



Transparency of CHI Research Artifacts: Results of a Self-Reported Survey

Chat Wacharamanotham¹, Lukas Eisenring¹, Steve Haroz², Florian Echtler³

¹University of Zurich, ²Université Paris-Saclay, Inria, ³Aalborg University (prev. BUW)



@chatchavan



@sharoz

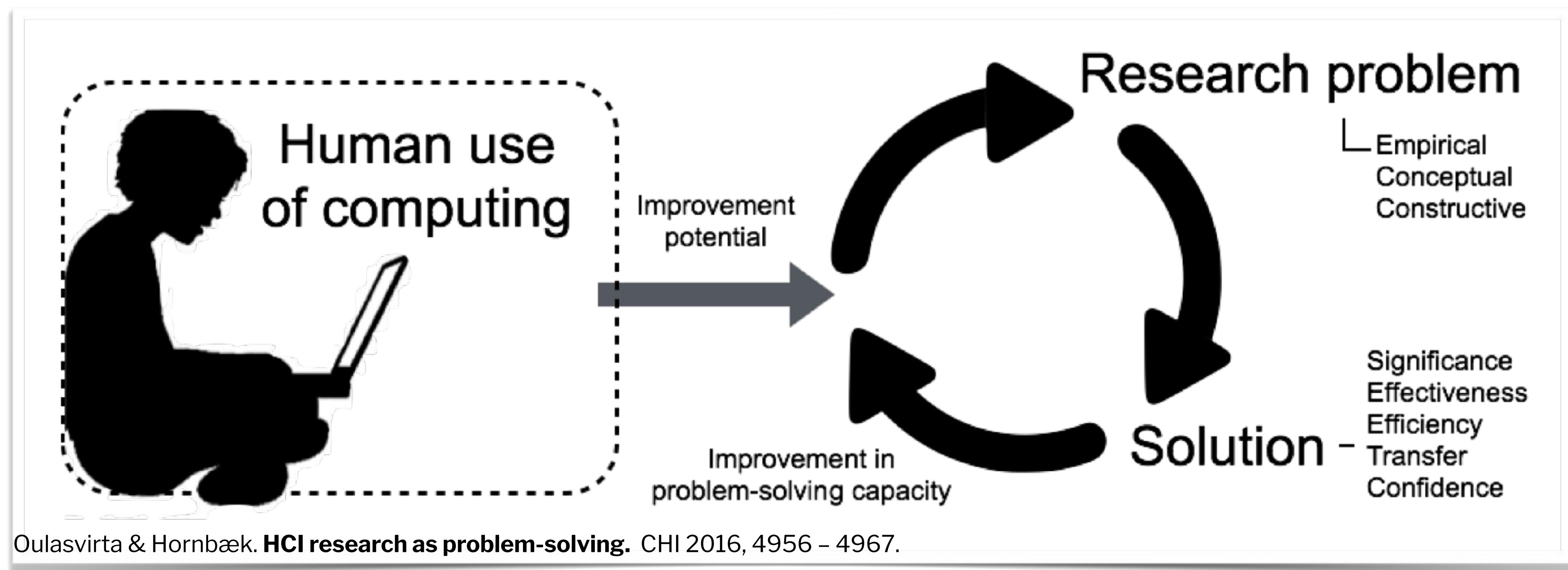


@floemuc



Why do you do HCI research?

