

Introduction to Research Data Management

Amanda L. Whitmire
Steven Van Tuyl
OSU Libraries

GOAL:

Achievable habits for implementing
data management best practices into
your workflow



Research data is:

“...the recorded factual material commonly accepted in the scientific community as necessary to **validate** research findings.”

U.S. Office of Management and Budget, Circular A-110



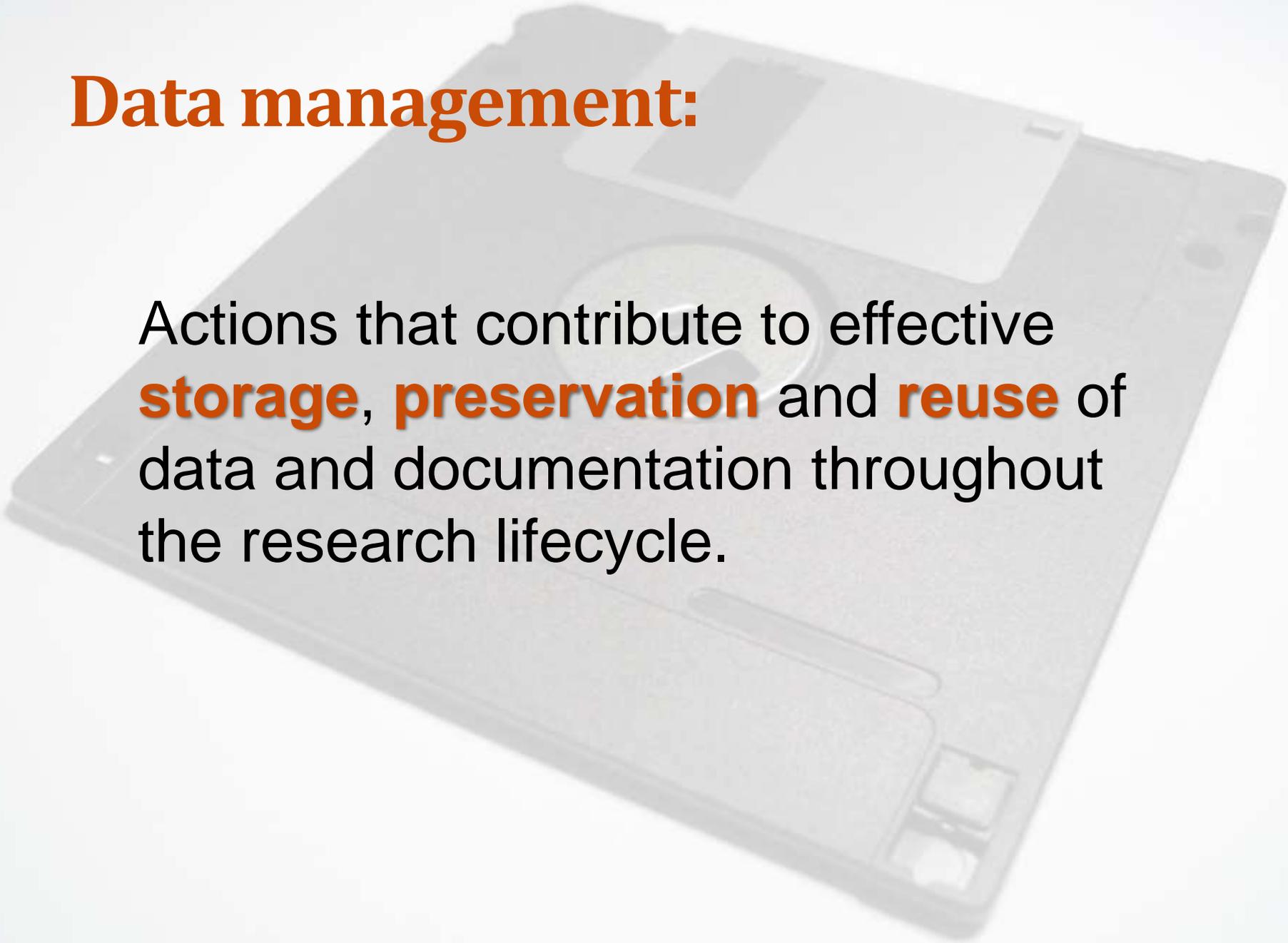
Research data is:

“Unlike other types of information, research data are collected, observed, or created, for the **purposes of analysis to produce and validate original research results.**”

University of Edinburgh
MANTRA Research Data Management Training,
'Research Data Explained'



Data management:



Actions that contribute to effective **storage**, **preservation** and **reuse** of data and documentation throughout the research lifecycle.

What data management is not:

Data/computational science

Database administration

A research method:

- what data to collect
- how to collect them
- how to design an experiment

Why data management?

Further your field

Increase visibility & impact

Saves time

Protects investment

Increases research efficiency

Preservation

Funding agency requirements

Further your field



Atul Butte
@atulbutte



Following

New markers for preeclampsia, found using #opendata from @NCBI GEO: ncbi.nlm.nih.gov/pubmed/24195779 #ISMB @Carment

Reply Retweet Favorite

RETWEETS FAVORITE

11

1



12:42 PM - 14 Jul 2014

PubMed.gov
US National Library of Medicine
National Institutes of Health

PubMed Advanced

Display Settings: Abstract

[BMC Med.](#) 2013 Nov 6;11:236. doi: 10.1186/1741-7015-11-236.

Integrating multiple 'omics' analyses identifies serological protein biomarkers for preeclampsia.

[Liu LY](#), [Yang T](#), [Ji J](#), [Wen Q](#), [Morgan AA](#), [Jin B](#), [Chen G](#), [Lyell DJ](#), [Stevenson DK](#), [Ling XB](#)¹, [Butte AJ](#).

Author information

Abstract

BACKGROUND: Preeclampsia (PE) is a pregnancy-related vascular disorder which is the leading cause of maternal morbidity and mortality. We identified novel serological protein markers to diagnose PE with a multi-'omics' based discovery approach.

METHODS: Seven previous placental expression studies were combined for a multiplex analysis, and in parallel, two-dimensional gel electrophoresis was used to compare serum proteomes in PE and control subjects. The combined biomarker candidates were validated with available ELISA assays using PE (n=32) and control (n=32) samples. With the validated biomarkers, a genetic algorithm was then used to construct and optimize biomarker panels.

RESULTS: In addition to the previously identified biomarkers, the angiogenic and antiangiogenic factors (soluble fms-like tyrosine kinase-1 (sFlt-1) and soluble endoglin (sEng)), we found 3 up-regulated and 6 down-regulated biomarkers in PE sera. Two optimal biomarker panels were developed for clinical assessment, respectively.

CONCLUSIONS: Both early and late onset PE diagnostic panels, constructed with our PE biomarkers, were superior over sFit-1/PlGF. The functional significance of these PE biomarkers and their associated pathways were analyzed which may provide new insights into the pathogenesis of PE.

PMID: 24195779 [PubMed - indexed for MEDLINE] [Free full text](#)

Increase visibility & impact

 OPEN ACCESS  PEER-REVIEWED

RESEARCH ARTICLE

34,276

VIEWS

97

CITATIONS

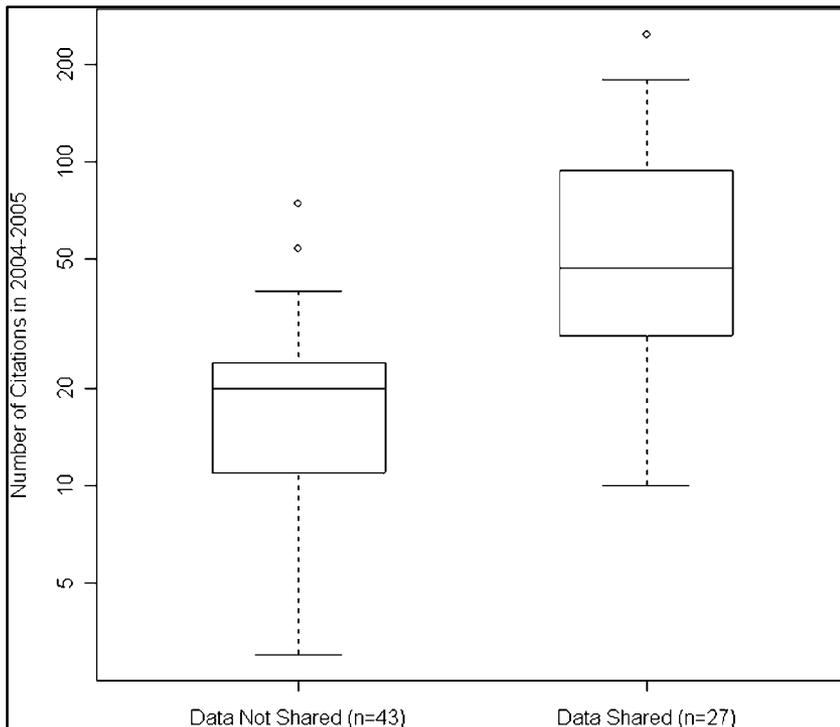
359

SAVES

Sharing Detailed Research Data Is Associated with Increased Citation Rate

Heather A. Piwowar , Roger S. Day, Douglas B. Fridsma

Published: Mar 21, 2007 • DOI: [10.1371/journal.pone.0000308](https://doi.org/10.1371/journal.pone.0000308) • Featured in [PLOS Collections](#)



85 cancer microarray clinical trial publications

69% increase in citations for articles w/publicly available data

Funder mandates



“...directed Federal agencies with more than \$100M in R&D expenditures to develop plans to make the published results of federally funded research freely available to the public within one year of publication and requiring researchers to better account for and manage the digital data resulting from federally funded scientific research.”

Posted by Michael Stebbins on February 22, 2013 at 12:04 PM EST



The Obama Administration is committed to the proposition that citizens deserve easy access to the results of scientific research their tax dollars have paid for. That's why, in a policy memorandum released today, OSTP Director John Holdren has directed Federal agencies with more than \$100M in R&D expenditures to develop plans to make the published results of federally funded research freely available to the public within one year of publication and requiring researchers to better account for and manage the digital data resulting from federally funded scientific research. OSTP has been looking into this issue for some time, soliciting broad public input on multiple occasions and convening an interagency working group to develop a policy. The final policy reflects substantial inputs from scientists and scientific organizations, publishers, members of Congress, and other members of the public—over 65 thousand of whom recently signed a *We the People* petition asking for expanded public access to the results of taxpayer-funded research.

Which agencies are affected?



