OPEN DATA, DATA SHARING AND ETHICS

Image by Adam Douglas

OPEN SCIENCE STUDENT SUPPORT GROUP

NOVEMBER 6, 2020

CHELSEA MORAN & JENELLE MORGAN

WHO WE ARE

Chelsea Moran

PhD (c), Clinical Psychology

Behavioural Medicine Lab

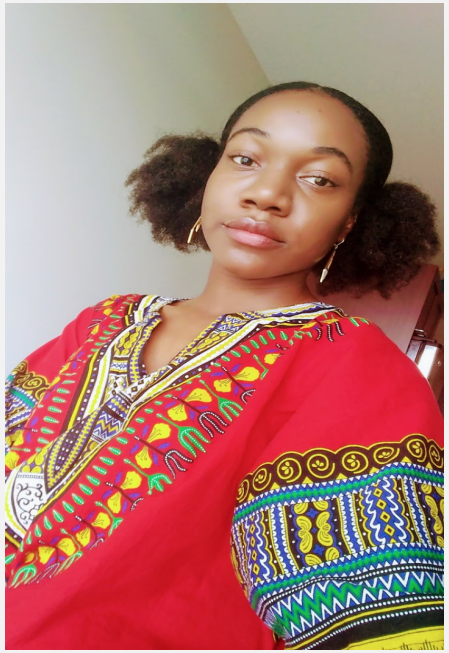
Twitter: @chelseavmoran

Jenelle Morgan

M.Sc. Student, I/O Psychology

Selection and Recruitment Lab

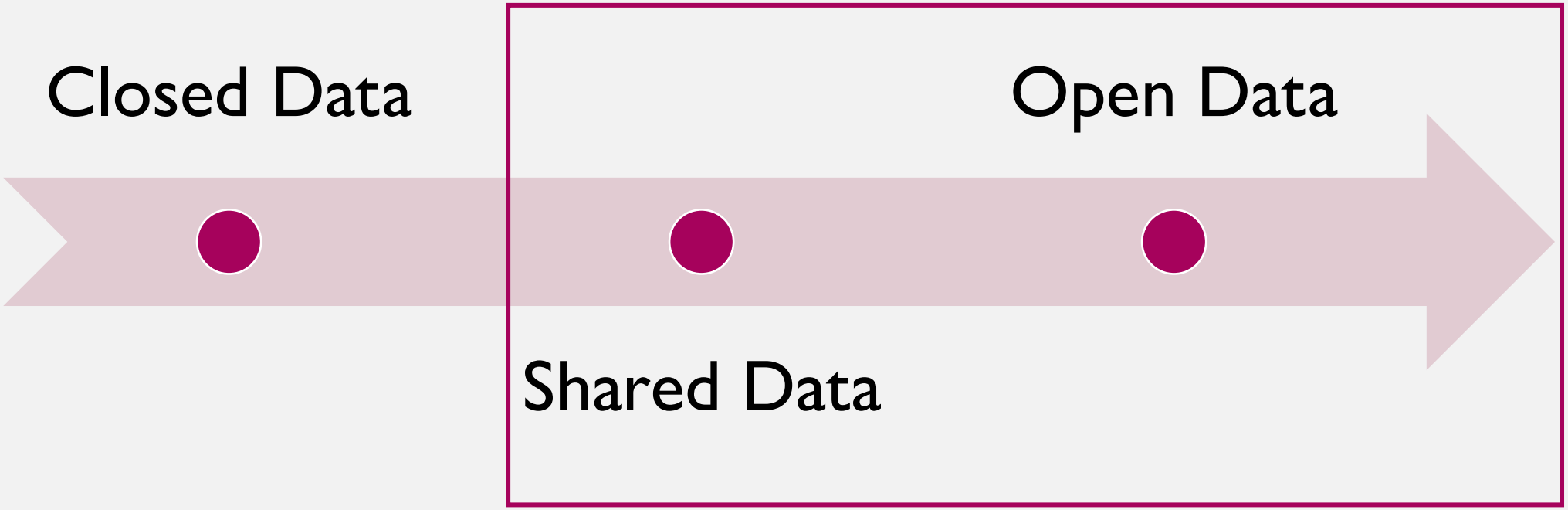
Facebook: Jenelle Morgan



DATA SHARING

Data is anything and everything produced during the research process.

LEVELS OF DATA SHARING & OPENNESS



WHAT IS OPEN DATA?

"Open data and content can be freely used, modified and shared by anyone for any purpose."

Open Knowledge Foundation

WHERE TO FIND OPEN DATA?