To increase our capacity to solve problems related to human use of technologies



Google Scholar

Articles Case law

Stand on the shoulders of giants

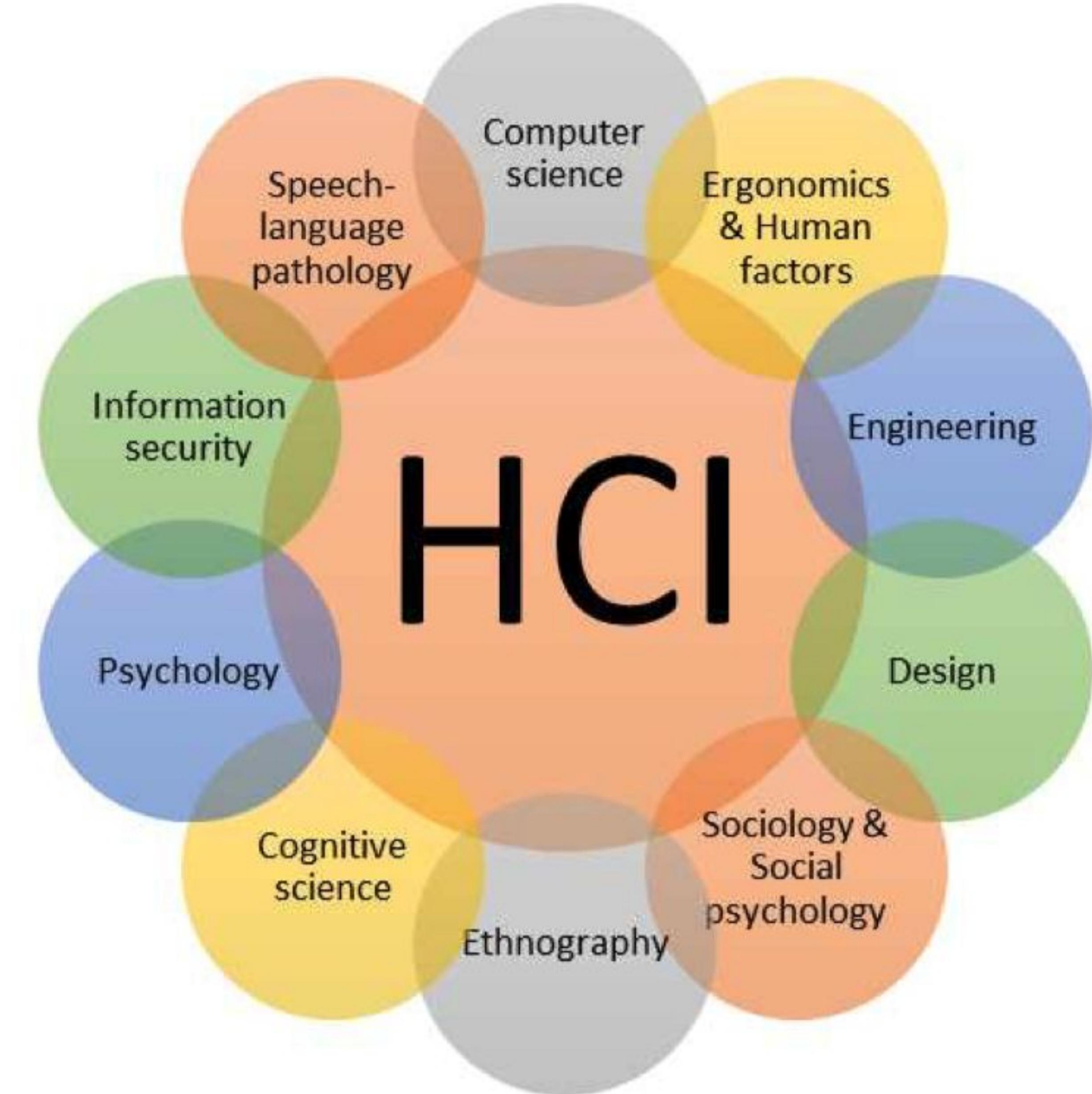


Image by Greg Splenka from <https://harvardmagazine.com/1997/01/freud.whole.html>



Why open science is critical to combatting COVID-19

“open science policies can remove obstacles to the free flow of research data and ideas, and thus accelerate the pace of research critical to combating the disease.”



HCI vs. Bioinformatics

	CHI 2016			CHI 2017			ISMB 2016	ISMB 2017
	Extended Abstracts	Papers/Notes	Total	Extended Abstracts	Papers/Notes	Total	Papers	Papers
Total	567	460	1027	458	599	1057	55	56
containing OS keywords	50	74	124	33	91	124	24	56
... supplemental material	2	22	24	4	36	40	-	-
releasing source code	12	10	22	6	21	27	20	49
Percentage	2.1%	2.0%	2.0%	1.3%	3.5%	2.6%	36.4%	87.5%

Table 1: Open-source in HCI and bioinformatics: mentions and releases

A HCI research project could generate...

- **Study materials:** Created by researchers, presented to participants to elicit their responses
- **Data:**
 - Raw data
 - Data processing procedure (including both software code and manual procedures)
 - Output from data processing
- **Prototypes:** Software or hardware

**“If researchers want to use
data or code from my
paper, they can contact me”**

Do you agree?

**“Putting code or data on GitHub
is a reliable way to share”**

Do you agree?