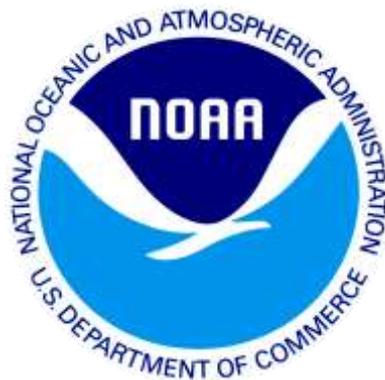
\$54M



\$35M



\$23M



Aspects of data management

DMPs/Planning

Storage & backup

File organization & naming

Documentation & metadata

Legal/ethical
considerations

Sharing & reuse

Archiving &
preservation

Data types & formats

Data types

State Illinois Incorporated place Chicago, Ill.
Block No. 08

Serial	PLACE OF ABODE	NAME	RELATION	HOME DATA	PERSONAL DESCRIPTION	EDUCATION	PLACE OF BIRTH	MOTHER TONGUE OR NATIVE LANGUAGE OF FOREIGN BORN	CITIZENSHIP, ETC.	OCCUPATION AND INDUSTRY		EMPLOYMENT	VETERAN
										INDUSTRY	CODE		
51		Stigman, Ellen R. Johnson	Wife	2 1/2, 000 R No	M W 33	Illinois	Illinois	German	1870	W	Yes	No	
52		William Jones	Head	2 1/2, 000 R No	M W 25	Illinois	Illinois	German	1870	W	Yes	No	
53		John Jones	Wife	2 1/2, 000 R No	F W 23	Illinois	Illinois	German	1870	W	Yes	No	
54		John Jones	Wife	2 1/2, 000 R No	F W 21	Illinois	Illinois	German	1870	W	Yes	No	
55		John Jones	Wife	2 1/2, 000 R No	F W 19	Illinois	Illinois	German	1870	W	Yes	No	
56		John Jones	Wife	2 1/2, 000 R No	F W 17	Illinois	Illinois	German	1870	W	Yes	No	
57		John Jones	Wife	2 1/2, 000 R No	F W 15	Illinois	Illinois	German	1870	W	Yes	No	
58		John Jones	Wife	2 1/2, 000 R No	F W 13	Illinois	Illinois	German	1870	W	Yes	No	
59		John Jones	Wife	2 1/2, 000 R No	F W 11	Illinois	Illinois	German	1870	W	Yes	No	
60		John Jones	Wife	2 1/2, 000 R No	F W 9	Illinois	Illinois	German	1870	W	Yes	No	
61		John Jones	Wife	2 1/2, 000 R No	F W 7	Illinois	Illinois	German	1870	W	Yes	No	
62		John Jones	Wife	2 1/2, 000 R No	F W 5	Illinois	Illinois	German	1870	W	Yes	No	
63		John Jones	Wife	2 1/2, 000 R No	F W 3	Illinois	Illinois	German	1870	W	Yes	No	
64		John Jones	Wife	2 1/2, 000 R No	F W 1	Illinois	Illinois	German	1870	W	Yes	No	
65		John Jones	Wife	2 1/2, 000 R No	F W 0	Illinois	Illinois	German	1870	W	Yes	No	
66		John Jones	Wife	2 1/2, 000 R No	F W 0	Illinois	Illinois	German	1870	W	Yes	No	
67		John Jones	Wife	2 1/2, 000 R No	F W 0	Illinois	Illinois	German	1870	W	Yes	No	
68		John Jones	Wife	2 1/2, 000 R No	F W 0	Illinois	Illinois	German	1870	W	Yes	No	
69		John Jones	Wife	2 1/2, 000 R No	F W 0	Illinois	Illinois	German	1870	W	Yes	No	
70		John Jones	Wife	2 1/2, 000 R No	F W 0	Illinois	Illinois	German	1870	W	Yes	No	
71		John Jones	Wife	2 1/2, 000 R No	F W 0	Illinois	Illinois	German	1870	W	Yes	No	
72		John Jones	Wife	2 1/2, 000 R No	F W 0	Illinois	Illinois	German	1870	W	Yes	No	
73		John Jones	Wife	2 1/2, 000 R No	F W 0	Illinois	Illinois	German	1870	W	Yes	No	
74		John Jones	Wife	2 1/2, 000 R No	F W 0	Illinois	Illinois	German	1870	W	Yes	No	
75		John Jones	Wife	2 1/2, 000 R No	F W 0	Illinois	Illinois	German	1870	W	Yes	No	
76		John Jones	Wife	2 1/2, 000 R No	F W 0	Illinois	Illinois	German	1870	W	Yes	No	
77		John Jones	Wife	2 1/2, 000 R No	F W 0	Illinois	Illinois	German	1870	W	Yes	No	
78		John Jones	Wife	2 1/2, 000 R No	F W 0	Illinois	Illinois	German	1870	W	Yes	No	
79		John Jones	Wife	2 1/2, 000 R No	F W 0	Illinois	Illinois	German	1870	W	Yes	No	
80		John Jones	Wife	2 1/2, 000 R No	F W 0	Illinois	Illinois	German	1870	W	Yes	No	
81		John Jones	Wife	2 1/2, 000 R No	F W 0	Illinois	Illinois	German	1870	W	Yes	No	
82		John Jones	Wife	2 1/2, 000 R No	F W 0	Illinois	Illinois	German	1870	W	Yes	No	
83		John Jones	Wife	2 1/2, 000 R No	F W 0	Illinois	Illinois	German	1870	W	Yes	No	
84		John Jones	Wife	2 1/2, 000 R No	F W 0	Illinois	Illinois	German	1870	W	Yes	No	
85		John Jones	Wife	2 1/2, 000 R No	F W 0	Illinois	Illinois	German	1870	W	Yes	No	
86		John Jones	Wife	2 1/2, 000 R No	F W 0	Illinois	Illinois	German	1870	W	Yes	No	
87		John Jones	Wife	2 1/2, 000 R No	F W 0	Illinois	Illinois	German	1870	W	Yes	No	
88		John Jones	Wife	2 1/2, 000 R No	F W 0	Illinois	Illinois	German	1870	W	Yes	No	
89		John Jones	Wife	2 1/2, 000 R No	F W 0	Illinois	Illinois	German	1870	W	Yes	No	
90		John Jones	Wife	2 1/2, 000 R No	F W 0	Illinois	Illinois	German	1870	W	Yes	No	
91		John Jones	Wife	2 1/2, 000 R No	F W 0	Illinois	Illinois	German	1870	W	Yes	No	
92		John Jones	Wife	2 1/2, 000 R No	F W 0	Illinois	Illinois	German	1870	W	Yes	No	
93		John Jones	Wife	2 1/2, 000 R No	F W 0	Illinois	Illinois	German	1870	W	Yes	No	
94		John Jones	Wife	2 1/2, 000 R No	F W 0	Illinois	Illinois	German	1870	W	Yes	No	
95		John Jones	Wife	2 1/2, 000 R No	F W 0	Illinois	Illinois	German	1870	W	Yes	No	
96		John Jones	Wife	2 1/2, 000 R No	F W 0	Illinois	Illinois	German	1870	W	Yes	No	
97		John Jones	Wife	2 1/2, 000 R No	F W 0	Illinois	Illinois	German	1870	W	Yes	No	
98		John Jones	Wife	2 1/2, 000 R No	F W 0	Illinois	Illinois	German	1870	W	Yes	No	
99		John Jones	Wife	2 1/2, 000 R No	F W 0	Illinois	Illinois	German	1870	W	Yes	No	
100		John Jones	Wife	2 1/2, 000 R No	F W 0	Illinois	Illinois	German	1870	W	Yes	No	

Observational | Can't be recaptured, recreated or replaced; Examples: sensor readings, sensory (human) observations, survey results

Experimental | Should be reproducible, but can be expensive; Examples: gene sequences, chromatograms, spectroscopy, microscopy, cell counts

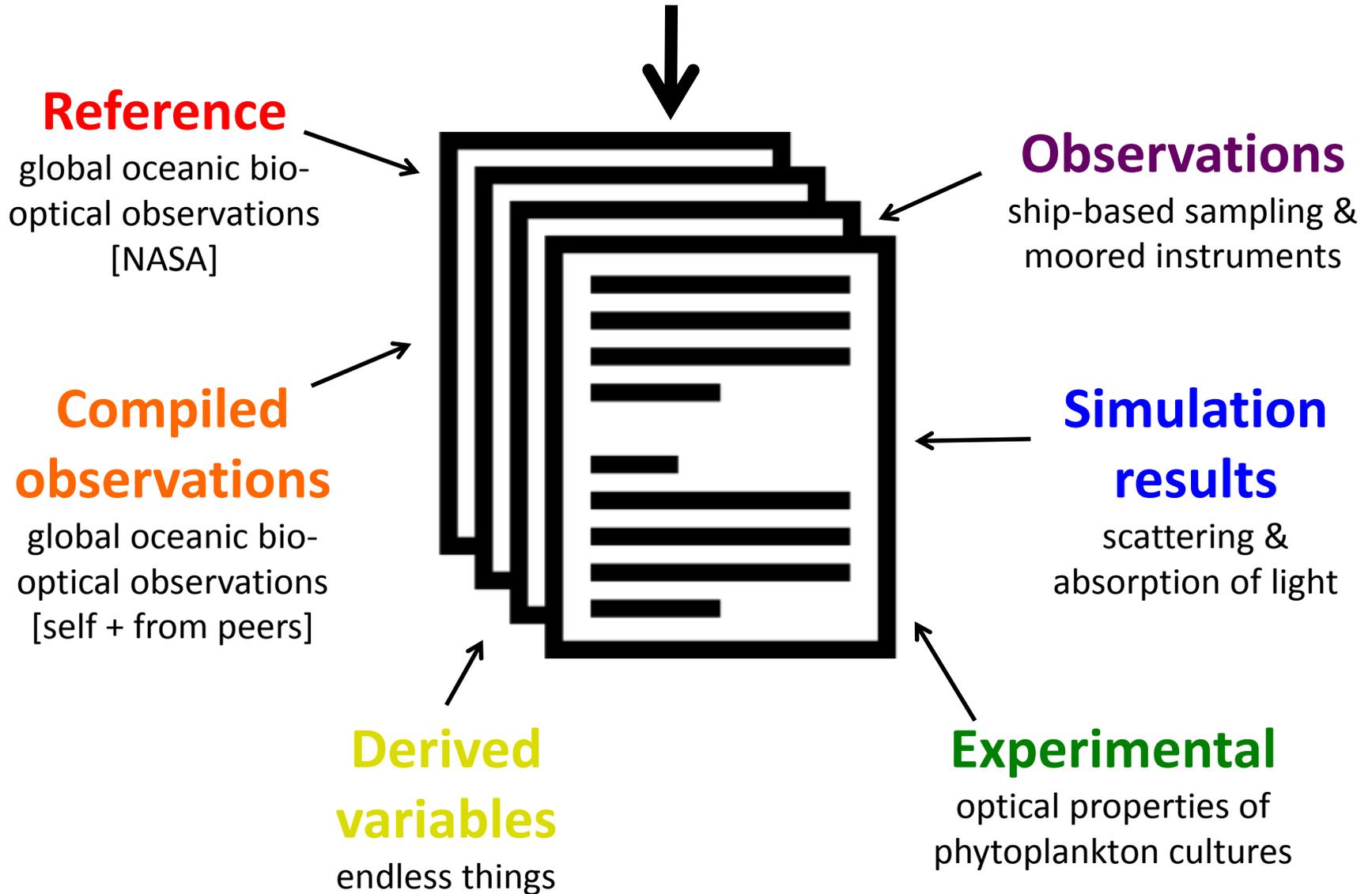
Derived or compiled | Reproducible, but can be very expensive; Examples: text and data mining, compiled database, 3D models

Simulation | Models and metadata, where the input can be more important than output data; Examples: climate models, economic models, biogeochemical models

Reference/canonical | Static or organic collection [peer-reviewed] datasets, most probably published and/or curated; Examples: gene sequence databanks, chemical structures, census data, spatial data portals

Amanda's dissertation

The spectral backscattering properties of marine particles



Data types: another classification

Qualitative data “is a categorical measurement expressed not in terms of numbers, but rather by means of a natural language description. In statistics, it is often used interchangeably with "categorical" data.” See also: **nominal**, **ordinal**

Quantitative data “is a numerical measurement expressed not by means of a natural language description, but rather in terms of numbers. However, not all numbers are continuous and measurable.”

“My favorite color is **blue-green**.” vs. “My favorite color is 510 nm.”

More common data types

Geospatial data has a geographical or geospatial aspect. Spatial location is critically tied to other variables.

Digital image, audio & video data

Documentation & scripts Sometimes, software code IS data; likewise with documentation (laboratory notebooks, written observations, etc.)

File Formats

“A **file format** is a standard way that information is encoded for storage in a computer file. It specifies how bits are used to encode information in a digital storage medium.” - Wikipedia

Data type

Qualitative, tabular
experimental data

Possible formats



- Excel spreadsheet (.xlsx)
- Comma-delimited text (.csv)
- Access database (.mdb/,accdb)
- Google Spreadsheet
- SPSS portable file (.por)
- XML file

Reflective Writing: 60 seconds

What types & formats of data will you be generating and/or using?

