Restricted but not forgotten: Rethinking restricted data

October 28th, 2025

Anneliese Eber, MSc, MLIS
Research Data Management Librarian
aeber@uwaterloo.ca

Alyssa Roussy
MLIS Co-op Student, Research Data Management
aroussy4@uwo.ca

UNIVERSITY OF WATERLOO | Library

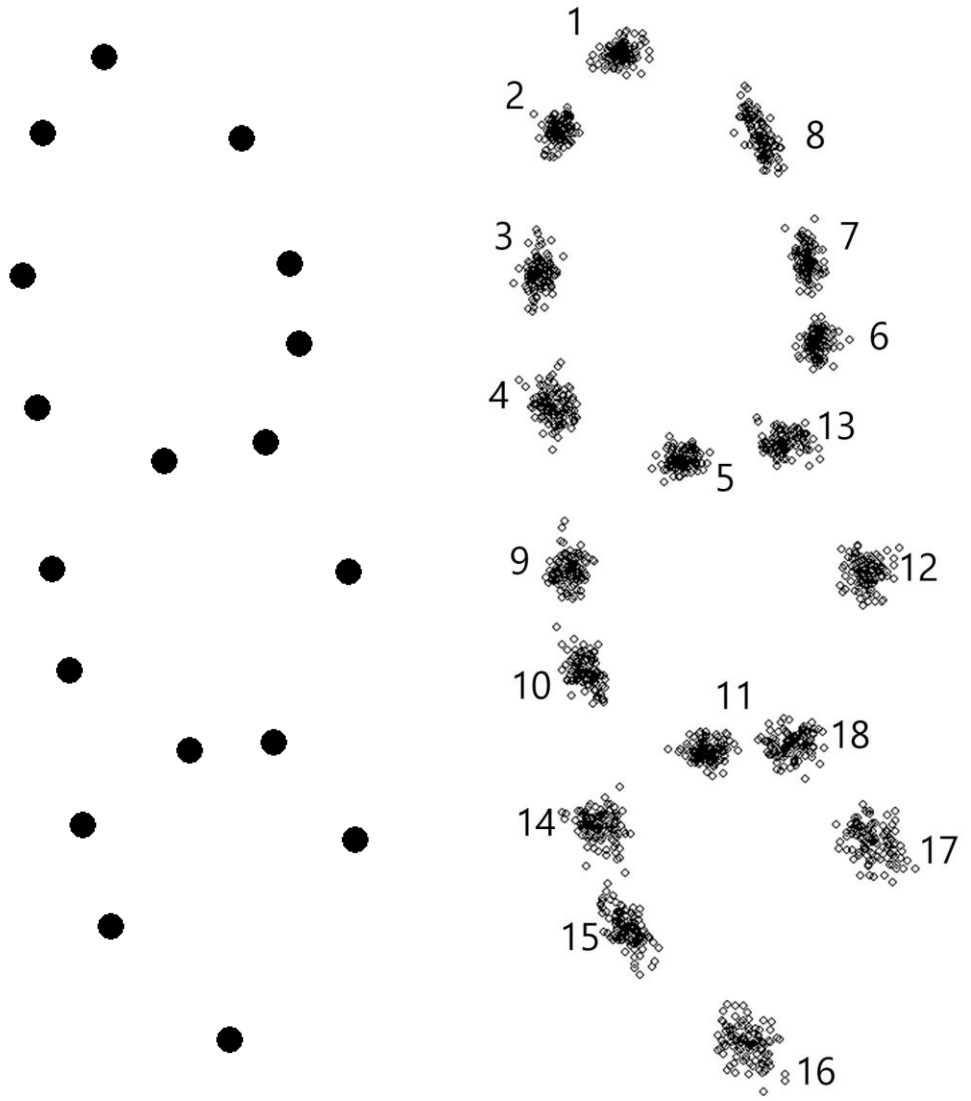
Opening Doors To Data
Canadian Research Data Centre Network
Connection 2025
Toronto

Introduction

- Research Data Management Librarian
- Background in anthropology & archaeology
- Research interests in addressing challenges & barriers to managing data in biological anthropology

Agenda

- What's the point?
- Rethinking restricted data
  - What is it?
  - How does it become restricted?
  - Case study
- Working with restricted data
- Where do we go from here: Changing the narrative



UNIVERSITY OF WATERLOO | Library

WHAT'S THE POINT?