**“Data from study
participants should not be
shared publicly online”**

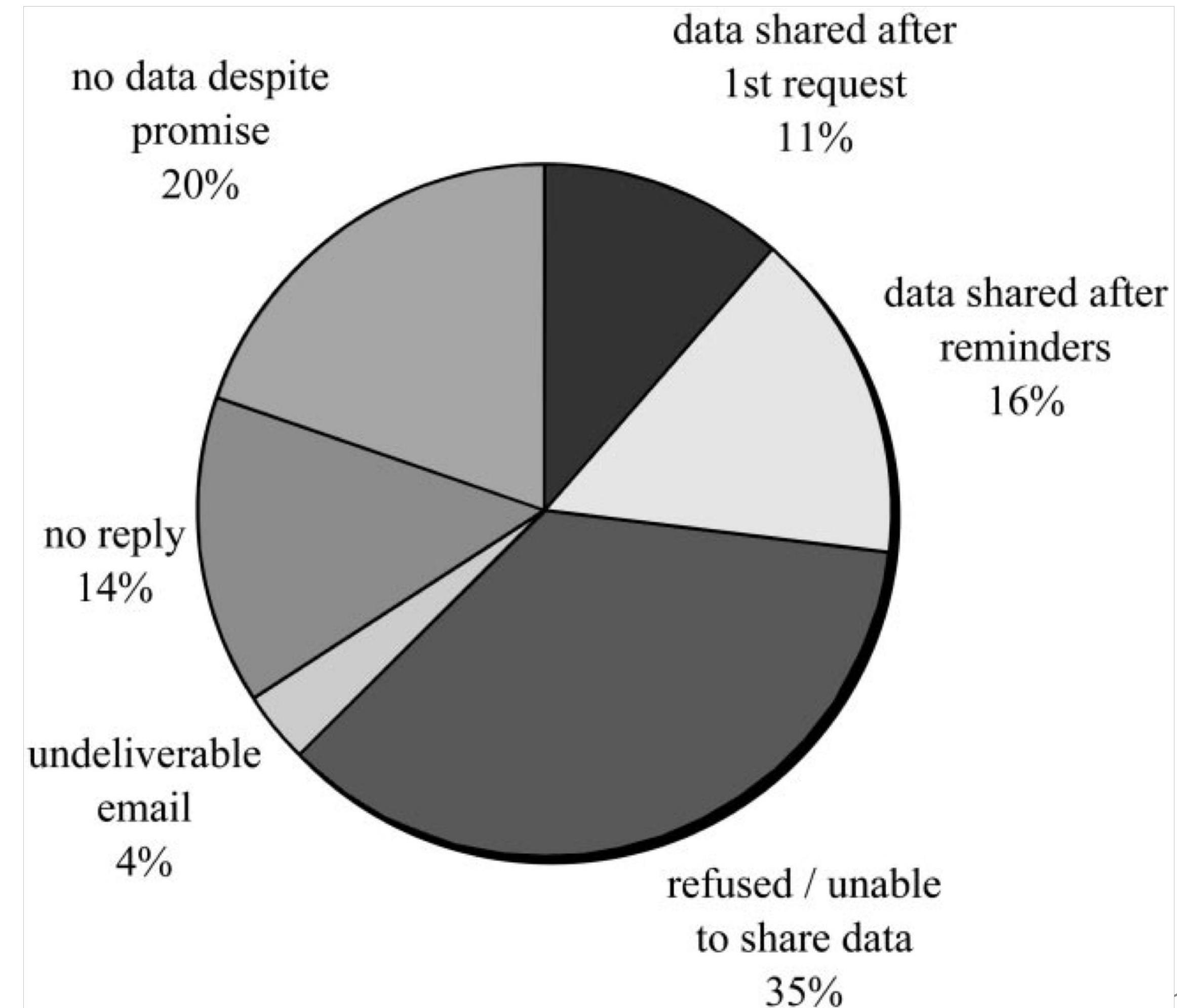
Do you agree?

“If researchers want to use data or code from my paper, they can contact me”

“If researchers want to use data or code from my paper, they can contact me”

A team of psychology researchers requested data from the authors of 141 articles published in prestigious psychology journals in the previous year.