Supplemental material published with a research article

(i.e., hosted by publisher)

Publicly-available website

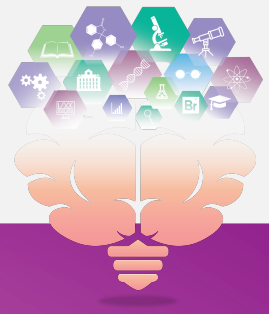
(e.g., personal or lab website)

Repository developed to support data sharing

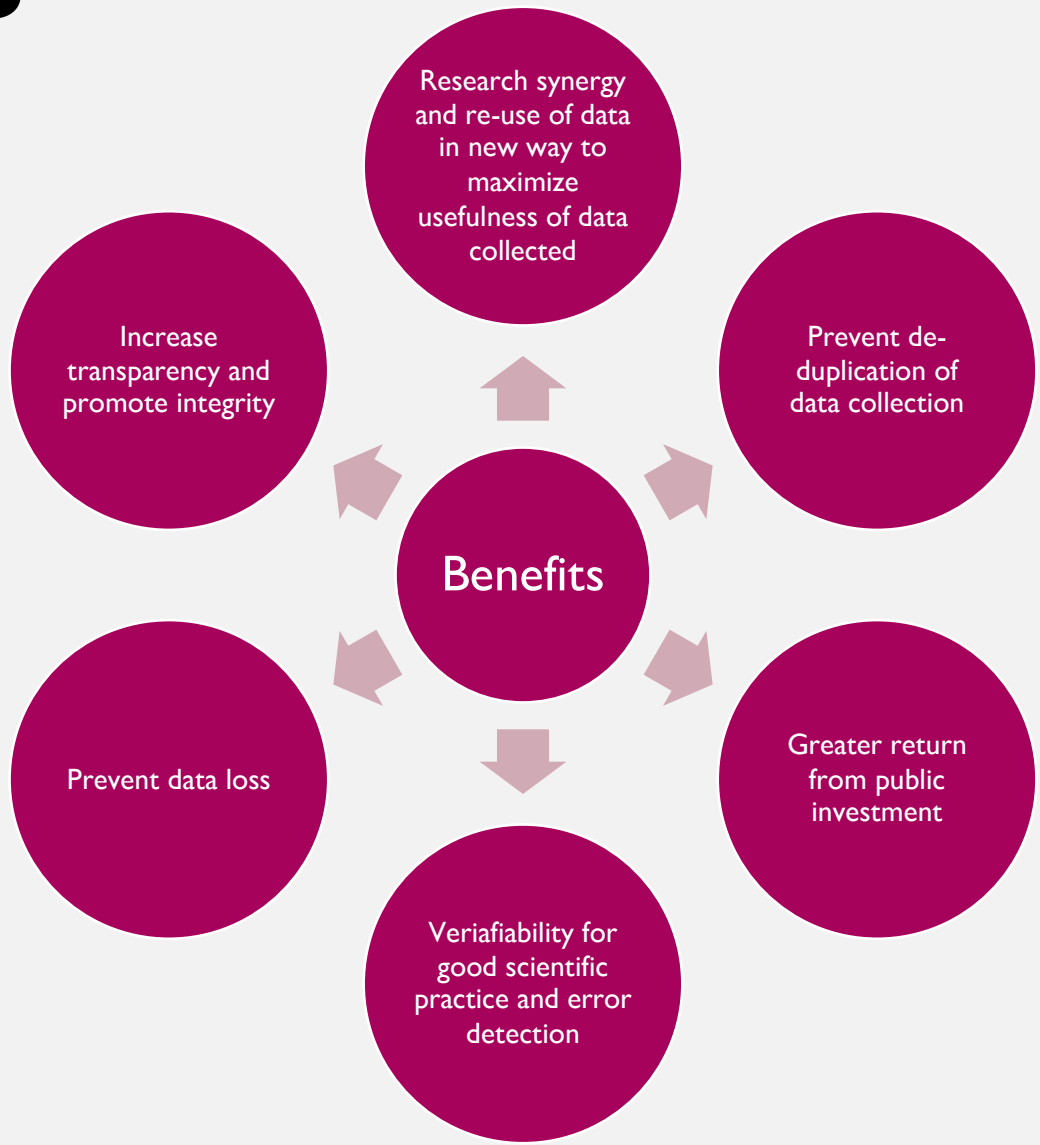
(e.g., Dataverse, Dryad, OSF, institutional/specialist options)

Data papers used to describe datasets, published in data journals

(e.g., Scientific Data, Data Science Journal)