What's the point? Why do we share data?

Responsible Research	Integrity
	Reproducibility
	Replicability
	Transparency
	Credibility
	Accountability

When we don't share data

Analysis of scientific paper retractions due to data problems: Revealing challenges and countermeasures in data management

The road to fraud starts with a single step

Millions of research papers at risk of disappearing from the Internet

An analysis of DOIs suggests that digital preservation is not keeping up with burgeoning scholarly knowledge.

More than 10,000 research papers were retracted in 2023 – a new record

The number of articles being retracted rose sharply this year. Integrity experts say that this is only the tip of the iceberg.

Medicine is plagued by untrustworthy clinical trials. How many studies are faked or flawed?

Investigations suggest that, in some fields, at least one-quarter of clinical trials might be problematic or even entirely made up, warn some researchers. They urge stronger scrutiny.

The Prevalence of Inappropriate Image Duplication in Biomedical Research Publications

An Unsettling Hint at How Much Fraud Could Exist in Science

Alberta medical dean accused of plagiarism resigns

'Avalanche' of spider-paper retractions shakes behavioural-ecology community

Allegations of fabricated data have prompted a university investigation and some soul-searching.

Biomedical paper retractions have quadrupled in 20 years – why?

Unreliable data, falsification and other issues related to misconduct are driving a growing proportion of retractions.

Disgraced McMaster scientist Jonathan Pruitt publishes fantasy novel in aftermath of research scandal

Investigation found Jonathan Pruitt 'engaged in data falsification and fabrication in several papers' while based at U.S. universities.

Facebook did not adequately protect user data in Cambridge Analytica scandal, FCA rules

Opening the black box of article retractions: exploring the causes and consequences of data management errors

Big picture problems

Leads to questionable credibility

All estimations of replicability are sub-optimal

Negative impact on health outcomes & public health

Erodes trust in science & research

Responsible(?) research in 2025

- Responsible & transparent research is more important than ever and *must* be:
  - Data-driven ***not*** belief-driven
  - Evidence-based
  - Transparent
  - Reproducible
- Correlation ≠ causation
- Data –> claims ***not*** claims –> data

> "It is not proof. We're doing the studies to make the proof."
>
> *RFK Jr.*
> *October 9th, 2025*