Observational | Experimental | Derived | Compiled |
Simulation | Reference/Canonical
Qualitative | Quantitative | Geospatial
Image/audio/video | Scripts/codes

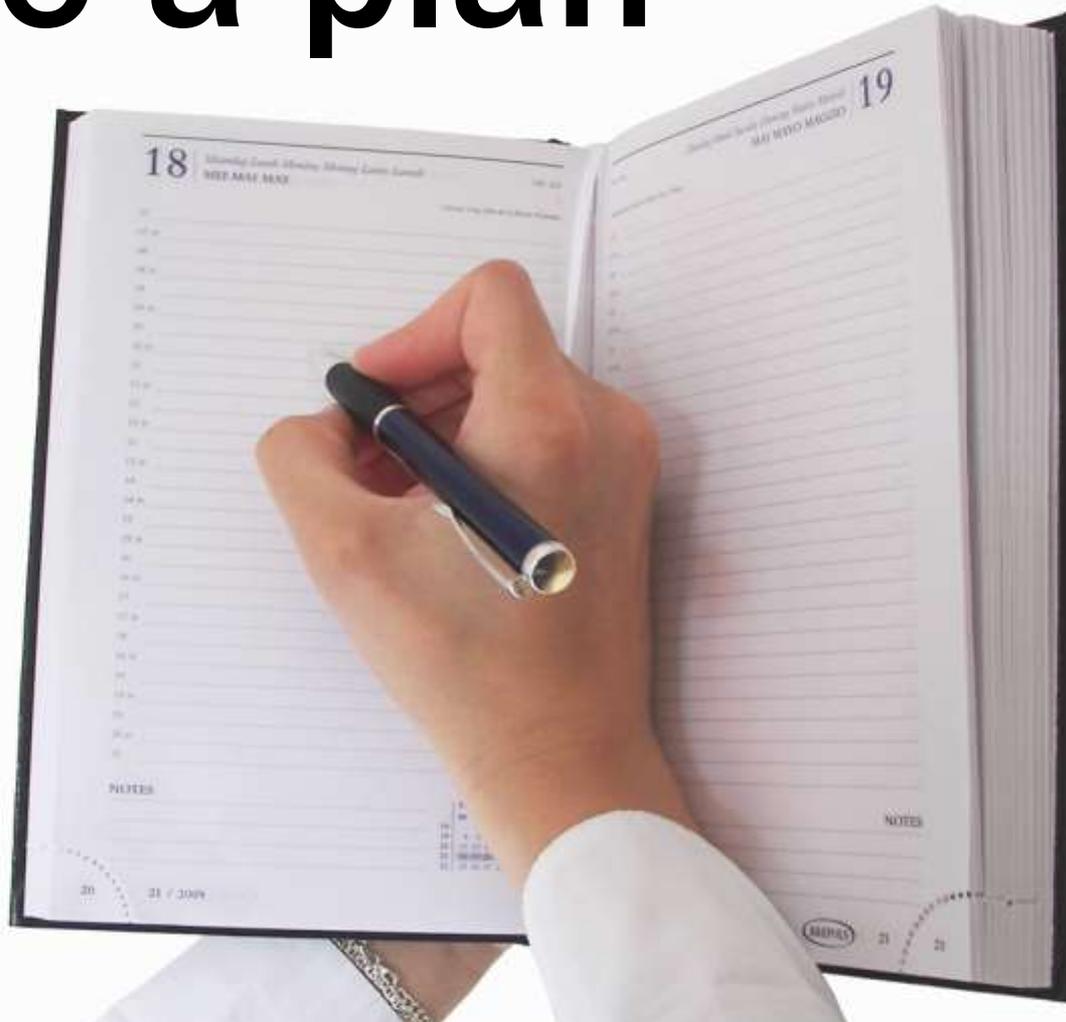


The “big picture”



Where do you start?

Make a plan



Data storage & backup



“Your data are the life blood of your research.
If you lose your data recovery could be slow, costly or even worse...

it could be impossible.”

Most common loss scenario: drive failure

Windows Error Recovery

Windows failed to start. A recent hardware or software change might be the cause.

If Windows files have been damaged or configured incorrectly, Startup Repair can help diagnose and fix the problem. If power was interrupted during startup, choose Start Windows Normally.

(Use the arrow keys to highlight your choice.)

Launch Startup Repair (recommended)

Start Windows Normally

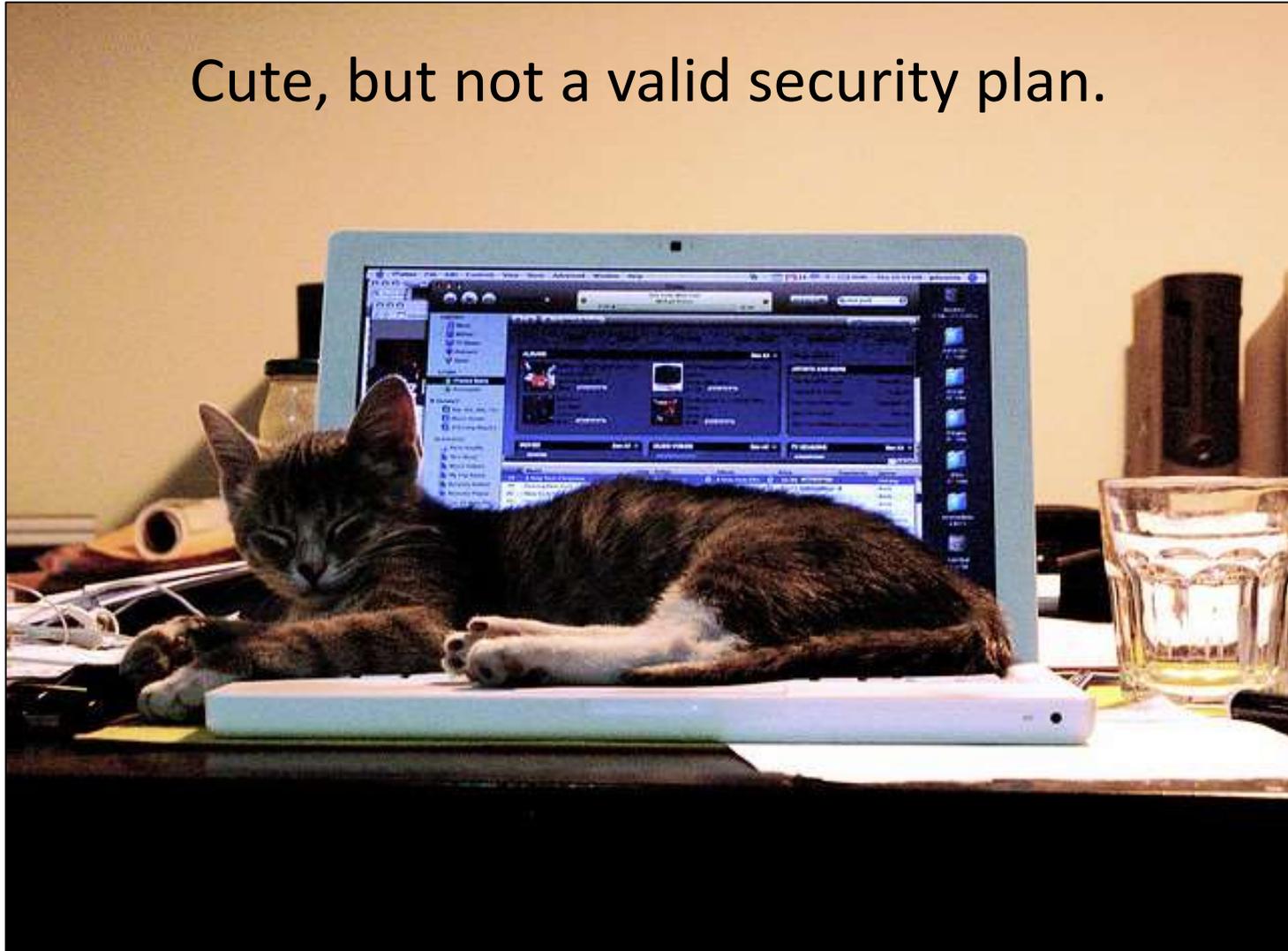
Seconds until the highlighted choice will be selected automatically: 20

Description: Fix problems that are preventing Windows from starting

ENTER=Choose

This happens a lot: physical theft & unintentional damage

Cute, but not a valid security plan.



Rare, unexpected events happen

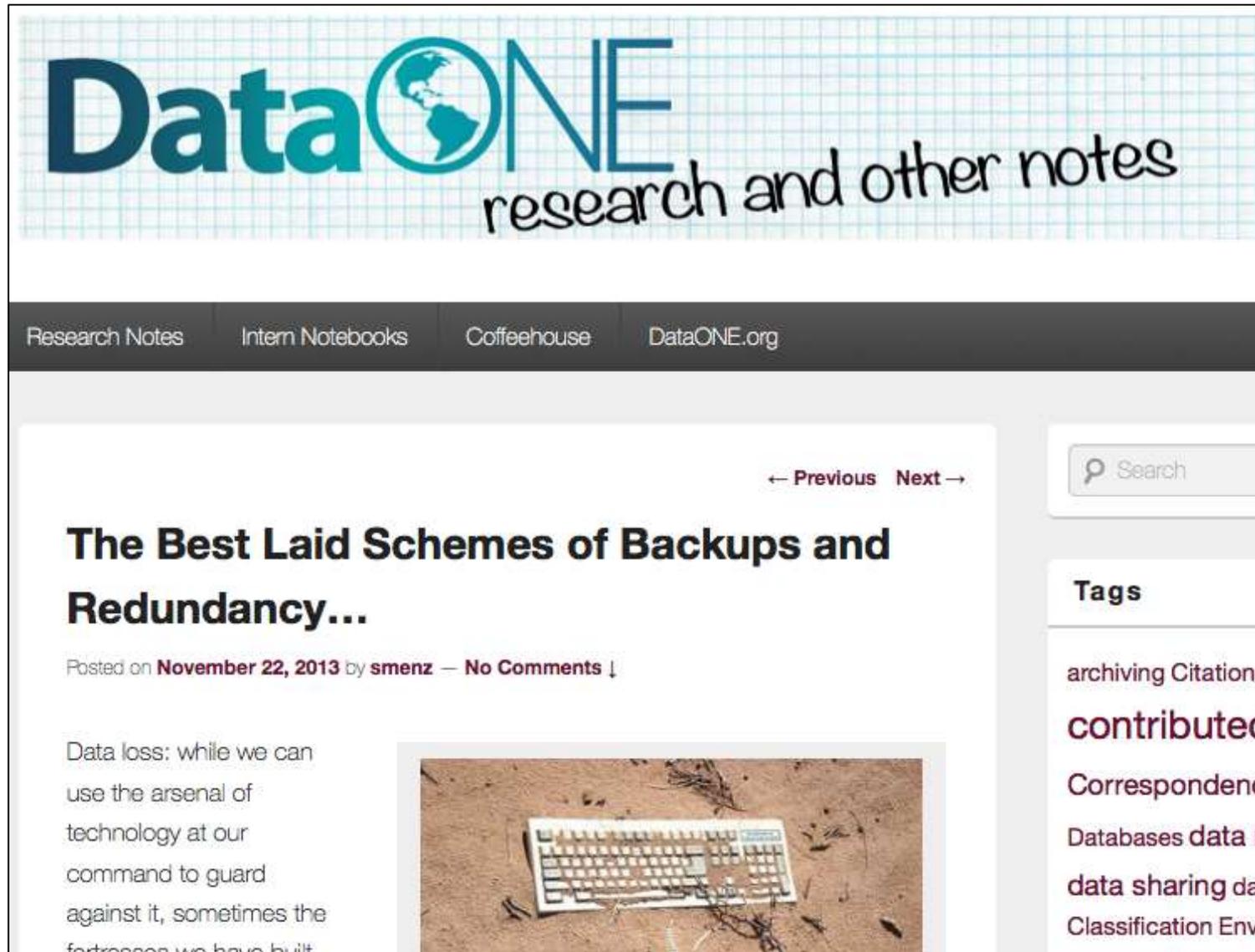
University of Southampton, School of Electronics and
Computer Science, Southampton, UK, 2005



It CAN happen to you



Real-world lesson: Audit your backups...



The screenshot shows the DataONE website header with the logo "DataONE" and the tagline "research and other notes". Below the header is a navigation bar with links for "Research Notes", "Intern Notebooks", "Coffeehouse", and "DataONE.org". The main content area features a blog post titled "The Best Laid Schemes of Backups and Redundancy..." by smenz, dated November 22, 2013. The post text begins with "Data loss: while we can use the arsenal of technology at our command to guard against it, sometimes the fortresses we have built". An image of a keyboard on the ground is partially visible. On the right side, there is a search bar and a "Tags" section with various tags like "archiving Citations", "contributed", "Correspondence", "Databases data", "data sharing da", and "Classification Env".

DataONE
research and other notes

Research Notes Intern Notebooks Coffeehouse DataONE.org

← Previous Next →

The Best Laid Schemes of Backups and Redundancy...

