Data sharing

Christian Behrends
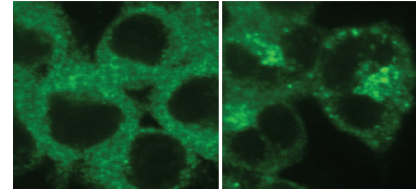
What kind of data do we have and share?
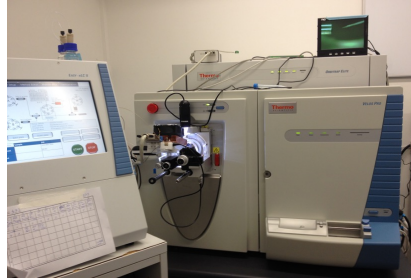
Proteomics
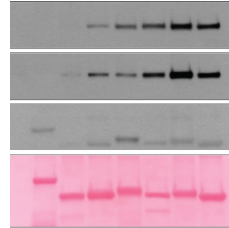
Biochemistry

Cell biology

ARTICLES

Network organization of the human autophagy system

Christian Behrends[1], Mathew E. Sowa[1], Steven P. Gygi[2] & J. Wade Harper[1]

- **Mass spec raw files -> PRIDE (EMBL), MassIVE (NIH funded)**

- Sequencing raw files ->GEO (NIH-funded)

- **Microscope images -> Mendeley data, Journals**
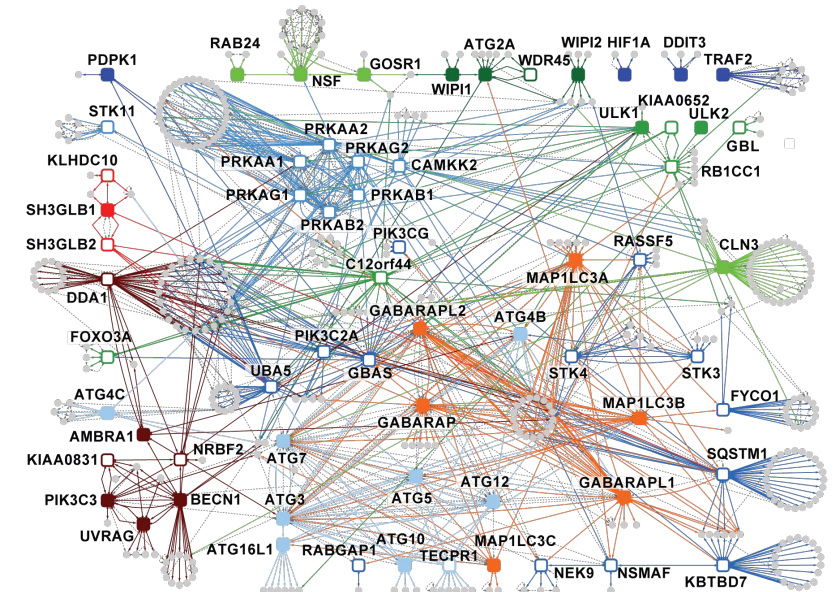
- **Immunoblots -> Mendeley data, Journals**