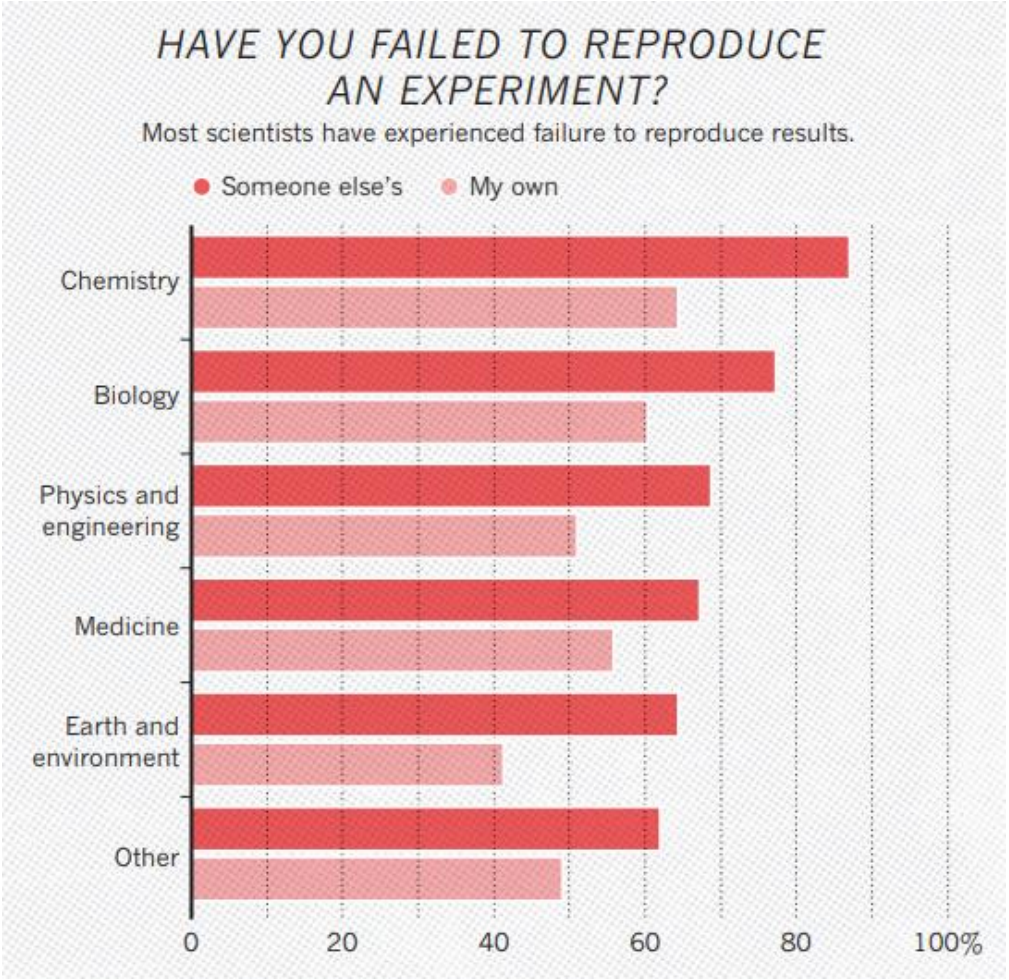
Responsible research

- Upholds ethical standards
- Minimizes potential harm(s)
- Considers wider impact of work on:
  - Society
  - Environment
  - Community
- Is transparent & reproducible
- Policy compliance
- Responsible data collection, analysis, storage, and sharing (where possible)

How did we get here?

Individual level

- Questionable research practices
  - Excluding outliers
  - Not reporting unsatisfactory results
  - Not reporting negative results
- Lack of access to methodological information & research data
- Mismanagement of data[sets]
- Cognitive bias



Baker, 2016

How did we get here?

Structural level

- Misaligned incentives
  - Tenure & promotion programs
- Publish or perish
- Publication metrics
- Prioritization of quantity over quality
- Better reception of hypothesis-supporting research
  - Less likely to report negative results
  - Journals resist publishing null results or replication studies
    - Rewards for novel research (nature pubs, etc.)
- Publishers
  - “lack of contribution”
  - Leads to publication bias