Posted on **November 22, 2013** by **smenz** — **No Comments** ↓

Data loss: while we can use the arsenal of technology at our command to guard against it, sometimes the fortresses we have built



archiving Citations
contributed
Correspondence
Databases data
data sharing da
Classification Env

Data backup

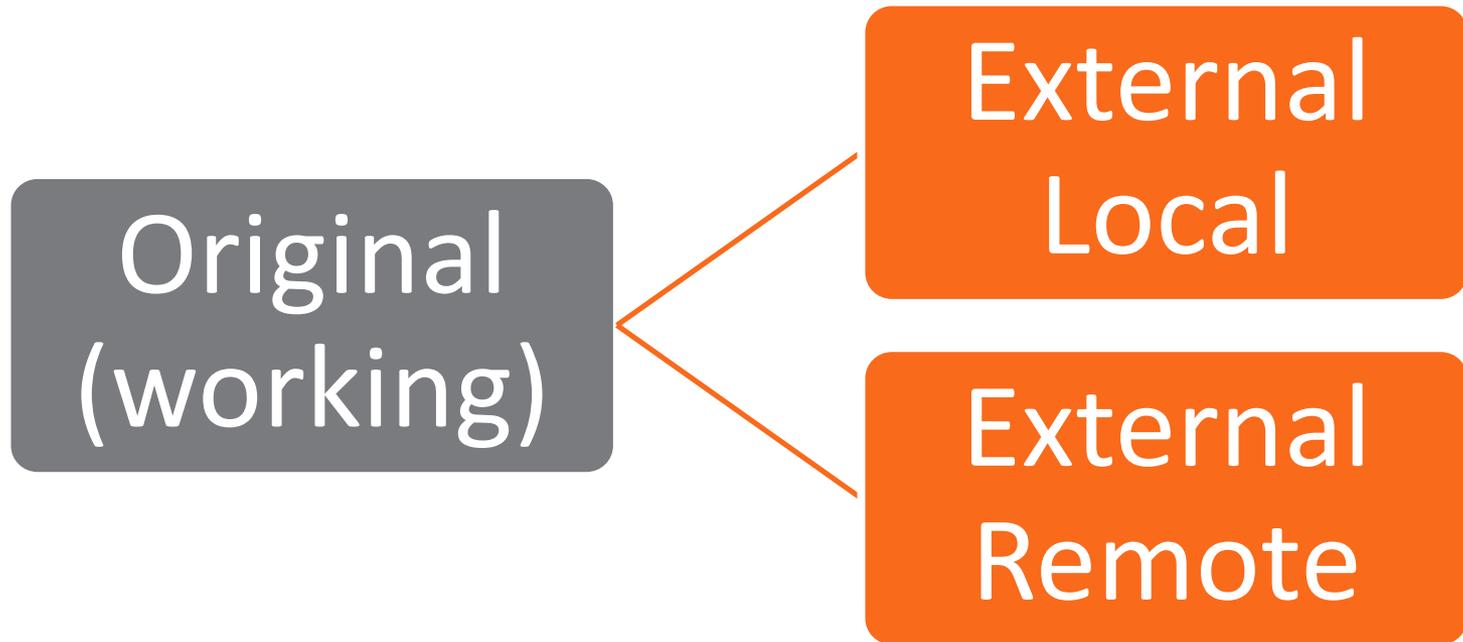
“Keeping backups is probably your most important data management task.”

-Everyone

1. Some data backup is better than none.
2. Automated backups are better than manual.
3. Your data are only as safe as your last backup.

Data backup

Best Practice: 3 Copies of datasets



Data storage options

1. Personal computers (PCs) & laptops
2. External storage devices
3. Networked Drives
4. Cloud servers

Storage: PC/laptop

Advantages

Convenient

Disadvantages

Drive failure common

Laptops: susceptible to theft & unintentional damage

Not replicated

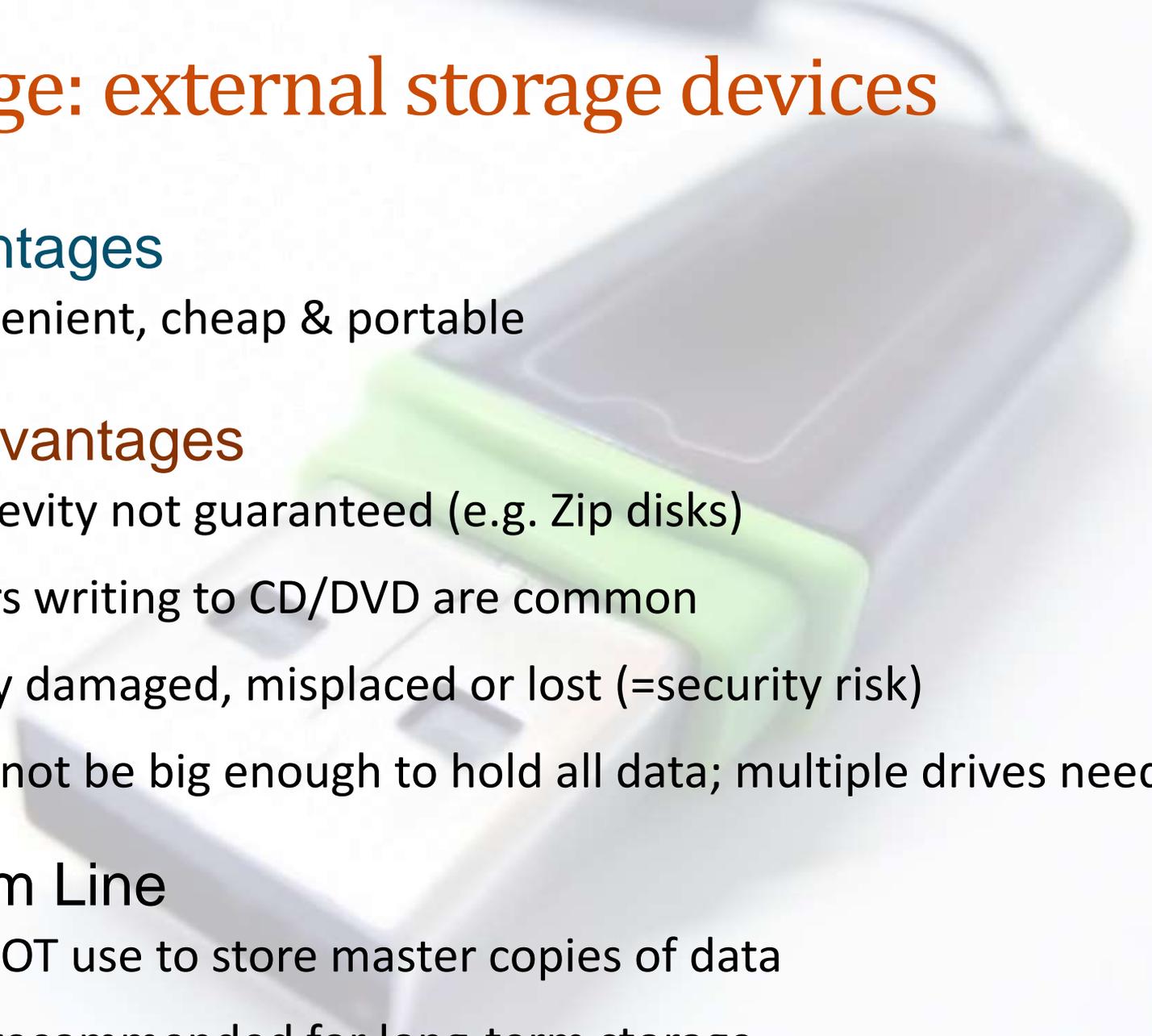
Bottom Line

Do NOT use to store master copies of data

Not a long term storage solution

Back up important data & files regularly

Storage: external storage devices



Advantages

Convenient, cheap & portable

Disadvantages

Longevity not guaranteed (e.g. Zip disks)

Errors writing to CD/DVD are common

Easily damaged, misplaced or lost (=security risk)

May not be big enough to hold all data; multiple drives needed

Bottom Line

Do NOT use to store master copies of data

Not recommended for long-term storage

Storage: networked drives

Advantages

Data in single place, backed up regularly

Replicated storage not vulnerable to loss due to hardware failure

Secure storage minimizes risk of loss, theft, unauthorized use

Available as needed (assuming network avail.)

Disadvantages

Cost may be prohibitive; export control

Bottom Line

Highly recommended for master copies of data

Recommended for long-term storage (~5 years)

Storage: cloud storage

Advantages

Data in single place, backed up regularly

Replicated storage not vulnerable to loss due to hardware failure

Secure storage minimizes risk of loss, theft, unauthorized use

Disadvantages

Cost may be prohibitive

Upload/download bottleneck & fees

Longevity?

