

Yale

Harvey
Cushing/
John Hay
Whitney
**MEDICAL
LIBRARY**

Reusing Data in a Thesis

Data Reuse Checklist

1. Does this data exist?
2. Is this data available?
3. Is this data usable?
4. What “rules” are associated with using this dataset?

Does this Data Exist?

- Why would this data be created?
- Who would be motivated to collect and maintain this data?
- Will this data exist in the future (think, COVID-19 contact tracing data)
- Is this data you will need to collect (think, social media data)

Data Availability

- Is the data creator providing access to this data? (All, or some)
- How is this data made available?

Is this Data Usable?

- Research considerations
 - Data contains necessary scope and variables to answer your question
- Computational considerations
 - File formats
 - Dataset size
- Data documentation
 - Variable definitions
 - Known units of measurements

Data use agreements, copyright and licencing.

Now that you
have the data,
what can you
do with it?

Data Use Agreements (DUAs)

- Terms of use how a dataset may be used and how it should be cited
- Typically associated with data procured more directly through a provider

Data and Copyright

- Data does not fall under U.S. copyright law (you cannot copyright facts)
- Databases (the organizational structure of the data) may fall under copyright
- Individual datum may fall under copyright (think, a database of pharmaceutical products)

Data Licences

- Terms of use how a dataset may be used and how it should be cited
- Typically associated with Open datasets downloaded through a web access portal

Copyright Services at the CWML

Consult with
your Scholarly
Communication
Librarian to
discuss your
thesis, licences
and copyright

Email
[lindsay.barnett@
yale.edu](mailto:lindsay.barnett@yale.edu) to get
started

Data Services at the CWML

Email
medicaldata@yale.edu to get
started

Attend classes
and workshops

Ask about data
visualization, data
management,
and data
transformation