Challenges to sharing data

Data sovereignty & ownership

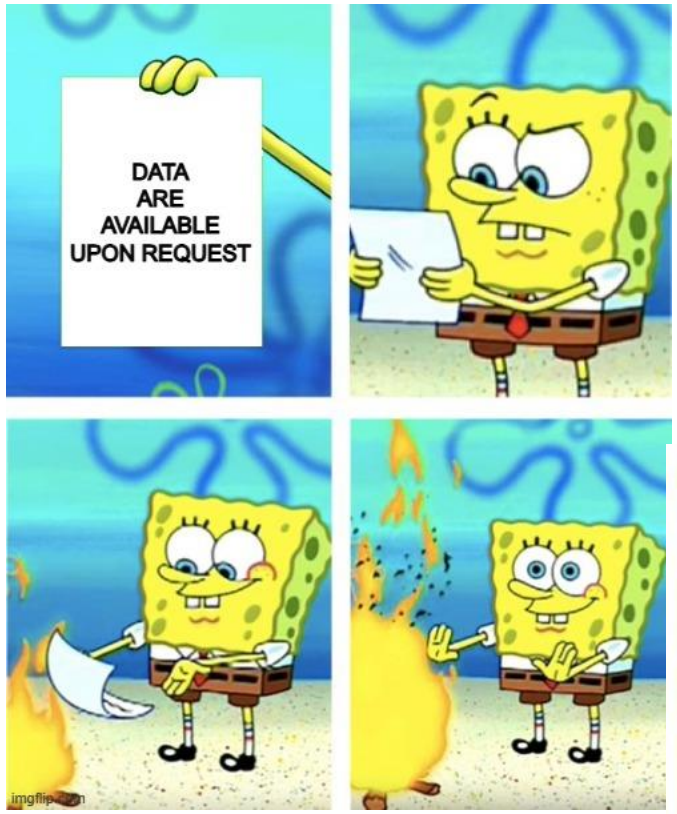
Ethical considerations

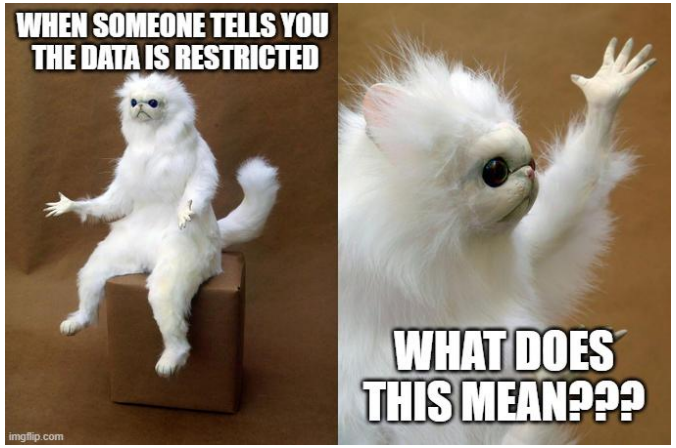
Legal obligations

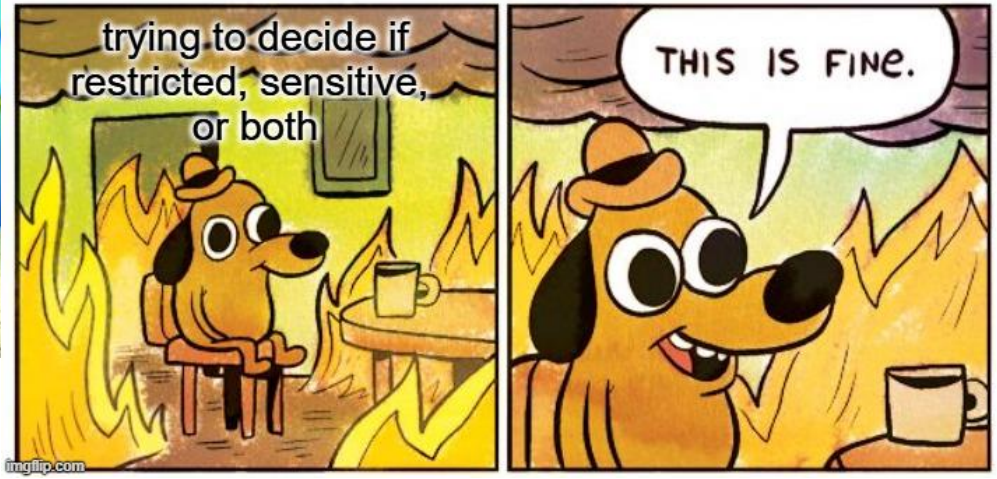
Commercial reasons

RETHINKING RESTRICTED DATA

What are restricted data?







Defining the unknown

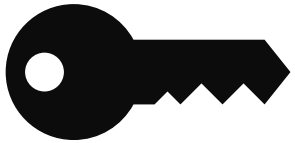
"data that are not immediately accessible because they are restricted due to commercial, ethical, or legal reasons, or they are only available upon request" (Read et al., 2024)

"data that cannot be released directly to the public research community due to possible risk(s) to study participants as well as the confidentiality promised to them" (National Institutes of Health)

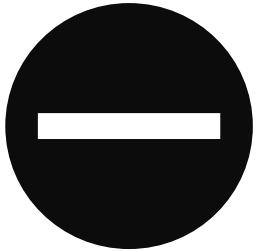
"Restricted access means that there is openly accessible information on how to obtain the data, but that the data are sensitive in some way, so access to them must be requested" (Swedish National Data Service)

Restricted what? What's the difference?

Restricted data



Restricted use



Restricted access

Why are things restricted?

- Confidentiality
- Varying privacy legislation
- Copyrighted material
- Collaborative research
- Embargo
- Career built on data
- Scrutiny from peers
- Unfamiliar with metadata
- Lack of shared metadata standards
- Potential data misuse
- Perceived limited value of data
- Lack of funding

Unique challenges

- Complex & time-consuming approvals
  - Difficult to navigate request processes
- Resource intensive to access
- Lack sufficient documentation (more than usual)
- Often poor quality
- Sparse or non-existent metadata limit discovery
- Inadequate infrastructure for data sharing
- Can’t use what you don’t know is there

Restricted data: Limited to health contexts?