Export control

Bottom Line

Possibly recommended for master copies of data

Not recommended for in-process data, large files

Storage: Google Drive for OSU

Advantages

All same advantages of network & cloud storage

File sharing & collaboration w/variable access levels

Unlimited storage (GD), 30 GB non-GD

Automatic version control on GD

Disadvantages

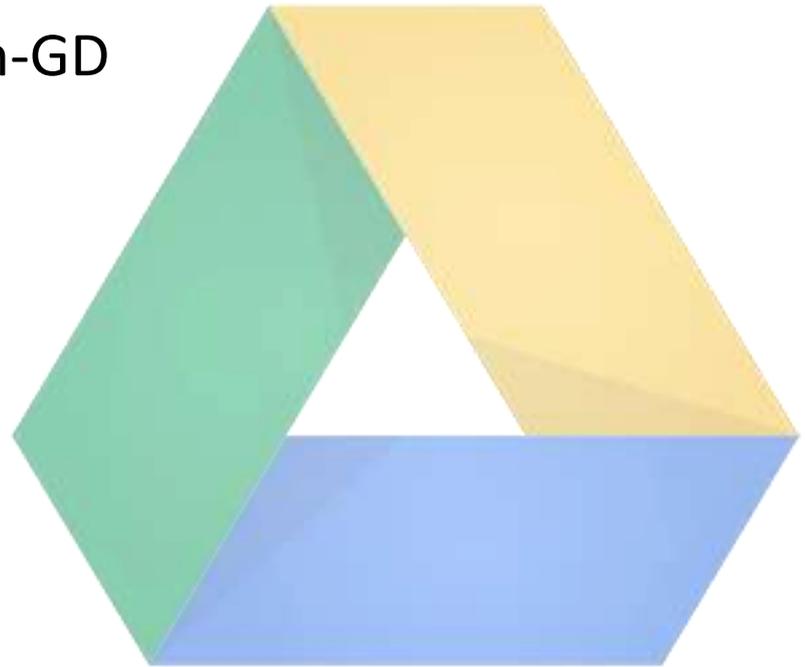
30 GB may not be enough

Upload/download bottleneck

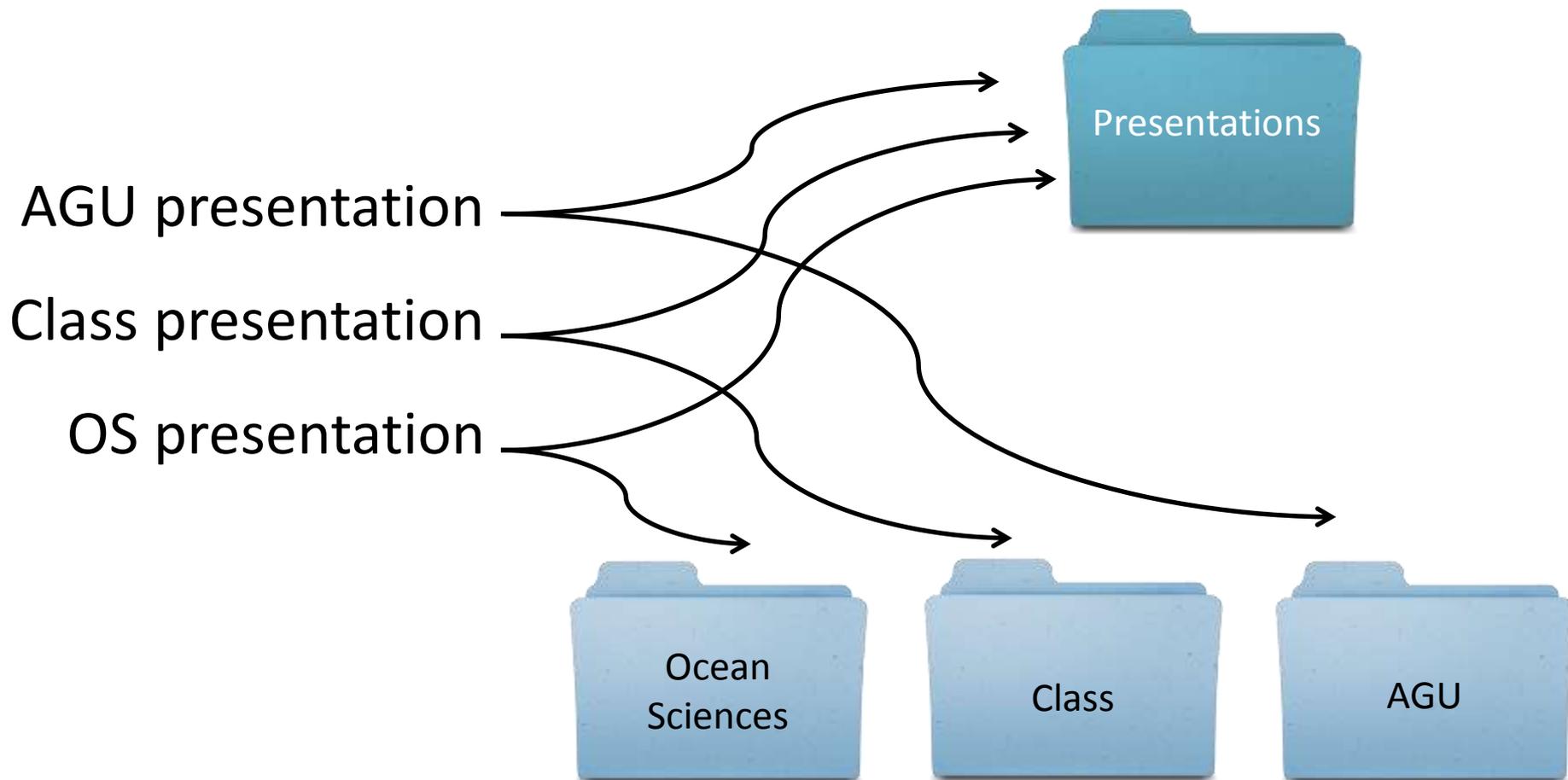
Bottom Line

Possibly recommended for master copies of data

Possibly not recommended for in-process data, large files



Data organization



Data storage options

Local

Computer

External storage

Network server

Remote

Network server

Cloud storage

Google Apps

Box, Dropbox,
etc.

File-naming conventions

File naming strategy?

Filename	Date Modified	Size	Type
data_2010.05.28_test.dat	3:37 PM 5/28/2010	420 KB	DAT file
data_2010.05.28_re-test.dat	4:29 PM 5/28/2010	421 KB	DAT file
data_2010.05.28_re-re-test.dat	5:43 PM 5/28/2010	420 KB	DAT file
data_2010.05.28_calibrate.dat	7:17 PM 5/28/2010	1,256 KB	DAT file
data_2010.05.28_huh??.dat	7:20 PM 5/28/2010	30 KB	DAT file
data_2010.05.28_WTF.dat	9:58 PM 5/28/2010	30 KB	DAT file
data_2010.05.29_aaarrgh.dat	12:37 AM 5/29/2010	30 KB	DAT file
data_2010.05.29_#*\$@*&!!.dat	2:40 AM 5/29/2010	0 KB	DAT file
data_2010.05.29_crap.dat	3:22 AM 5/29/2010	437 KB	DAT file
data_2010.05.29_notbad.dat	4:16 AM 5/29/2010	670 KB	DAT file
data_2010.05.29_woohoo!!.dat	4:47 AM 5/29/2010	1,349 KB	DAT file
data_2010.05.29_USETHISONE.dat	5:08 AM 5/29/2010	2,894 KB	DAT file
analysis_graphs.xls	7:13 AM 5/29/2010	455 KB	XLS file
ThesisOutline!.doc	7:26 AM 5/29/2010	38 KB	DOC file
Notes_Meeting_with_ProfSmith.txt	11:38 AM 5/29/2010	1,673 KB	TXT file
JUNK...	2:45 PM 5/29/2010		Folder
data_2010.05.30_startingover.dat	8:37 AM 5/30/2010	420 KB	DAT file

#OverlyHonestMethods



File naming conventions

s/n, variable

Retain

order

Project_instrument_location_YYYYMMDDhhmmss_e

xtra.ext

Index/grant

conditions

Leading zero!

File naming strategies

Order by date:

1955-04-12_notes_MassObs.docx
1955-04-12_questionnaire_MassObs.pdf
1963-12-15_notes_Gorer.docx
1963-12-15_questionnaire_Gorer.pdf

Order by subject:

Gorer_notes_1963-12-15.docx
Gorer_questionnaire_1963-12-15.pdf
MassObs_notes_1955-04-12.docx
MassObs_questionnaire_1955-04-12.pdf

Order by type:

Notes_Gorer_1963-12-15.docx
Notes_MassObs_1955-04-12.docx
Questionnaire_Gorer_1963-12-15.pdf
Questionnaire_MassObs_1955-04-12.pdf

Forced order with numbering:

01_MassObs_questionnaire_1955-04-12.pdf
02_MassObs_notes_1955-04-12.docx
03_Gorer_questionnaire_1963-12-15.pdf
04_Gorer_notes_1963-12-15.docx

Version control



File Edit View Insert Format Data

Share...

New ▶

Open... ⌘O

Rename...

Make a copy...

Import...

See revision history ⌘⇧G

Spreadsheet settings...

Download as ▶

Publish to the web...

Email collaborators...

Email as attachment...

Print ⌘P

Revision history ✕

Jun 13, 3:30 PM PT
■ Nicole Vasilevsky

May 30, 12:45 PM PT
■ Nicole Vasilevsky

May 29, 7:53 PM PT
■ anonymous

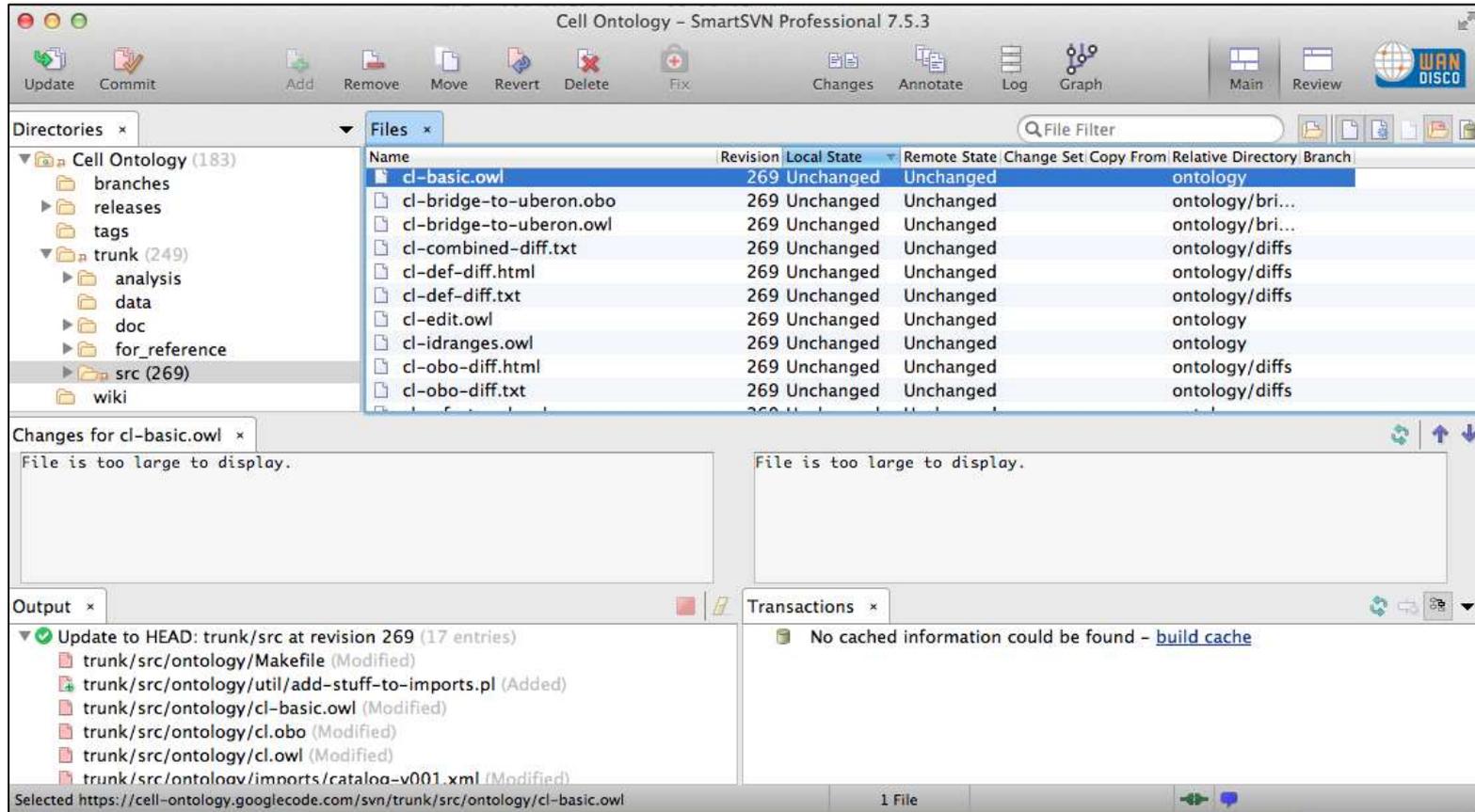
May 29, 9:51 AM PT
■ Nicole Vasilevsky

May 28, 10:14 AM PT
■ Nicole Vasilevsky

May 21, 7:48 AM PT
■ Nathan Urban

May 16, 6:38 PM PT
■ Nicole Vasilevsky

Version control





Random suggestion

Disambiguate yourself

ORCID

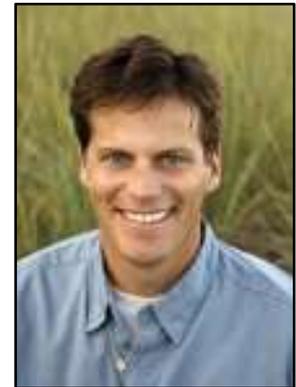
Connecting Research
and Researchers

Open Researcher & Contributor ID



John L. Campbell
Forest Research Ecologist
Oregon State University, Corvallis, OR

John L. Campbell
Forest Research Ecologist
Center for Research on Ecosystem Change
US Forest Service, Durham, NC





[Winter in northeastern North America: a critical period for ecological processes](#)

JL Campbell, MJ Mitchell, PM Groffman... - *Frontiers in Ecology* ..., 2005 - Eco Soc America

Ecological research during winter has historically been a low priority in northeastern North America, an oversight that stems from the commonly accepted notion that there is little biological activity when temperatures drop below freezing. However, recent research has ... Cited by 108 Related articles All 10 versions Cite

[Post-wildfire logging hinders regeneration and increases fire risk](#)

DC Donato, JB Fontaine, JL Campbell, WD Robinson... - *Science*, 2006 - sciencemag.org

... However, our data indicate that delay was responsible for ~10% of the woody fuel present after logging. ← JK Agee, *Fire Ecology* of Pacific Northwest Forests (Island Press, Washington, DC, 1993). ... More in Collections. **Ecology**. Related Jobs from ScienceCareers. ... Cited by 156 Related articles All 63 versions Cite

[Site-level evaluation of satellite-based global terrestrial gross primary production and net primary production monitoring](#)

DP Turner, WD Ritts, WB Cohen... - *Global Change* ..., 2005 - Wiley Online Library

... AA, Running, SW, Zhao, M., Wofsy, SC, Dunn, AL, Law, BE, **Campbell, JL**, Oechel, WC ... 3 Department of Forest **Ecology** and Management, University of Wisconsin, Madison, WI 53706, USA., 4 ... HARV is within the Harvard Forest Long Term **Ecological** Research (LTER) site in ... Cited by 175 Related articles All 25 versions Cite

[Supply-side controls on soil respiration among Oregon forests](#)

JL Campbell, OJ Sun, BE Law - *Global Change Biology*, 2004 - Wiley Online Library

... Additional Information. How to Cite. **Campbell, JL**, Sun, OJ and Law, BE (2004), Supply-side controls on soil respiration among Oregon forests. *Global Change Biology*, 10: 1857-1869. doi: 10.1111/j.1365-2486.2004.00850.x. Author Information. ... Cited by 48 Related articles All 2 versions Cite

[Ecosystem processes and human influences regulate streamflow response to climate change at long-term ecological research sites](#)

JA Jones, JF Creed, KL Hatcher, RJ Warren, MB Adams... - *BioScience*, 2012 - JSTOR

Analyses of long-term records at 35 headwater basins in the United States and Canada indicate that climate change effects on streamflow are not as clear as might be expected, perhaps because of ecosystem processes and human influences. Evapotranspiration was ... Cited by 18 Related articles All 28 versions Cite

[Can fuel-reduction treatments really increase forest carbon storage in the western US by reducing future fire emissions?](#)

JL Campbell, ME Harmon... - *Frontiers in Ecology and* ..., 2011 - Eco Soc America

It has been suggested that thinning trees and other fuel-reduction practices aimed at reducing the probability of high-severity forest fire are consistent with efforts to keep carbon (C) sequestered in terrestrial pools, and that such practices should therefore be rewarded ...