27% success rate

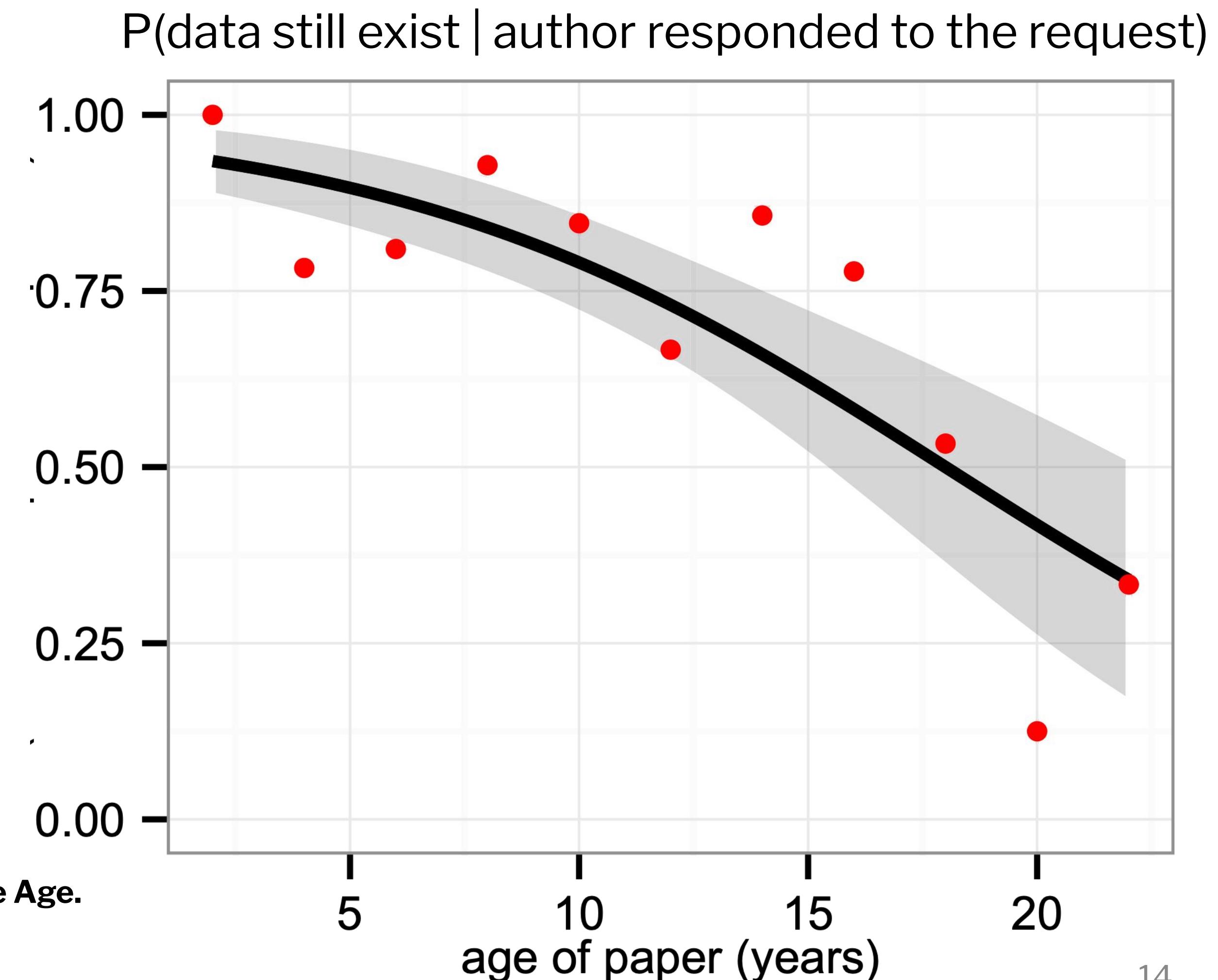


Wicherts et al. **The poor availability of psychological research data for reanalysis.** American Psychologist 61, 7 (2006), 726 – 728.

“If researchers want to use data or code from my paper, they can contact me”

Researchers from the field of biology requested data from 516 articles published between 2–20 years

The odds of data still exist fall 17% per year



Vines et al. **The Availability of Research Data Declines Rapidly with Article Age.**
Current Biology 24, 1 (2019), 94–97.

“Putting code or data on GitHub is a reliable way to share”

Danger Zone

Make this repository private
Hide this repository from the public.

Transfer ownership
Transfer this repository to another user or to an organization where you have the ability to create repositories.

Archive this repository
Mark this repository as archived and read-only.

Delete this repository
Once you delete a repository, there is no going back. Please be certain.

[Make private](#)[Transfer](#)[Archive this repository](#)[Delete this repository](#)

“Putting code or data on GitHub is a reliable way to share”

nature

Microsoft’s GitHub buyout raises fears

Users worry popular data-sharing site will become less open.

BY ANDREW SILVER

GitHub — a website that has become popular with scientists collaborating on research data and software — is to be acquired by Microsoft for US\$7.5 billion. In the wake of the takeover announcement

on 4 June, some scientists and programmers

Scholar cited the website which is free to use
for projects uses a version which tracks changes. This allows researchers to work on track changes. Although



TheScientist
EXPLORING LIFE, INSPIRING INNOVATION

Home / News & Opinion

Researchers React to Microsoft’s Acquisition of GitHub

Computational biologists are optimistic that the purchase of the world’s largest hub for open-source computer code will not affect the way they use GitHub for science.

<http://dx.doi.org/10.1038/d41586-018-05426-0>

<https://www.the-scientist.com/news-opinion/researchers-react-to-microsofts-acquisition-of-github-48397>

“Putting code or data on GitHub is a reliable way to share”

What if you run out of funding? What happens to my data?