- Restricted data appears across many disciplines
- Disciplines may have different reasons for restricting data
- Inter-disciplinary research may add additional challenges
- Extractive research appears in many disciplines
- Historically extractive disciplines may have additional considerations

Sharing fossil data: A case study

- Fossils are unique—literally one of a kind
- Often considered cultural heritage
- Subject to modern geopolitics based on modern borders
- Extremely fragile to work with
- Researchers often build their careers on a small set of data
  - Single site
  - Single fossil or set of fossils

Brown et al., 2004

Sharing fossil data: Sociocultural challenges

- Anthropology & archaeology have a deeply colonial history
  - Historically extractive research & well-known for it
- Research often done by Western researchers and excludes local communities and researchers
- Artifacts & fossil remains often leave their country of origin

Sharing fossil data: Physical challenges

- Fossils don't travel well, if at all
- Extremely expensive to transport fossils
  - Securely packing & air travel
  - Research & import permits
  - Storage
- Extremely expensive to travel to fossils
  - Research permits
  - Travelling with [expensive] equipment

Masao et al., 2016

Sharing fossil data: Questions of ownership

- Found within a country (e.g., Kenya)
- Found by a researcher (likely not from the same country)
  - Finders keepers this is not
- May represent human history & lineage

- Who 'owns' history? Can someone 'own' something that old?
- What about scans or replicas of fossils?
  - Who owns those?
  - Ownership vs. authorship?

Dispute Over Java Man Raises a Question: Who Owns Prehistory?

An Indonesian claim against a Dutch museum has pushed the debate about restitution into the realm of the natural history museum — where it hasn't been much of an issue until now.

Siegal, 2022

Science

The Netherlands returns 'Java Man,' thousands of colonial-era fossils taken from Indonesia

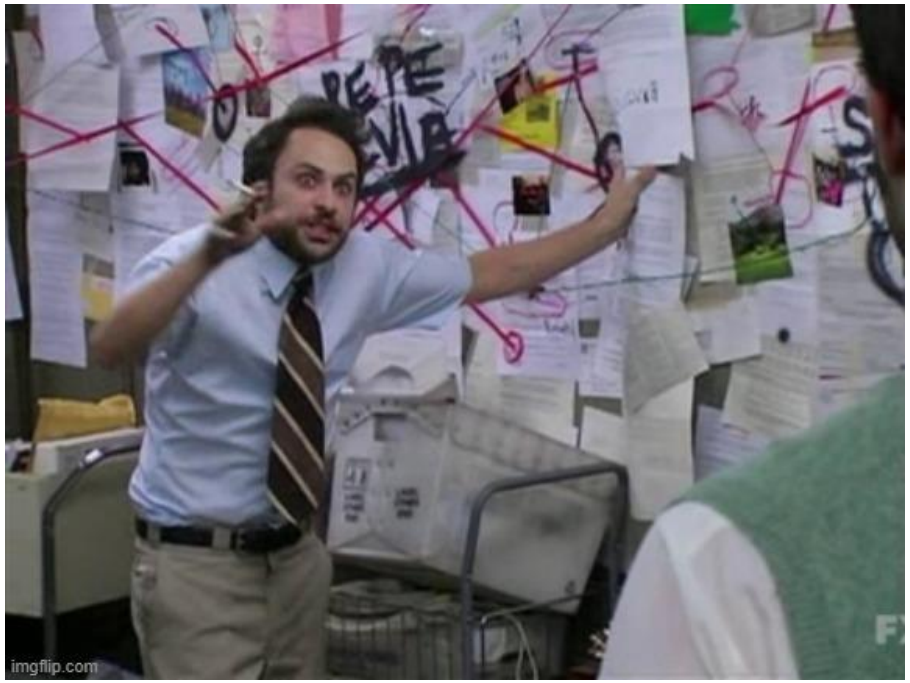
More than 28,000 fossils to be returned to Indonesia that were taken when it was a Dutch colony

The Associated Press · Posted: Sep 26, 2025 3:49 PM EDT | Last Updated: September 26

Associated Press

Sharing fossil data: Complex to navigate

- Complications add up on top of one another
- Eventually it becomes easier to just call the data restricted
- No single solution
- Continually evolving policies
- “Data available upon request”: Maybe…
- Shifts in research culture



Sharing fossil data: Data sovereignty

Whose data are they? Whose data should they be?