- include patents
- include citations
- Create alert

Another Doppelgänger

Chris Langdon
Studies ocean acidification
OSU HMSC



OSU
Oregon State
UNIVERSITY

Oregon Sea Grant

About Us ▾ Research ▾ Outreach & Engagement ▾ Education ▾ Publications & Vi

Resistance of Pacific Oyster Larvae and Juveniles to Ocean Acidification



Resistance of Pacific Oyster Larvae and Juveniles to the Effects of Ocean Acidification (2014-17)

Chris Langdon
Hatfield Marine Science Center
Coastal Oregon Marine Experiment Station
Oregon State University
2030 SE Marine Science Drive
Newport, Oregon 97365
Phone: 541-867-0231
Fax: 541-867-0345
Email

ROSENSTIEL SCHOOL of
MARINE & ATMOSPHERIC SCIENCE

INFO FOR: GRADUATE + UNDERGRADUATE

About Academic People Research Newsroom Outreach Resources Blog Support RSMAS

Home People Contact Us Chris Langdon

People

Faculty Index:
Department of Atmospheric Sciences
Department of Marine Biology and Ecology
Department of Marine Ecosystems and Society
Department of Marine Geosciences
Department of Ocean Sciences



Chris Langdon
Professor, Department of Marine Biology and Ecology
RSMAS/MBF
University of Miami
4600 Rickenbacker Causeway
Miami, FL 33149

S Gros 211
Tel: 305.421.4614
clangdon@rsmas.miami.edu
Lab Website
Google Scholar Profile

Research Publications Students Courses

Albright R, Mason B, Langdon C (2008) Effect of aragonite saturation state on the settlement and post-settlement growth of *Pavites astrooides* larvae. *Coral Reefs* doi:10.1007/s00338-008-0392-5

Manzello D, Kleypas J, Budd DA, Eakin CM, Glynn PW, Langdon C (2008) Poorly cemented coral reefs of the eastern tropical Pacific: possible insights into reef development in a high-CO₂ world. *Proceedings of the National Academy of Science*, 105(30), doi:10.1073/pnas.0712167105



Chris Langdon
Studies ocean acidification
University of Miami



Metadata

What is metadata?

- Data about data
- Structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource.



NISO, *Understanding Metadata*

Metadata

“The metadata accompanying your data should be written for a user **20 years into the future** -- what does that person need to know to use your data properly? Prepare the metadata for a user who is unfamiliar with your project, methods, or observations.”

Oak Ridge National Laboratory Distributed Active
Archive Center for Biogeochemical Dynamics
(ORNL DAAC)



OAK RIDGE NATIONAL LABORATORY

Managed by UT-Battelle for the Department of Energy

Major metadata standards

Darwin Core | **biological diversity, taxonomy**

Dublin Core | **general**

DDI (Data Documentation Initiative) | **social & behavioral sciences**

DIF (Directory Interchange Format) | **environmental sciences**

EML (Ecological Metadata Language) | **ecology, biology**

ISO 19115 | **geographic data**

01101101 01100101 01110100 01100001
01100100 01100001 01110100 01100001

Metadata examples

Santa Barbara Coastal Long Term Ecological Research (LTER)

[web link](#)

Bureau of Labor Statistics, Consumer Price Index, 1913-1992

[web link](#)



Legal & ethical considerations



Data sharing & reuse

“...digitally formatted scientific data resulting from unclassified research supported wholly or in part by Federal funding should be stored and **publicly accessible** to search, retrieve, and analyze.”

Office of Science and Technology Policy
The White House



How to preserve & share data



How to preserve & share data

Login

ScholarsArchive@OSU

SA@OSU Home



Open Access at OSU

Increase the visibility and impact of your scholarship with ScholarsArchive@OSU.

About ScholarsArchive@OSU

ScholarsArchive@OSU is Oregon State University's digital service for gathering, indexing, making available and storing the scholarly work of the Oregon State University community. It also includes materials from outside the institution in support of the university's land, sun, sea and space grant missions and other research interests.

- [FAQ](#)
- [Open Access](#)

Search ScholarsArchive@OSU

[Advanced Search](#)

Browse

All of ScholarsArchive@OSU

- [Communities & Collections](#)
- [Issue Date](#)
- [Authors](#)
- [Titles](#)
- [Subjects](#)

My Account

- [Login](#)
- [Register](#)

Discover

Author

- [Oregon State University. Extension Service \(2848\)](#)
- [Forest Products Laboratory \(U.S.\) \(1912\)](#)
- [Oregon State University. Agricultural](#)

Data journals



Earth System Science Data
The Data Publishing Journal

| Contact

Home

Online Library ESSD

Recent Final Revised Papers

Volumes and Issues

Special Issues

Full Text Search

Title and Author Search

Online Library ESSD

Alerts & RSS Feeds

General Information

Submission

Review

Production

Subscription

Comment on a Paper

Earth Syst. Sci. Data, 4, 47-73, 2012
www.earth-syst-sci-data.net/4/47/2012/
doi:10.5194/essd-4-47-2012
© Author(s) 2012. This work is distributed
under the Creative Commons Attribution 3.0 License.

Article

Related Articles

Database of diazotrophs in global ocean: abundance, biomass and nitrogen fixation rates

Y.-W. Luo¹, S. C. Doney¹, L. A. Anderson², M. Benavides³, I. Berman-Frank⁴, A. Bode⁵, S. Bonnet⁶, K. H. Boström⁷, D. Böttjer⁸, D. G. Capone⁹, E. J. Carpenter¹⁰, Y. L. Chen¹¹, M. J. Church⁸, J. E. Dore¹², L. I. Falcón¹³, A. Fernández¹⁴, R. A. Foster¹⁵, K. Furuya¹⁶, F. Gómez¹⁷, K. Gundersen¹⁸, A. M. Hynes^{19,*}, D. M. Karl⁸, S. Kitajima¹⁶, R. J. Langlois²⁰, J. LaRoche²⁰, R. M. Letelier²¹, E. Marañón¹⁴, D. J. McGillicuddy Jr.², P. H. Moisander^{22,**}, C. M. Moore²³, B. Mouríño-Carballido¹⁴, M. R. Mulholland²⁴, J. A. Needoba²⁵, K. M. Orcutt¹⁸, A. J. Poulton²⁶, E. Rahav⁴, P. Raimbault⁶, A. P. Rees²⁷, L. Riemann²⁸, T. Shiozaki¹⁶, A. Subramaniam²⁹, T. Tyrrell²³, K. A. Turk-Kubo²², M. Varela⁵, T. A. Villareal³⁰, E. A. Webb⁹, A. E. White²¹, J. Wu³¹, and J. P. Zehr²²

¹Department of Marine Chemistry and Geochemistry, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA

²Department of Applied Ocean Science and Engineering, Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA

³Instituto de Oceanografía y Cambio Global, Universidad de Las Palmas de Gran Canaria, 35017, Las Palmas de Gran Canaria, Spain

⁴The Mina and Everard Goodman Faculty of Life Sciences, Bar-Ilan University, Ramat Gan 52900, Israel

⁵Instituto Español de Oceanografía, Centro Oceanográfico de A Coruña, 15080 A Coruña, Spain

⁶IRD-INSU-CNRS, Laboratoire d'Océanographie Physique et Biogéochimique, UMR 6535, Centre d'Océanologie de Marseille, Aix Marseille Université, France

⁷Department of Natural Sciences, Linnaeus University, 39182 Kalmar, Sweden

⁸School of Ocean and Earth Science and Technology, University of Hawaii, Honolulu, Hawaii 96822, USA

⁹Department of Biological Sciences and Wrigley Institute for Environmental Studies, University of Southern California, Los Angeles, California 90089, USA

¹⁰Romberg Tiburon Center, San Francisco State University, Tiburon, California 94920, USA

¹¹Department of Marine Biotechnology and Resources, National Sun Yat-sen University, Kaohsiung 80424, Taiwan

¹²Department of Land Resources and Environmental Sciences, Montana State University, Bozeman, MT 59717, USA

¹³Laboratorio de Ecología Bacteriana, Instituto de Ecología, Universidad Nacional Autónoma de México, Mexico

¹⁴Laboratorio de Ecología Bacteriana, Instituto de Ecología, Universidad Nacional Autónoma de México, Mexico

Search ESSD

Search

Full Text

Special Issue

MAREDAT - Towards a world atlas of marine plankton functi...

Final Revised Paper

Supplement (79 KB)



Citation

BibTeX
EndNote

Discussion Paper

Published on 2012-02-13

Share



Data management plans



What is a data management plan?

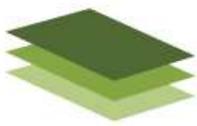
...Physical specimens consist of carbonate biothems and filtered water samples are stored in Spilde or Northup's labs until completion of analyses. At this time they are returned to the federal agency for museum curation or destroyed during analysis. Field notes will be scanned into pdf files, with a copy sent to Carlsbad Caverns National Park or other federal cave manager. SEM images will be saved in the tif format. ...



Sections of a data management plan

1. Types of data
2. Data & metadata standards
3. Archiving & preservation
4. Sharing (access provisions)
5. Transition from collection to reuse

See resources on your handout + use DMPTool



Your go-to
resource for data
management plans

DMPTool
Guidance and Resources for your Data Management Plan

Home About DMP Tool DMP News My Plans Funder Requirements Help

My Data Management Plans

Create a new plan:

Existing plans: 3

Plan name: NSF-ENG test
Solicitation No.:
Funder: NSF-ENG: Engineering
Status: You provided responses for 6 out of 6 questions
Comment:
 Export to: Plain Text Rich Text

Plan name: Pallavi_test
Solicitation No.: 123456789
Funder: NSF-DMR: Materials Research
Status: You provided responses for 2 out of 5 questions
Comment: This is a test plan for Pallavi Dhagat in Materials Sciences
 Export to: Plain Text Rich Text

NSF-BIO: Biological Sciences: 1. Products of Research

What kind of data will be collected, standards employed, and for how long will data be retained?

Progress

Click on a section below to edit it at any time.

- complete
- in progress
- not started

Plan description

1. Products of Research
2. Data Storage and Preservation
3. Data Formats and Metadata
4. Data Dissemination & Policies for Data Sharing and Public Access
5. Roles and Responsibilities

Help box size:

Describe the data to be collected (actual observations) during your research including amount (if known). Name the type of data, the instrument or collection approach, and how the data will be sampled. If actual data are interpreted, note the interpretation. Describe any quality control measures. Also describe the final derivative products and the analysis used including analytical software packages. Indicate what data will be made available and preserved and how long the data will be retained after the life of the project. Consider these questions:

- What data will be generated in the research?

Rich Text Editor

This is just a test.