Data sharing – benefits

- **Citations**

- **Carrer recognition, visibility**

- **New collaboration**

- **Validation of data in research papers (reproducibility)**

- **Easier to find useful data**

- **Data archived and preserved for the future (data disaster)**

- **New research made possible**

- **More efficient research (data duplication)**

- **Re-use in teaching**
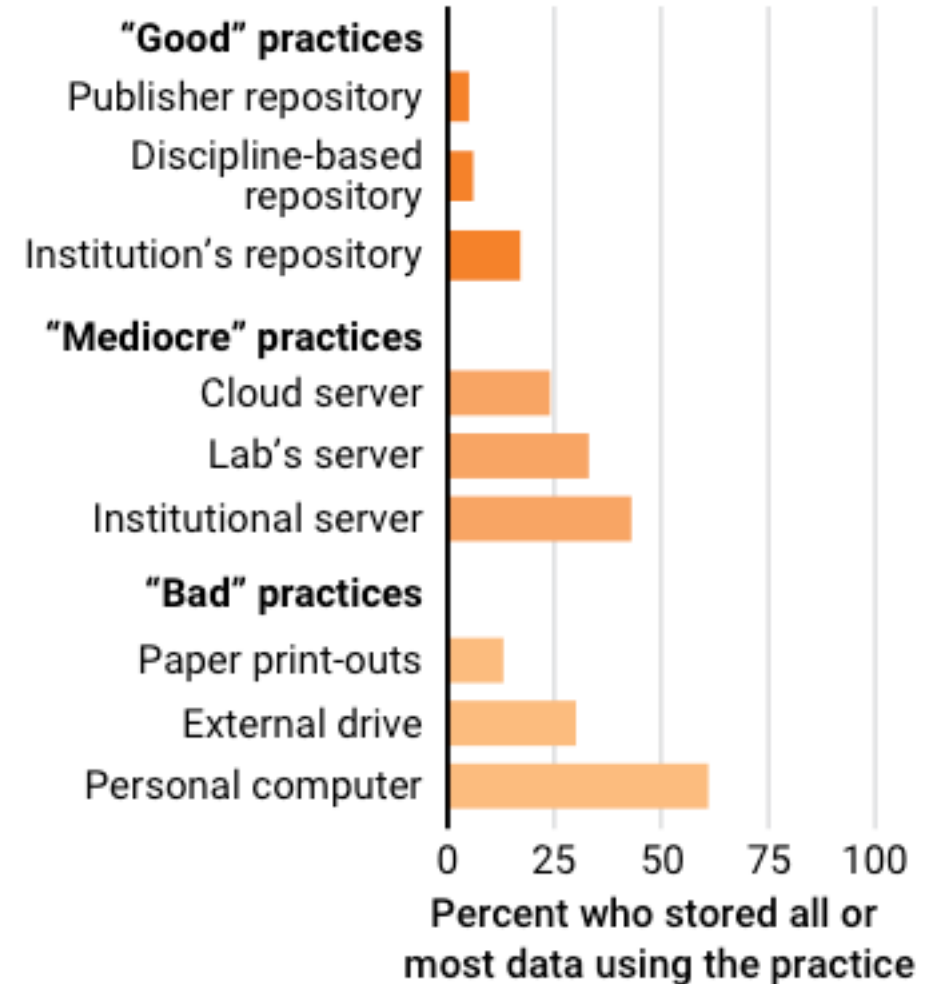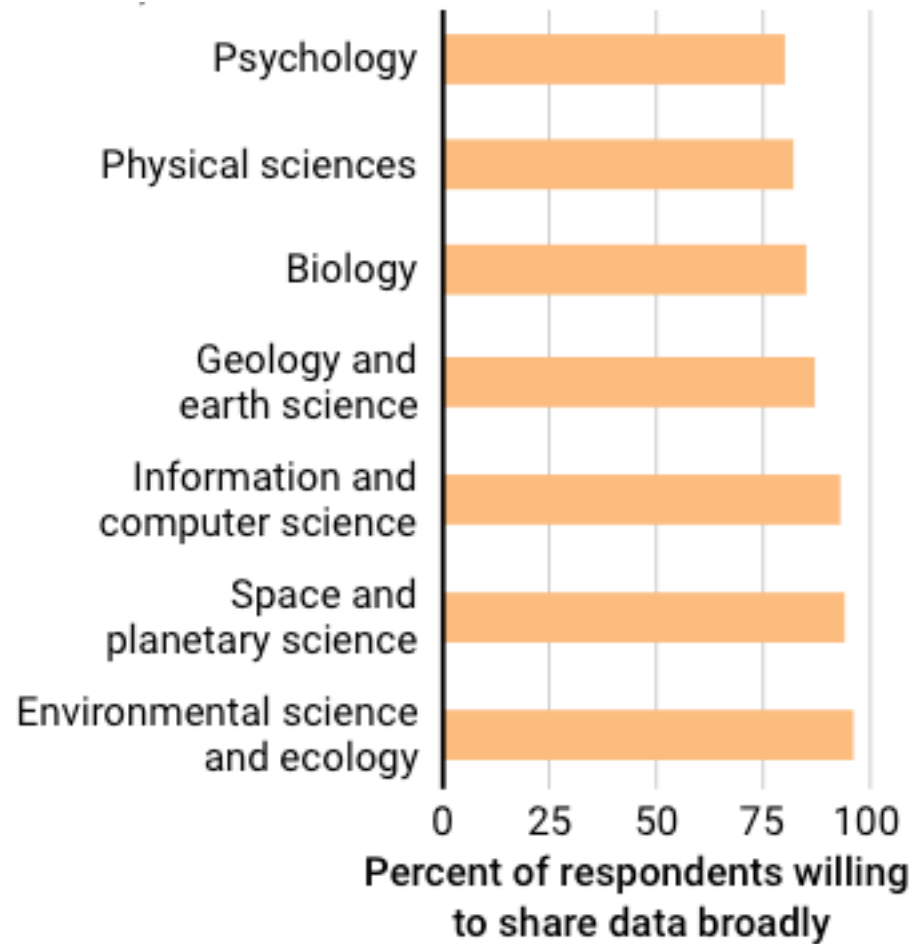
- **Compliance to funding body, publisher, university**

Benefits for sharing PI

Benefits for the research community

Public benefits

Data sharing – intentions and reality

Psychology
Physical sciences
Biology
Geology and earth science
Information and computer science
Space and planetary science
Environmental science and ecology

Percent of respondents willing to share data broadly

"Good" practices
Publisher repository
Discipline-based repository
Institution's repository

"Mediocre" practices
Cloud server
Lab's server
Institutional server

"Bad" practices
Paper print-outs
External drive
Personal computer

Percent who stored all or most data using the practice

Data sharing – challenges

- **Time and effort:** Making data sharable is labor intensive (e.g. organization, metadata, file naming, look-up tables, …)

- **Lack of incentives for sharing data:** current academic culture lacks the incentive structure to properly reward researchers who share their data.

- **Loss of control over data:** Scoop, errors in the data, contradictory conclusions, misuse

- **Personal data:** Anonymizing/pseudonymizing not always easy. Trade off with utility.

- **Restrictions:** Privacy, confidential or sensitive data, third party data, copyright, commercial valorization potential

- **Incomplete or poor sharing:** Research articles miss links to depositories

- **Publisher:** Journals struggle enforcing sharing requirements, time-consuming examination, no extra revenue

Data sharing – how to share

- **Mail/transfer data upon request (avoid)**

- **Make data available via a personal or project website**

- **Provide data as supplementary materials to a journal article**

- **Share data via a data repository/data archive (Preferred)**

Data sharing – data depositories

Advantages:

● No handling of data reuse queries and managing data acces

● Guarantees in terms of sustainable access to data

● Make your data discoverable and citable

● Make your data FAIR

● publish dataset using a persistent identifier (e.g. DOI)

● service of long-term preservation

● repositories may host data from various disciplines (e.g. Mendeley Data) or are discipline specific (e.g. PRIDE)

Data sharing

Questions?