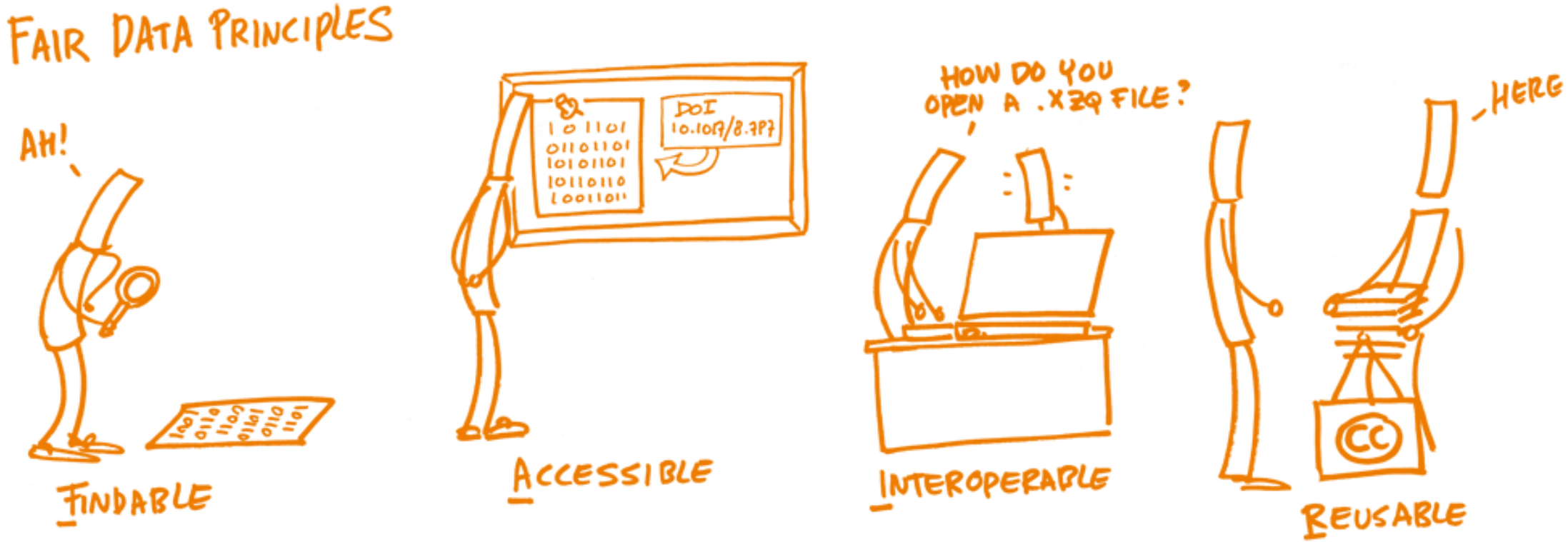


BENEFITS


Research synergy and re-use of data in new way to maximize usefulness of data collected
Prevent de-duplication of data collection
Greater return from public investment
Veriafiability for good scientific practice and error detection
Prevent data loss
Increase transparency and promote integrity
Benefits

HOW DOES IT WORK?

Things to consider	Tips and details
1. Am I allowed to share this data, and under which conditions? 2. Am I allowed to use the data?	• Ethical approvals/informed consent • Who owns the data (license restrictions)
1. What data will I be sharing? 2. What data will I be using?	• Entire dataset, part of dataset, or only data/variables associated with analyses from a specific manuscript
How will I share this data?	• Choose storage option/research data repository that will meet your needs (e.g., access, permanence, file size limits) • Make sure data is identified using a permanent link like a DOI
How will I license my data?	• Choose license that best suits how you wish your data to be used by others (see here for a guide on license types) • Panton Principles
What information will others need in order to understand/use this data?	• Prepare a package describing the features and context of the dataset (e.g., how data was collected, sample characteristics, size, format type, field definitions and data creation scripts or textual documentation as applicable) and include reproducible code/syntax
1. What do I need to do to prepare the data for sharing? 2. What do I need to do before using open data?	• Deidentification (e.g., using Safe Harbor Method)



TRY TO MAKE IT "FAIR"

SOURCE: OPEN SCIENCE TRAINING HANDBOOK

ETHICAL CONSIDERATIONS

Ethical Considerations for Using Open Data	Ethical Considerations for Data Sharing
Consider Terms of Use	Ethics Proposal
Ethical Review Board, Copyrighting, Licenses and Potential Fraudulent Use of Data	Ethics Review Board
Informed consent	Informed consent
De-identification	De-identification

BRIEF EXAMPLE

- **Ethics application "language":**

  5.0 If applicable, describe your plans to link the data in this study with data associated with other studies (e.g. within a data repository) or with data belonging to another organization:

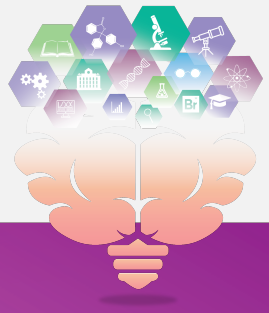
  "We plan to publish the final dataset on open science framework (OSF) after the publication of study results. We will remove gender, ethnicity and province of study variables from the dataset before posting to remove the small likelihood that individuals could be identified."

