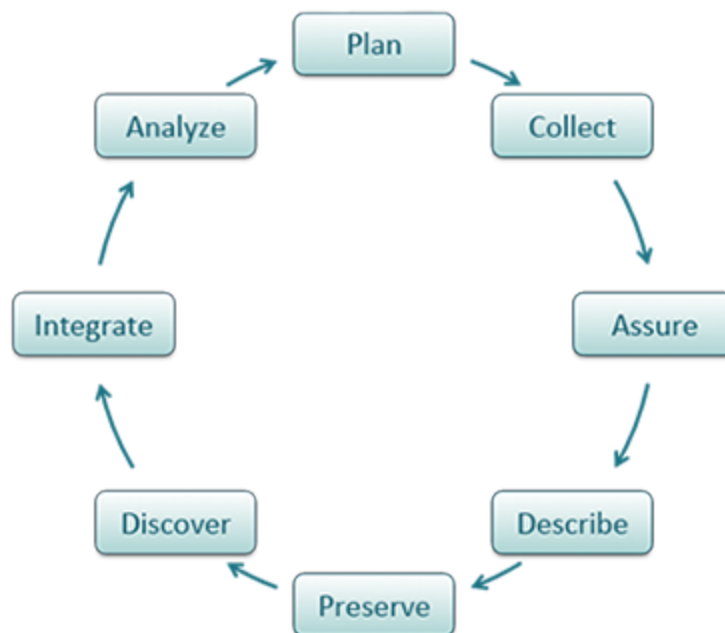


Describe



- A. Why is this part of the Data Lifecycle Important to graduate students?
- B. Sharing data that you collect and analyze for your thesis is really important to justify your decisions and results. Describing it as you go is the only way to remember what you actually did. Do you remember what you ate for breakfast three weeks ago? Do you think you will remember what you did to your data three weeks from now?
 - C. Describing your data well will help a lot when the time to publish the data comes along, or when you leave the research lab and somebody else will carry on your work.
 - D. Data without metadata can be useless (if you take a sample from human gut or from the ocean may just be very important information)
 - E. Learning about domain metadata standards early is helpful for future data discovery and reuse.
 - F. Many modern projects include many institutions - communicating across institutions requires good metadata.
 - G. Journals (or funding agencies) often require that the data are published alongside a paper; metadata are needed to describe these data for depositing into repo and for reuse by others
- H. How can this be integrated into Graduate School?
- Introduction to the topic throughout grad school (through talks?) +1
 - Should be a required section of thesis proposal
 - ESIP - learning things that go back to lab/advisor
 - Grad students often learn from their lab group to keep a paper or electronic journal, they could also describe their data in their lab notebooks (example: can

you remember what you had for breakfast 3 weeks ago Friday?)- can these practices be shared or homogenized across lab groups or wider programs?

- Can we develop fun activities in student groups at conferences or in a domain-oriented meeting (concrete examples in the domain)
- Informal training opportunities offered by library or informatics groups (e.g., for ocean sciences domain <https://tinyurl.com/WHOI-Data-Science-syllabus>)
-

I. What are the resources graduate students can use for this part of the Data Lifecycle?

(Distilled version of) Domain standard descriptions

MIXS, Darwin Core

ESIP - learning things that go back to lab/advisor

Openscapes (cohort training, long-term bitesize learning and application)

Attendance/presentations at domain meetings (AGU, ESA, GSA, ...)

NCEAS and DataONE virtual trainings

https://dataoneorg.github.io/Education/video_step/describe/

Record "Fabulous Fails in Data Management" - Light hearted lightning talks from

ESIP community that can be shared more broadly.

Data One Stories: <https://old.dataone.org/data-stories>