COS established a \$250,000 preservation fund for hosted data in the event that COS had to curtail or close its offices. If activated, the preservation fund will preserve and maintain read access to hosted data. This fund is sufficient for 50+ years of read access hosting at present costs. COS will

<https://help.osf.io/hc/en-us/articles/360019737894-FAQs>



- **Replicas:** All data files are stored in CERN Data Centres, primarily Geneva, with replicas in Budapest. Data files are kept in multiple replicas in a distributed file system, which is backed up to tape on a nightly basis.
- **Retention period:** Items will be retained for the lifetime of the repository. This is currently the lifetime of the host laboratory CERN, which currently has an experimental programme defined for the next 20 years at least.
- **Succession plans:** In case of closure of the repository, best efforts will be made to integrate all content into suitable alternative institutional and/or subject based repositories.

<https://about.zenodo.org/policies/>

“Data from study participants should not be shared publicly online”

- Do you need to collect identifiable information?
- Could they be anonymized?
- Deposit in protected repositories
- In medicine, a systematic review of 25 qualitative studies found that “*in cases where participants perceived there to be actual or potential public benefits from research and had trust in the individuals or organisations conducting and/ or overseeing data linkage/sharing, they were generally supportive [to the sharing].*”

Aitken et al. **Public responses to the sharing and linkage of health data for research purposes: a systematic review and thematic synthesis of qualitative studies.** BMC Medical Ethics (2016) 17:73.

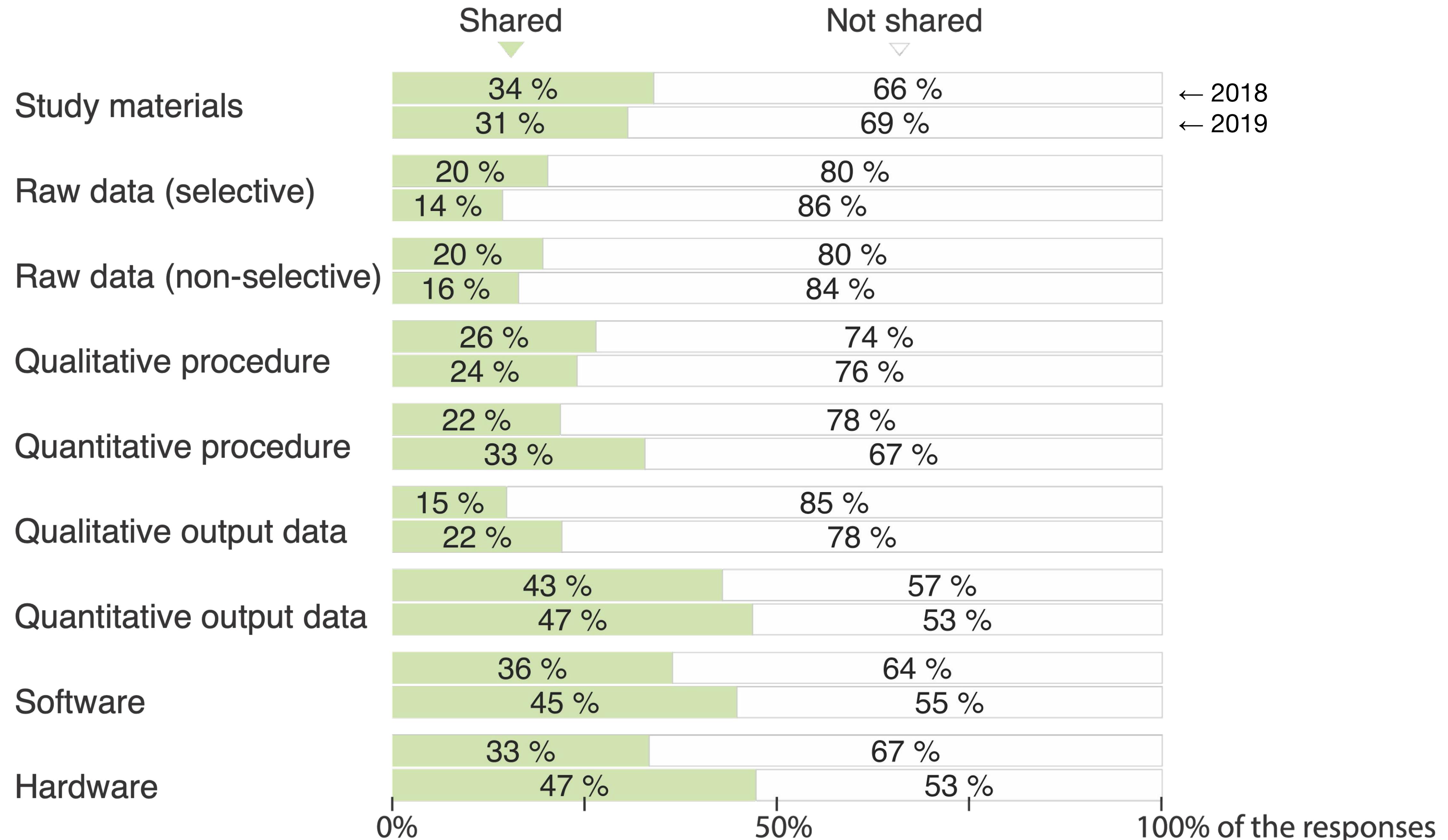
**How is the research
artifact sharing in HCI?**