- **From informed consent form:** "Following analysis and publication of the results, data collected from your participation in this research study will be de-identified and will be held in a database will be posted on online on the Open Science Framework webpage (https://osf.io/) for future use by other researchers. Any future use of this research data is required to undergo review by a Research Ethics Board. "

- OSF Project with Data: https://osf.io/ftw6v/

BARRIERS

- Legal issues, misuse of data, insufficient time, lack of funding, incompatible data types
- Some data cannot be open – not all data should be open: privacy protection, national security, commercialization
- Ethics and confidentiality when working with human data
- Standards are different across sub-discipline and type of data collected (e.g., qualitative)
- Technological barriers – resources to set up database



CHALLENGES

Learn more about it!

- Check out the resources shared in today's presentation

Talk about it!

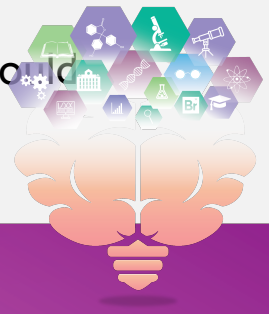
- Talk to your lab supervisor and collaborators about the pros and cons of sharing data and using shared data sets in your field.

Try it out!

- Find a dataset that would be interesting for you to use, download it, explore it - just to see what it would be like!
- Explore different data sharing platforms (e.g., OSD, Dataverse, Dryad, among many others!) and identify 1 or 2 options that you might be interested in using in the future.

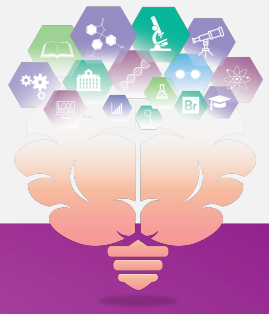
Implement it!

- Make a data sharing plan for your next project, according to the FAIR principles.
- Create an account on a data sharing platform.
- Identify an open dataset that you'd be interested in using and create a plan of ethical issues to consider if you should use that dataset.



RESOURCES

- Watch (videos):
  - Managing and Sharing Research Data (lecture by Sarah Jones on behalf of Foster Open Science, approximately 55 mins)
  - Open Access to Research Data in Horizon 2020 (lecture explaining importance of research data management and open data, examples, and resources to help support implementation, approximately 15 minutes)
- Read (articles/papers/blogs):
  - Open Data Handbook (Website with collection of guides, resource library and value stories highlighting examples of the social and economic value of open data)
  - The FAIR Guiding Principles for scientific data management and stewardship (Peer-reviewed article)
  - Data reuse and the open data citation advantage (Peer-reviewed article)
  - Open Licensing and File Formats (From the Open Science Training Handbook)
  - Practical Data Anonymization (Slides with How-to)
  - Open Science, Open Data (Slides with tips on how to answer the most commonly heard objections to data sharing starting on slide #32)
  - Open Science Training Handbook (see section called "Open Research Data and Materials")
  - Managing and Sharing Research Data (Free online course through University of Leeds with links to multiple resources. Can click through content quickly, no need to complete all modules of the course)
  - Data Ethics Canvas (tool to help identify and manage ethical issues for anyone who collects, shares or uses data)



REFERENCES

Bezjak, S., et al. (2018). Open Research Data and Materials. In Open Science Training Handbook (Eds). Retriveved from: https://book.fosteropenscience.eu/en/

Burgelman, J., Pascu, C., and Szkuta, K., Von Schomberg, R., Karalopoulos, A., Repanas, K & Schouppe, M. (2020). Open science, open data and open scholarship: European policies to make science fit for the 21st century. Frontiers in Big Data

Hand, D. (2018). Aspects of Data Ethics in a Changing World: Where are we now? Big Data, 6(3), 176-190

Huston, P., Edge, V. & Bernier, E. (2019). Reaping the benefits of open data in public health. Canada Communicable Disease Report, 45(11), 252-256.

Krotov. V. (2018). Legality and Ethics of Web Scraping, presented at 24th Americas Conference on Information Systems, New Orleans.

Piwowar, H. A., & Vision, T. J. (2013). Data reuse and the open data citation advantage. PeerJ, 1, e175.

Open Knowledge Foundation (2020). The Open Data Handbook. Retrieved from: https://opendatahandbook.org/

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., ... & Bouwman, J. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific data, 3(1), 1-9.