Museum & institutional data agreements

Ambiguous authorship & ownership of data

Colonial history of data collection

Trying and [sort of] succeeding

IMAGE ACQUISITION AND PROCESSING IN DETAIL

STEP 1: IMAGE ACQUISITION

Physical object usnm:anthropology-tc:621R was imaged to create undeposited raw media.

IMAGING DETAILS Field Definitions

Modality	Laser Scan
Device	NextEngine 3D Laser Scanner
Device facility	Integrative Anatomical Sciences (University of Southern California)
Creator	--
Event date	--
Software	--
Description	--
Description Attachment	None
Photographic Reference	None

OWNERSHIP AND PERMISSIONS

Field Definitions

Data managed by	Biren Patel
Data uploaded by	Biren Patel
Publication status	Restricted Download
Download reviewer	Biren Patel
IP holder	United States National Museum
Copyright statement	In Copyright
Creative Commons license	--
Morphosource use agreement type	Standard
Permits commercial use	Commercial Use Not Permitted
Permits 3D use	3D Printing Limited
Required archival of published derivatives	On Any Repository
Funding attribution	National Science Foundation [BCS 1317047 (to Biren Patel), BCS 1539741 (to Caley Orr)], the L.S.B. Leakey Foundation (to Biren Patel), the Wenner-Gren Foundation (to Caley Orr), and the University of Southern California (to Biren Patel).
Publisher	--