We asked authors of CHI 2018–19

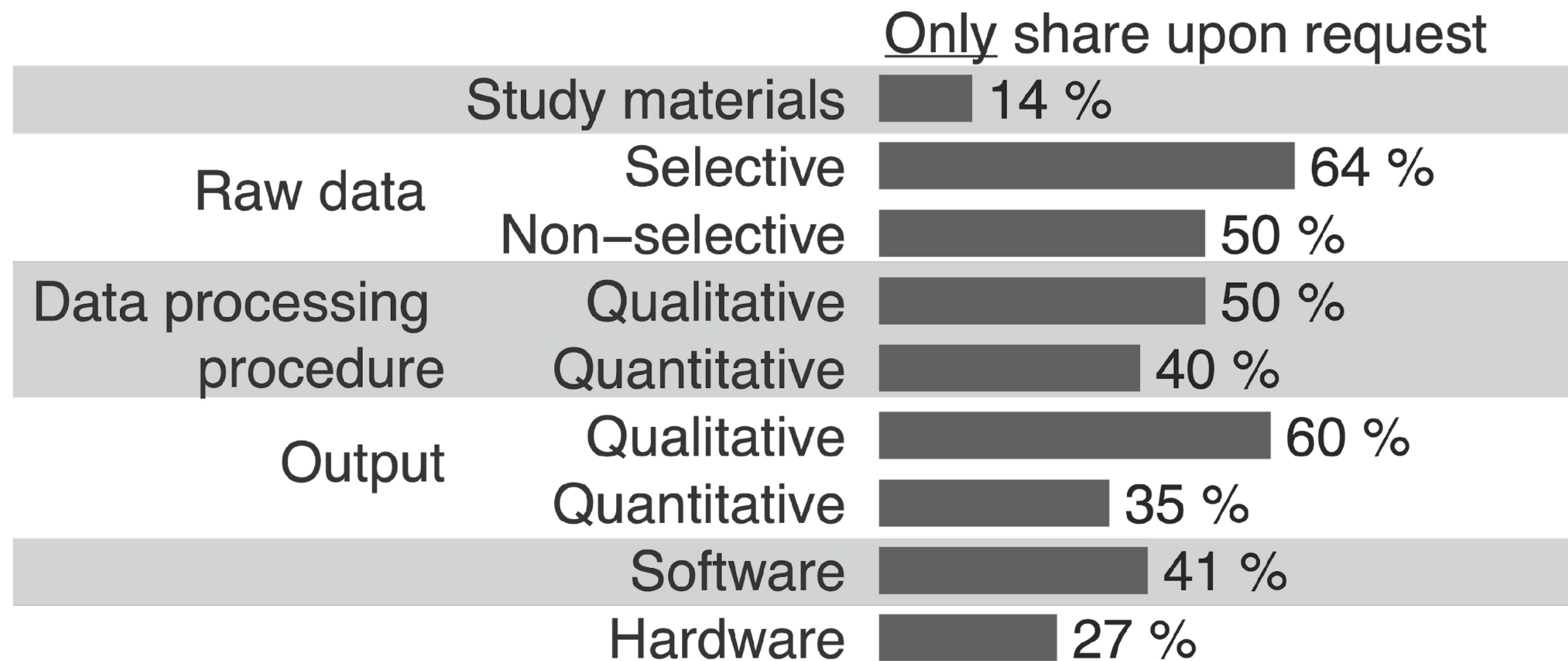
- What types of research artifacts they generate?
- If they share it, how?
- If not, why?