Contact information

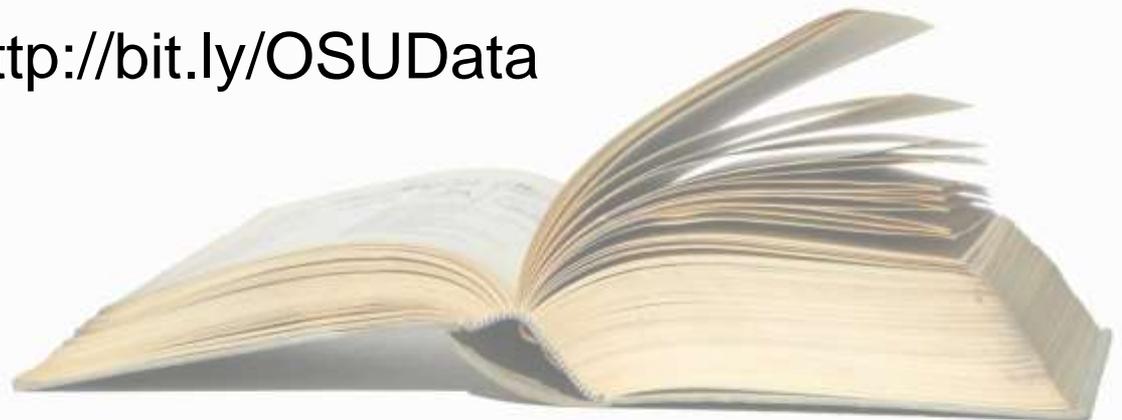
Amanda Whitmire | Data Management
Specialist

amanda.whitmire@oregonstate.edu

Steve Van Tuyl | Data and Digital Repository
Librarian

steve.vantuyl@oregonstate.edu

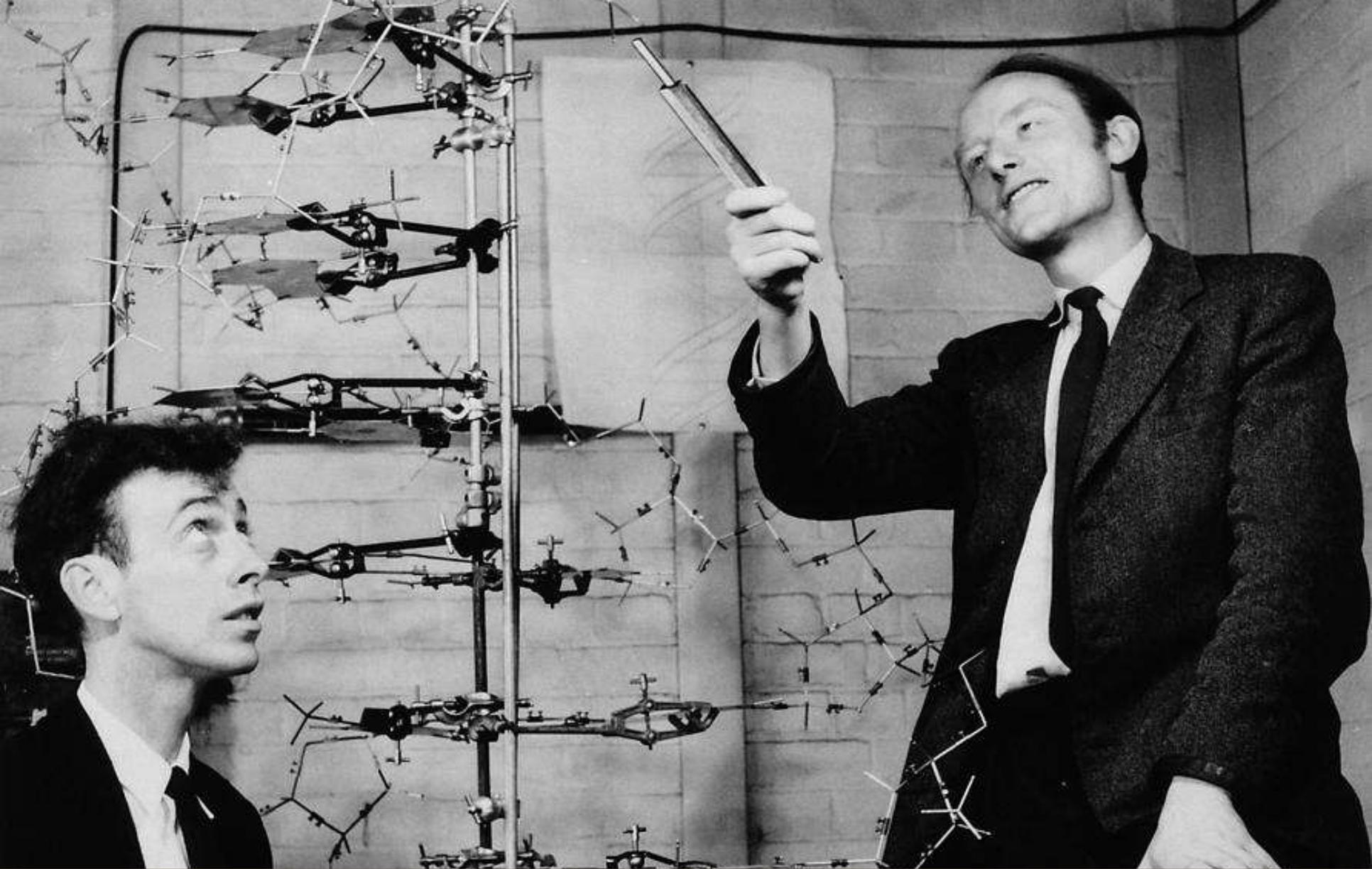
<http://bit.ly/OSUData>



end



Extra / Outdated Slides



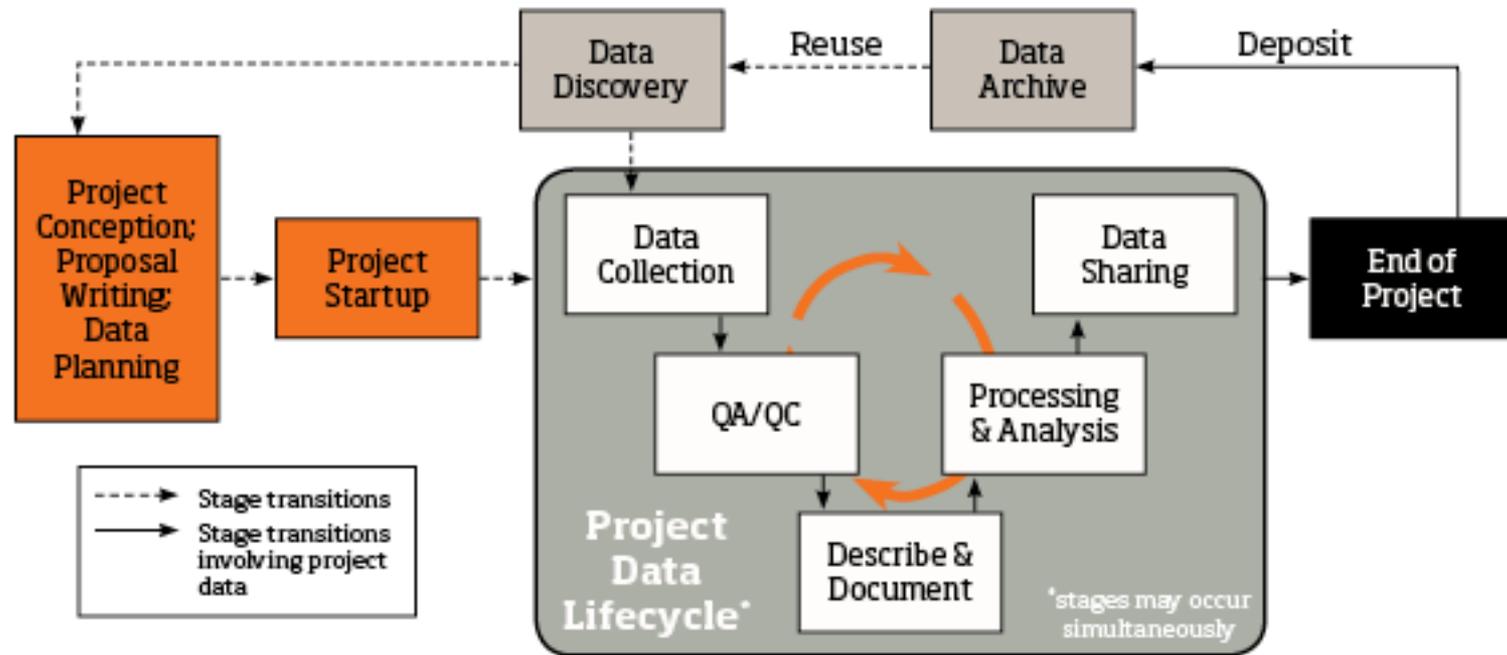
YOU speak for YOUR data

But first, you need to manage it



Data management in the lifecycle

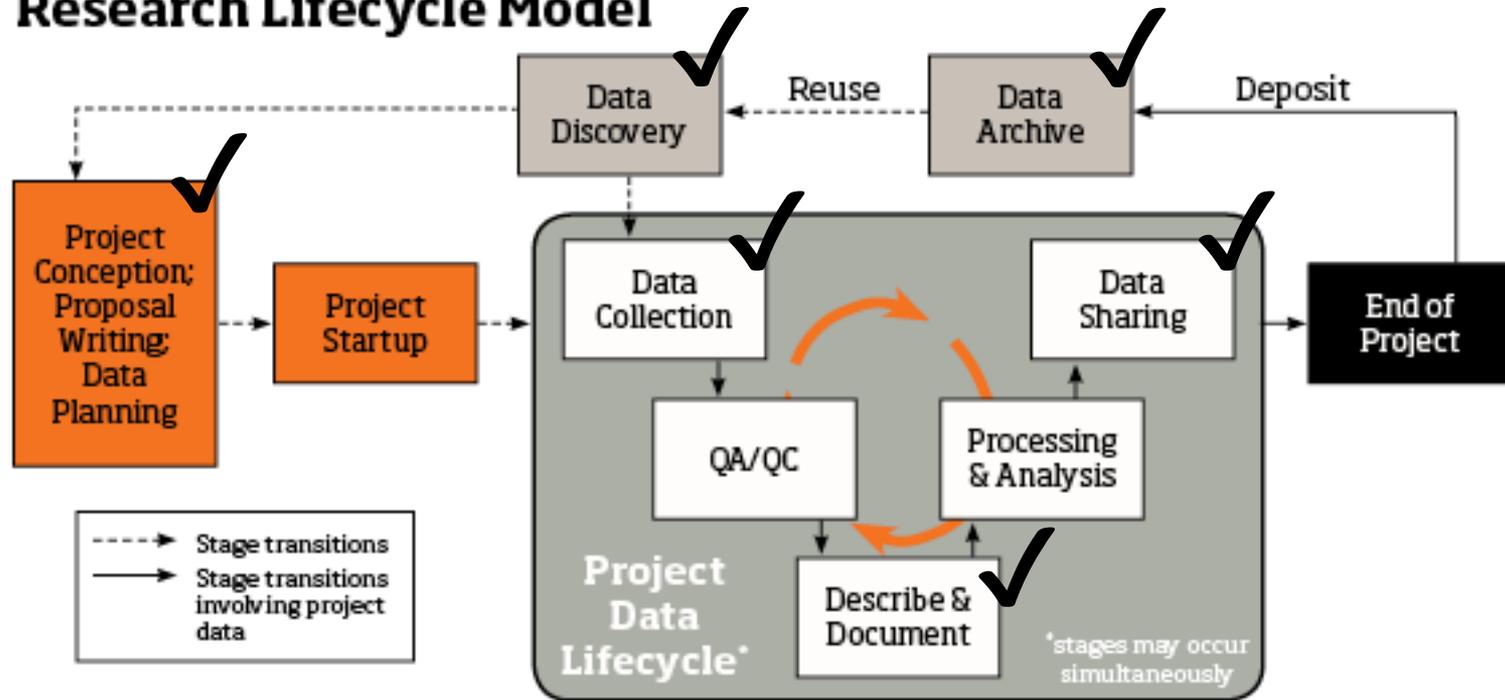
Research Lifecycle Model



Adapted from: <http://dmconsult.library.virginia.edu/lifecycle>
<http://www.icpsr.umich.edu/icpsrweb/content/deposit/guide/#cycle>

How can OSU Libraries help?

Research Lifecycle Model



Adapted from: <http://dmconsult.library.virginia.edu/lifecycle>
<http://www.icpsr.umich.edu/icpsrweb/content/deposit/guide/#cycle>

Copyright – who owns the data?