ark:/87602/m4/M56923

WORKING WITH RESTRICTED DATA

Research data lifecycle

NFDI4Microbiota Knowledge Base, 2024

FAIR data

F A I R

Findable

Accessible

Interoperable

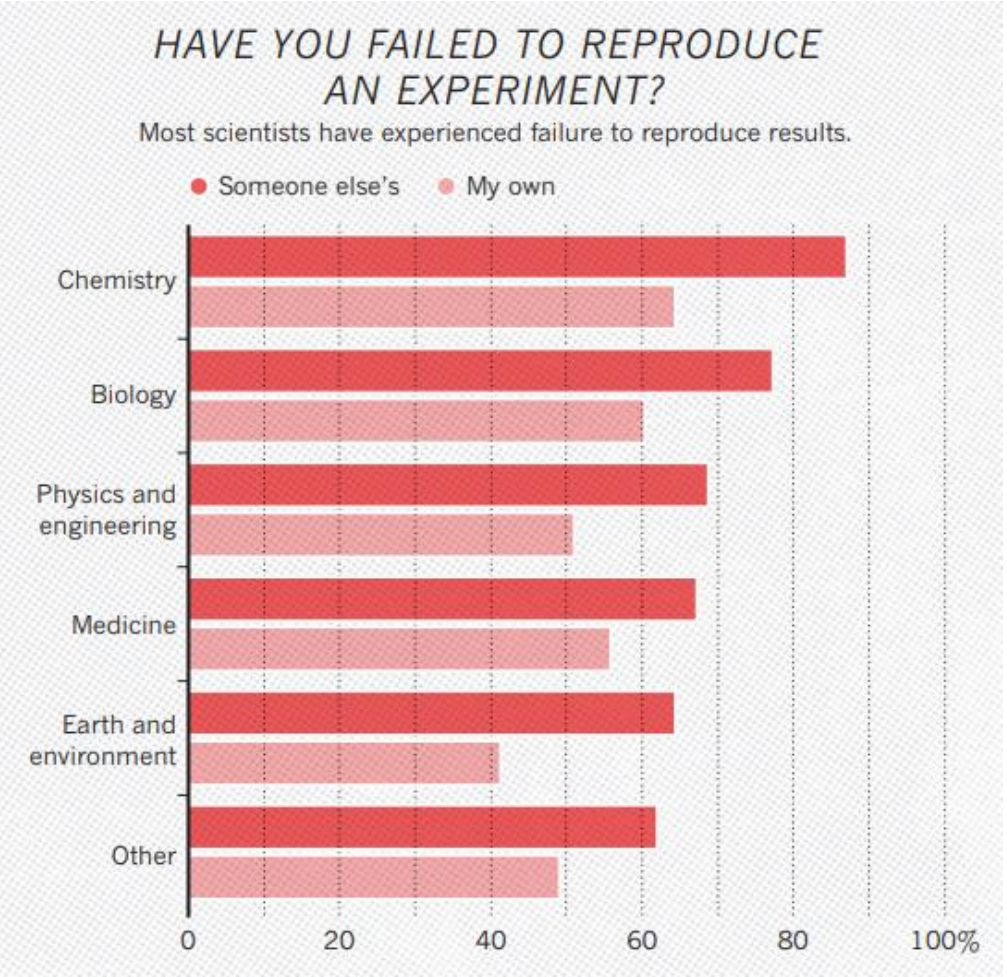
Reusable

Present- and future-proofing data

Findable	Accessible	Interoperable	Reusable
Humans and computers should be able to easily find **metadata** and data	Once the user finds the data, they need to know how to access them and may require details around authentication and authorization	Data usually need to be integrated with other data and need to interoperate with applications or workflows for analysis, storage, and processing	Metadata and data should be well-described so that they can be replicated and/or combined in different settings

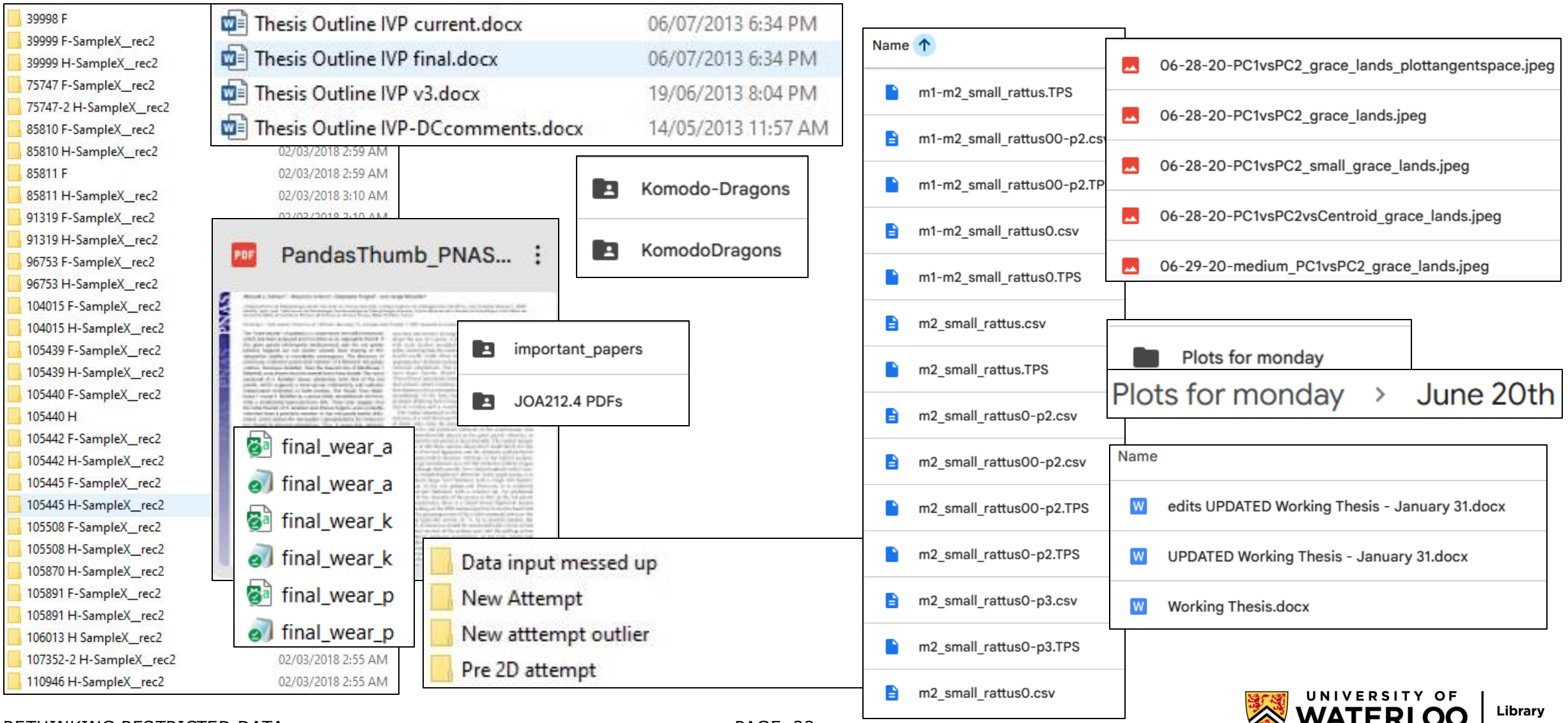
Fair for who?