460 responses out of 1356 invitations

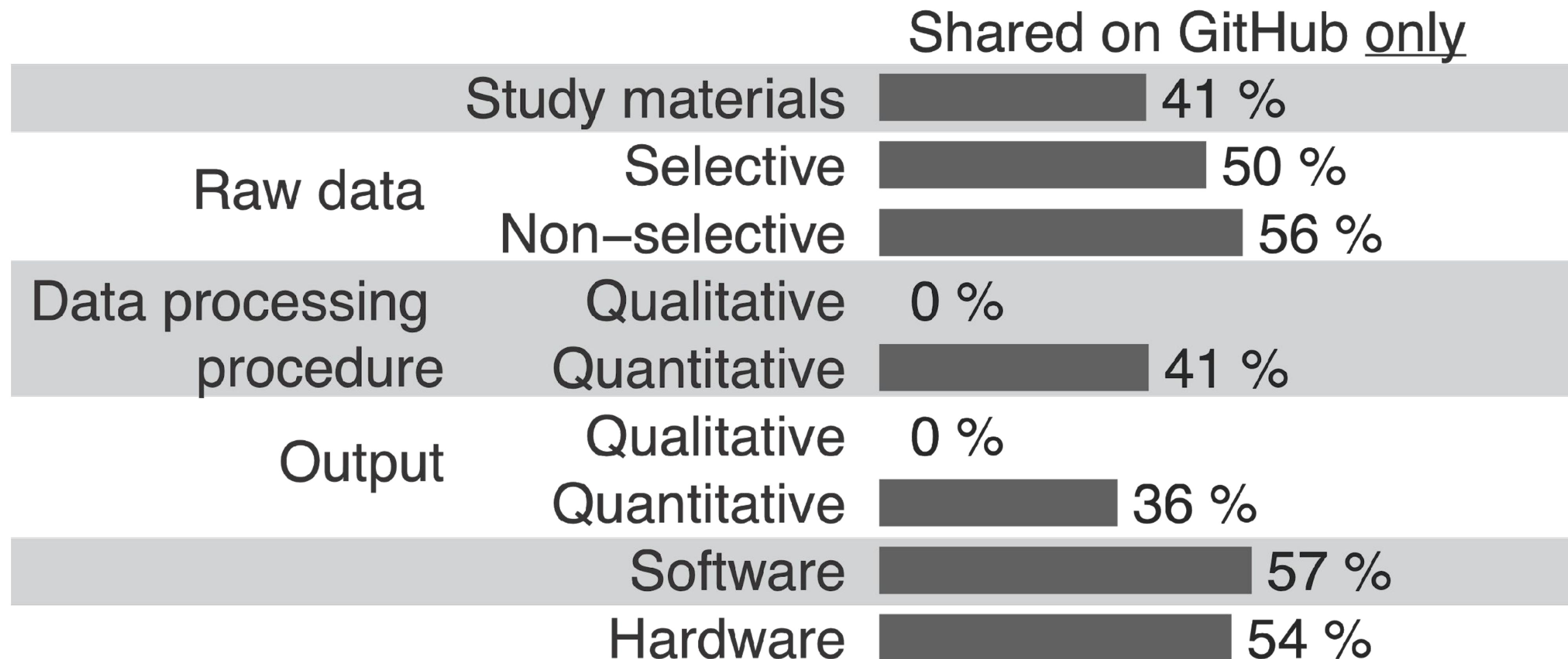
34% response rate! (quite high)



“If researchers want to use data or code from my paper, they can contact me”

