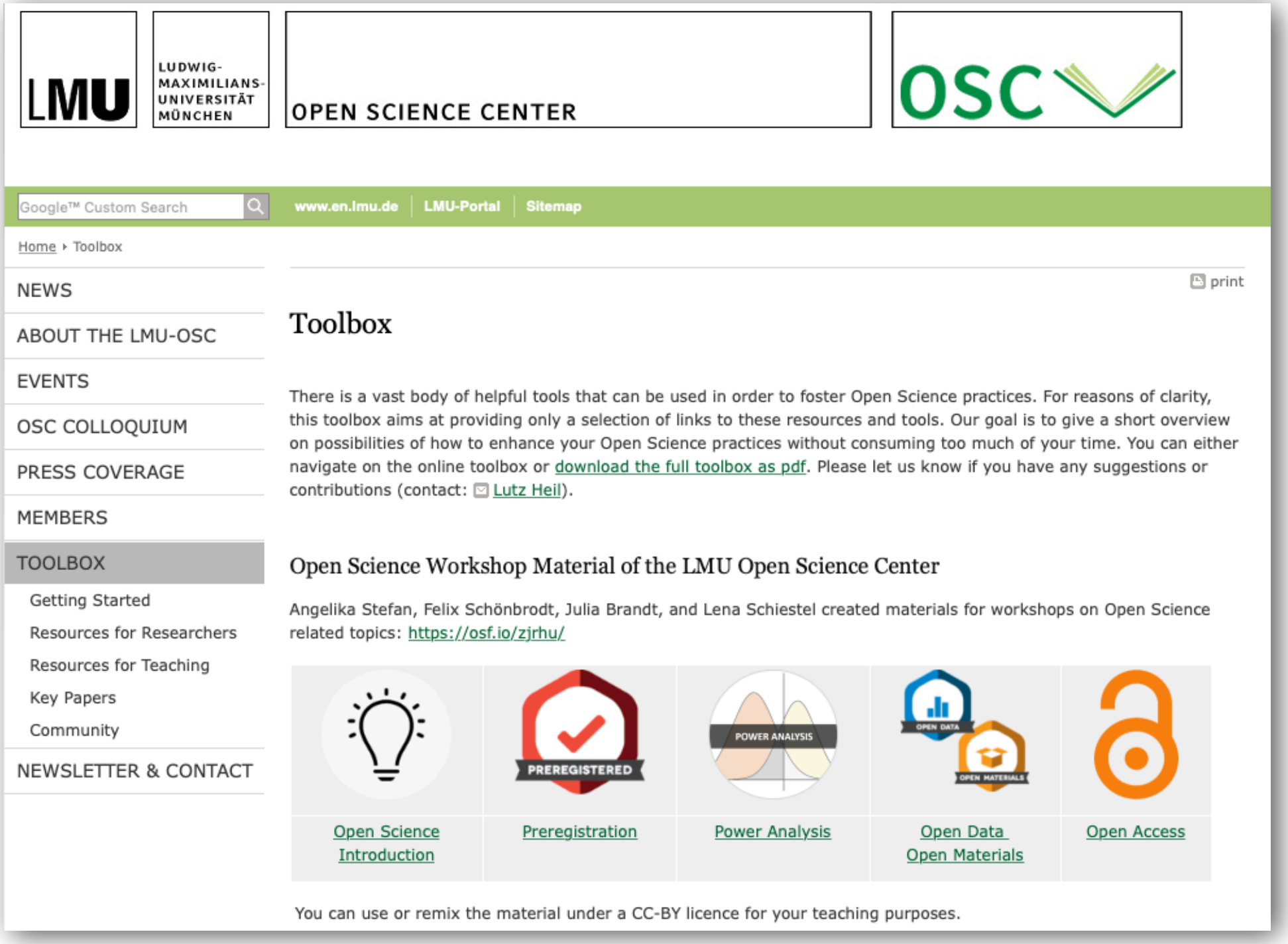
Helen M. Berman,[1,*] Gerard J. Kleywegt,[2] Haruki Nakamura,[3] and John L. Markley[4]



https://www.cell.com/action/showPdf?pii=S0969-2126%2812%2900018-4

Resources

OSC
LMU Open Science Center



https://www.osc.uni-muenchen.de/toolbox/index.html