



# Glossary

---

**Research Data:** “Units of information observed, collected, or created in the course of research”

Erway, Ricky. 2013. Starting the Conversation: University-wide Research Data Management Policy. Dublin, Ohio: OCLC Research.

<http://www.oclc.org/content/dam/research/publications/library/2013/2013-08.pdf>.

**Collaborating:** Sharing materials within a research group on a limited basis

**Sharing:** providing access to materials publicly or semi-publicly (i.e. restricted, but with a clear request process)

**Archiving:** Long-term (greater than 5 years) preservation, often with public or semi-public sharing

# Decisions, decisions, decisions...

---

- Keep or delete?
- Wanted, needed, or absolutely essential?
- Can I find it later?

It's a Monday morning and your hard drives are crashing. The IT person says she has enough time to save three files; which three are most important? What are their names?

# Data Management Education

---

## Theory vs. Reality

“how [do] we connect discipline-agnostic material to your discipline-specific workflows”?

Whitmire, Amanda. GRAD 521: Research Data Management: Syllabus.  
<http://guides.library.oregonstate.edu/course-guide/GRAD521/grad521syllabus>

# Why do we care?

PLOS ONE retracts breast cancer genetics paper after claim of misappropriated data

Wrong in the tooth: Faked data, authorship issues force retraction of dental paper

with one comment

**“We are living in hell:” Authors retract 2nd paper due to missing raw data**

with 15 comments

A 2006 paper investigating the effects of docosahexaenoic acid (DHA) and celecoxib on prostate cancer cells has been retracted because it appears to contain panels that were duplicated, and the authors could not provide the raw data to show otherwise.

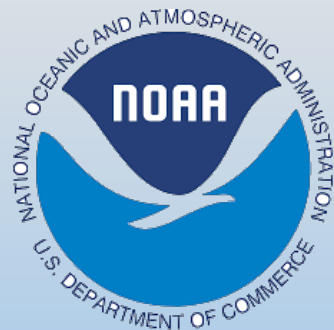


This is the second paper the authors have lost [because they couldn't furnish the original data](#) to defend their work against allegations of image manipulation. The reason: the Institute for Cancer Prevention in New York, where the authors did the work, shut its doors abruptly in 2004, co-author [Bhagavathi A. Narayanan](#) told us. (The institute closed thanks to \$5.7 million in grant that was misspent, [the New York Post reported](#) at the time.)

<http://retractionwatch.com>

# Federal Data Sharing Policies

---



# What is good data practice?

---

- Know what data you have
- **Keep relevant metadata**
- **Have proper storage, backup, security**
- Observe privacy and ethics principles
- *Know your policies and legal obligations*
- Plan for access and sharing
- Properly archive or preserve data



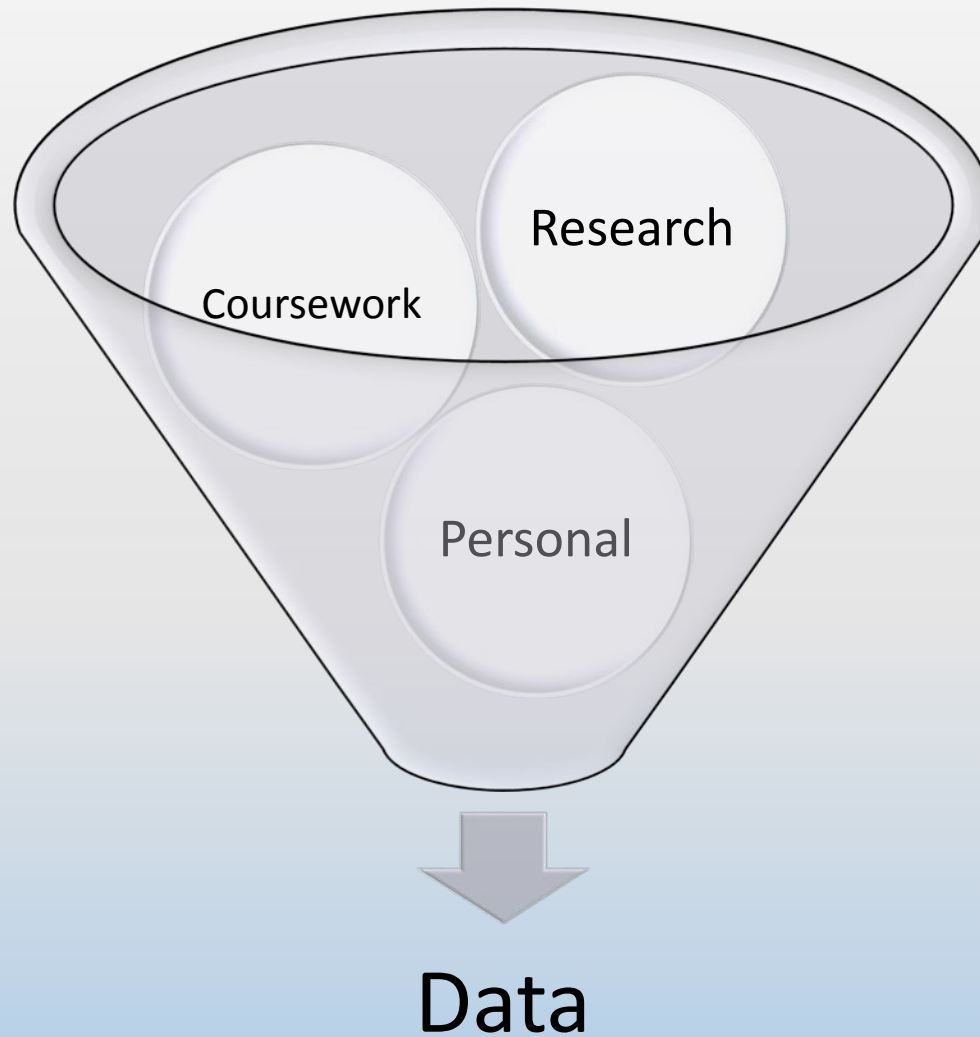
# What does it look like?