- Before data can be FAIR for others, it must first be FAIR for you
  - If you can't find or access your data, you can't make it FAIR for others
- Prioritize making sure you can find, access, open, and reuse your own data
  - This will help make it reusable in the future



Baker, 2016

File name hall of shame



Finding the unfindable

Home / Collections / Search results

Choose filters | homo floresiensis | Keyword | Album | Empty filters

7 Tags found | 175 Albums found

No results are available.

You can try to edit your filters and perform a new search.

Collections

- Ditsong National Museum of Natural History
- Espace muséal d'Andenne
- Institute of Archeology, CAS, Brno
- Jersey Museum
- McGregor Museum
- MiC-SABAP Imperia Savona
- Musée Archéologie Nationale
- Musée des Confluences
- Museum für Naturkunde
- National Museum of Lao PDR
- National Museum of Natural History
- Natural History Museum, United Kingdom
- Naturalis Biodiversity Center
- Publications
- Rockefeller Museum
- Senckenberg Gesellschaft für Naturforschung
- State Museum of Natural History Stuttgart
- Taï Chimpanzee Project
- University of Kent
- University of Liège
- University of the Witwatersrand
- Yale Peabody Museum

FAIR-ify as you go

- Create documentation—lots of it
- Saves files in interoperable formats
- Avoid acronyms—document those you do use
- Avoid missing values—document why values are missing
- Keep track of any ethical, legal, or commercial restrictions you have & what data they apply to
- Consider any consent required for data sharing ***early***

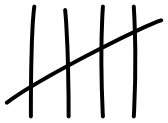
Key tips

Create a folder structure

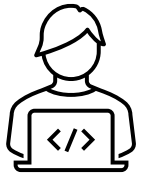
Cleanup unsorted files regularly

Name files [well]

Backup files



Version files numerically



Document ✨ everything ✨

Data management planning

- Create a data management plan
- Think about your research holistically
- Planning will help you anticipate challenges before they happen
  - May be able to avoid some challenges
- Think about consent and data sharing early

Data collection

Documentation & metadata

Storage & backup

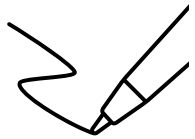
Preservation

Sharing & reuse

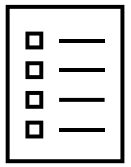
Responsibilities & resources

Ethics & legal compliance

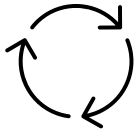
Creating a DMP



- As detailed as possible/needed
  - Different levels of details for different templates, funders, etc.



- Project-specific details
  - The more project-specific the better
    - File naming conventions
    - File versioning details



- Considers the lifespan of a project
  - Before a project starts until after it finishes

Research data management

Organization, maintenance, and archiving of research data throughout the project's life

Includes processes undertaken to create organized, documented, accessible, and reusable quality research data

Benefits of RDM

Workflow Efficiency

- Saves time long-term
- Easier to find files
- Organized data

Academic Reputation

- Signifies research excellence
- Increased citations
- Visibility & recognition of scholarly work

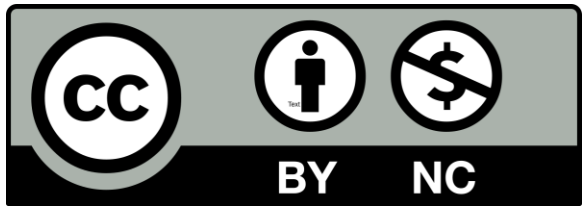
Contribution to research community

- Better research replicability
- Long-term cost savings
- Increased data integrity

WHERE DO WE GO FROM HERE?

Credits & licensing

Creative Commons



UNIVERSITY OF WATERLOO

Library

YOU+WATERLOO

Our greatest impact happens together.