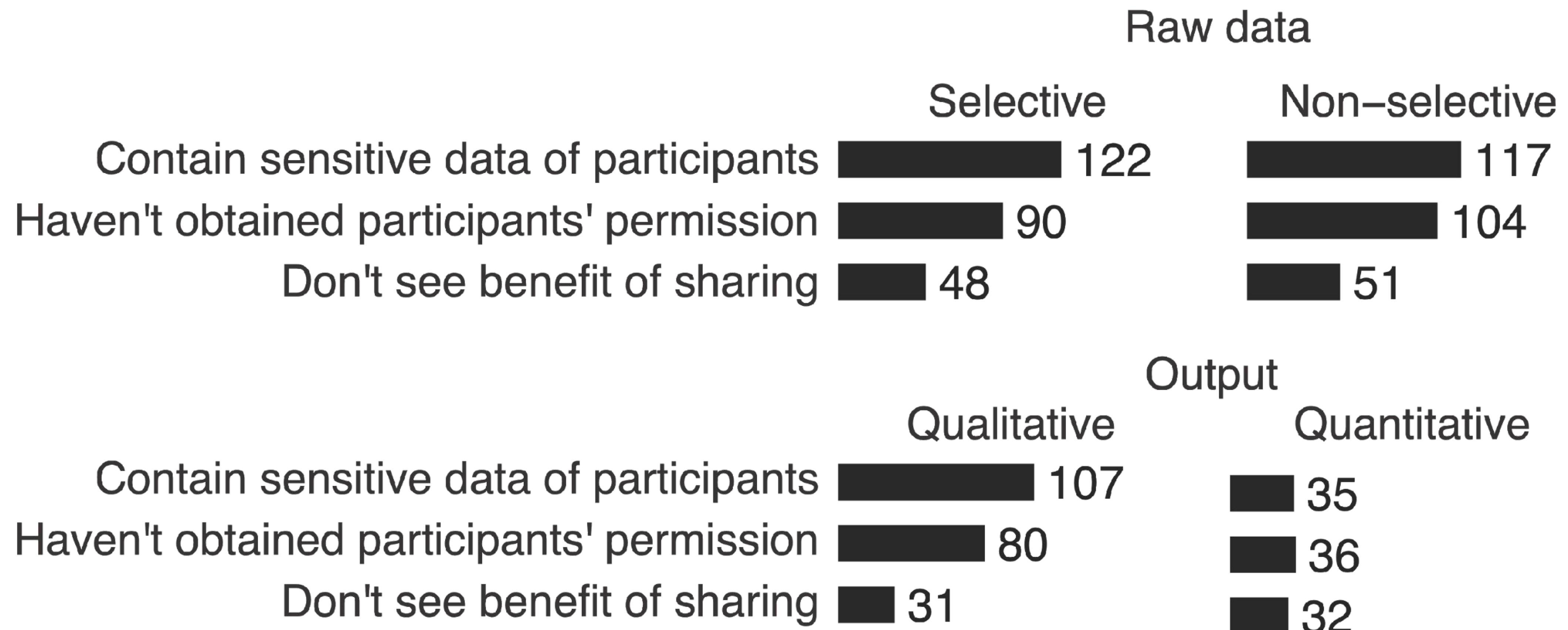


“Putting code or data on GitHub is a reliable way to share”



“Data from study participants should not be shared publicly online”

(Top three concerns about data sharing from the survey)



Sharing is good for you!

1€ filter: a simple speed-based low-pass **filter** for noisy input in interactive systems

[G Casiez, N Roussel, D Vogel](#) - ... of the SIGCHI Conference on Human ..., 2012 - dl.acm.org

The **1€ filter** (" one Euro **filter**") is a simple algorithm to **filter** noisy signals for high precision and responsiveness. It uses a first order low-pass **filter** with an adaptive cutoff frequency: at low speeds, a low cutoff stabilizes the signal by reducing jitter, but as speed increases, the ...

☆ 99 Cited by 145 Related articles ↗

1€ Implementations

- [Example how to check your 1€ filter implementation](#)
- [Python version](#) (BSD Licence)
- [C version](#) by [Jonathan Aceituno](#)
- [C++ version](#) (BSD Licence) by Nicolas Roussel
- [C++ version using templates](#) by [Jonathan Aceituno](#)
- [Java version](#) (BSD Licence) by [Stéphane Conversy](#) from [Université de Toulouse](#)
- [JavaScript version](#) by Florian Renaut
- [JavaScript version using class](#) (BSD Licence) by Géry Casiez
- [C# version](#) by [Mitsuru Furuta](#) from [Sensorit > more code](#)
- [F# version](#) by [Mitsuru Furuta](#) from [Sensorit > more code](#)
- [Objective C version](#) by [Stefan Ivanov](#) from [RWTH Aachen University](#)
- [Non object-oriented version](#) by [Stefan Ivanov](#) from [RWTH Aachen University](#)

Sharing is caring, if shared with care

- **Plan early**
- **Prepare how to store research artifacts**
- **Collect only necessary data**
- When sharing, weigh the trade-off between **transparency, privacy, effort, and rawness**
- Share at a **FAIR location**
 - **Findable, Accessible, Immutable, Reusable**
 - If sharing is not possible, describe the reasons

(See page 8-10 of our paper for details and pointers to resources that help you plan and make decisions)

Sharing research artifacts is a part of research craftsmanship



The field of HCl is in a unique position

- We produce diverse types of research materials
- We have expertise in understanding work practices and context
- We can build tools, especially ones that are enjoyable to use
- HCl can improve scientific practice as a whole!

Transparency of CHI Research Artifacts: Results of a Self-Reported Survey

~~“If researchers want to use data or code from my paper, they can contact me”~~

You should deposit research artifacts on open or protected repositories.

~~“Putting code or data on GitHub is a reliable way to share”~~

In addition to GitHub, put snapshots at the publication time on OSF or Zenodo

~~“Data from study participants should not be shared publicly online”~~

Collect only the data the research project needs, anonymize them, and share them at a FAIR location

Slides and paper: <http://tiny.cc/chi-sharing>



Chat Wacharamanotham¹, Lukas Eisenring¹, Steve Haroz², Florian Echtler³

¹University of Zurich. ²Université Paris-Saclay, Inria, ³Bauhaus-Universität Weimar