# Keep them separate

---



# From reality to good practice

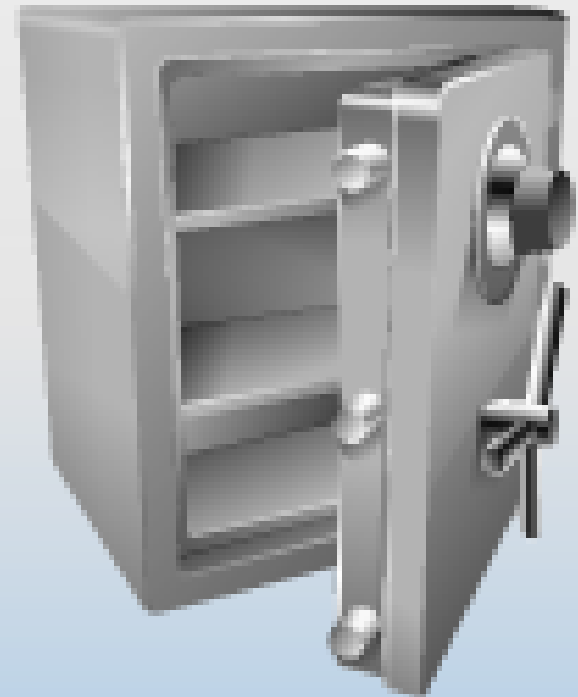
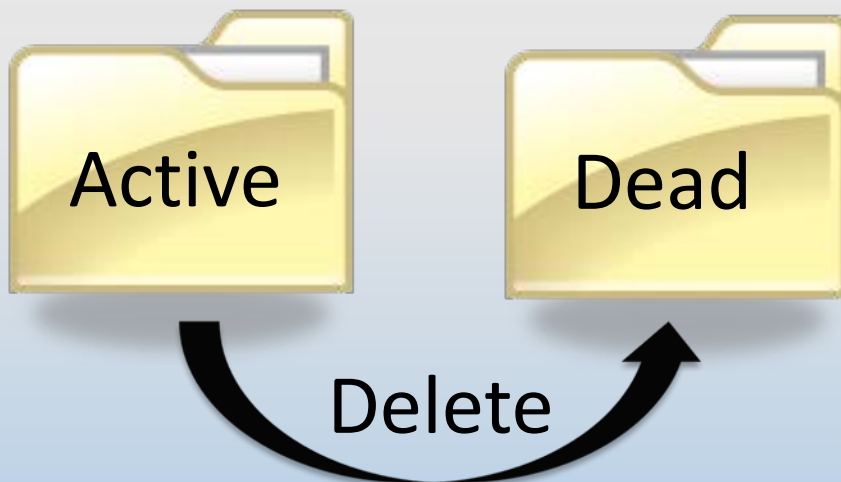
---

Have a system

Prioritize – have a ‘vault’

Purge or archive regularly

Backup regularly



# Document, document, document!

- Who, what, when, why, where...
- Have consistent file organization
- Describe the data – README.TXT files
- Perform quality assurance
- Prioritize raw data



**Hans Hassell**  
@hjghassell



@emilythorson @AlbertsonB2 @sgadarian  
We collected data until RA quit & none of us  
wanted to do it so we called it good  
[#overlyhonestmethods](#)



**Mihaela Bereza**  
@Atunci\_CoV



Subject withdrawn from study because we  
couldn't figure out why her data were in a  
biologically implausible range  
[#Overlyhonestmethods](#)



**BriAnna Garza**  
@Hybridbrii



Had to throw out that outlier data bc I was  
testing scent and memory and someone came  
in smelling like weed that day.  
[#OverlyHonestMethods](#)