Copyright applies to text, images, recordings, videos etc. in a digital form, in the same way that it would to analog versions of those works. The code of computer programs (both the human readable source code and the machine readable object code) is protected by copyright as a literary work. Data compilations such as datasets and databases can be protected by copyright in the literary works category, which includes ‘tables’ or ‘compilations’. A table or compilation, consisting of words, figures or symbols (or a combination of these) is protected if it is a literary work and has the required degree of originality.



Types, formats & stages of data

Raw data

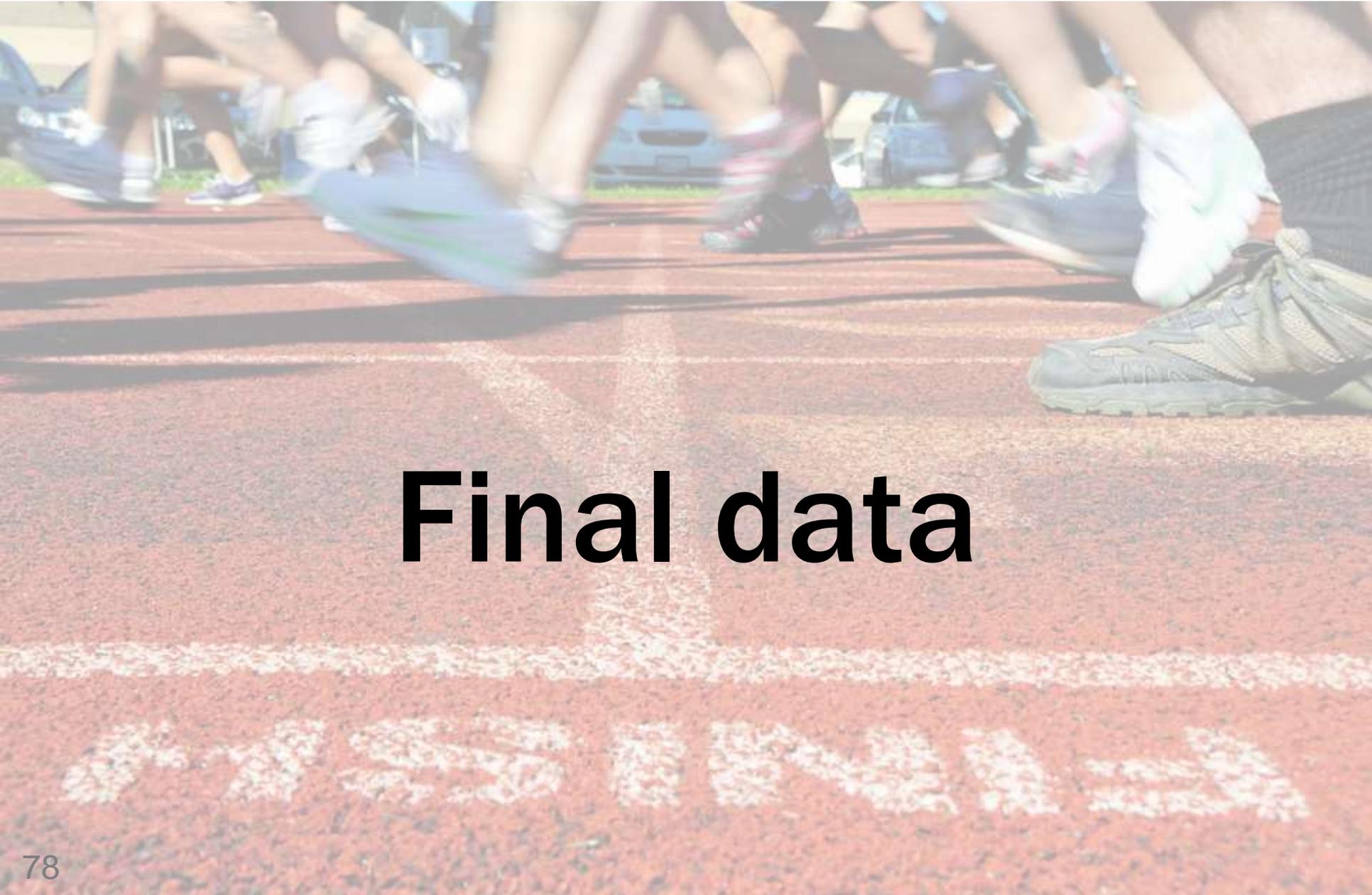


Types, formats & stages of data



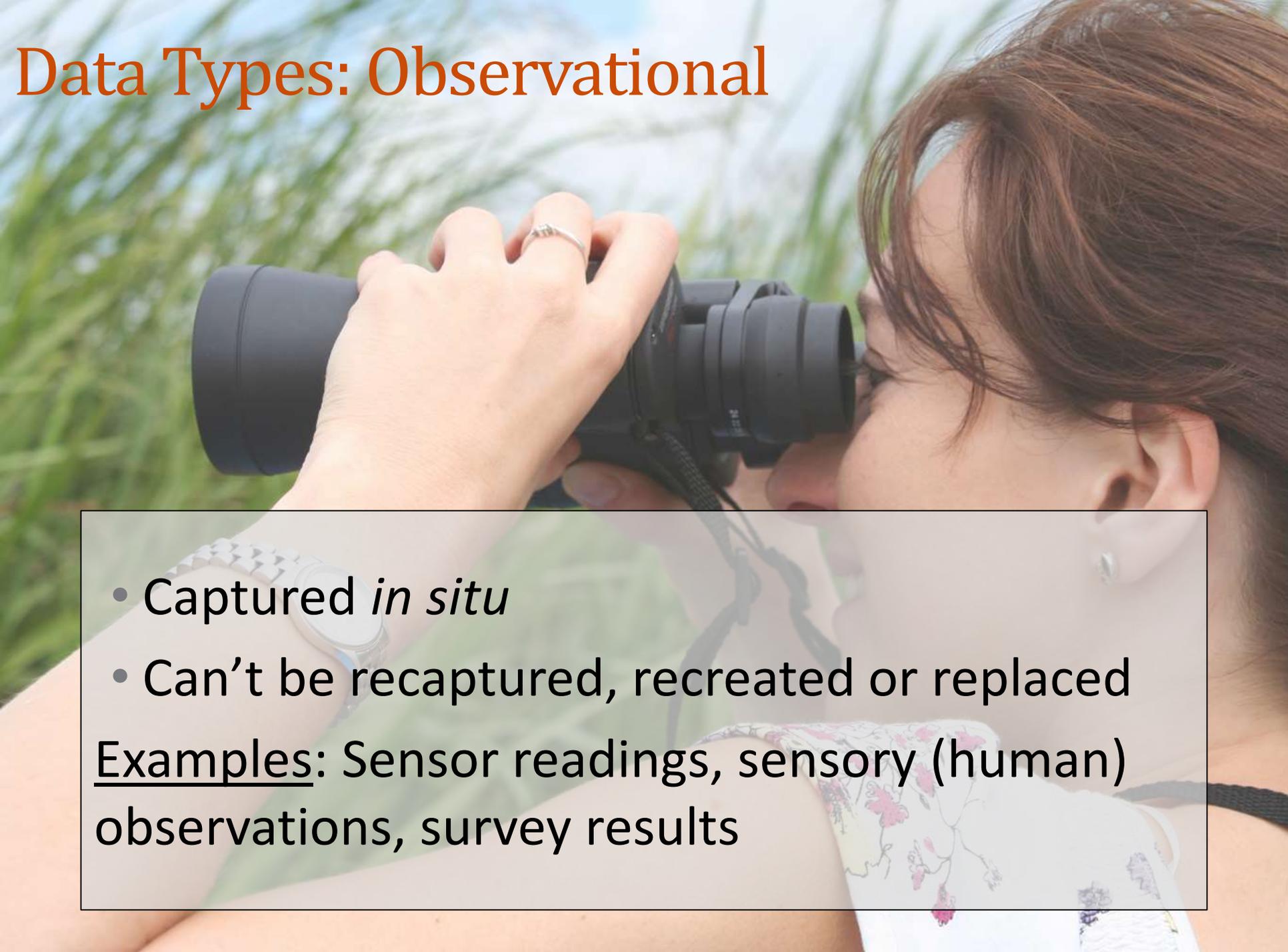
Intermediate data

Types, formats & stages of data



Final data

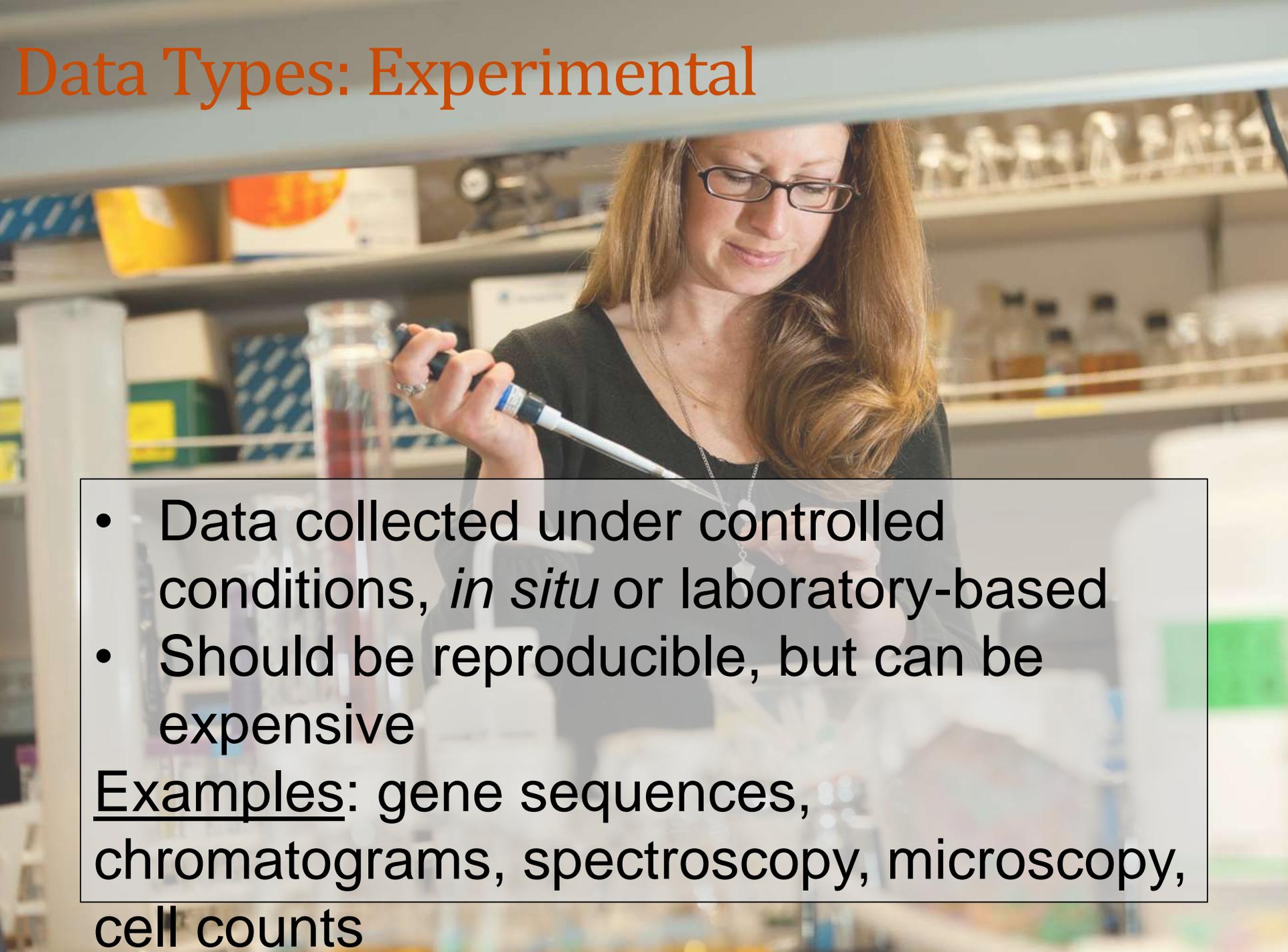
Data Types: Observational

A close-up photograph of a woman with brown hair looking through black binoculars. She is wearing a silver ring on her finger and a watch on her wrist. The background is a blurred natural setting with green grass and a blue sky.

- Captured *in situ*
- Can't be recaptured, recreated or replaced

Examples: Sensor readings, sensory (human) observations, survey results

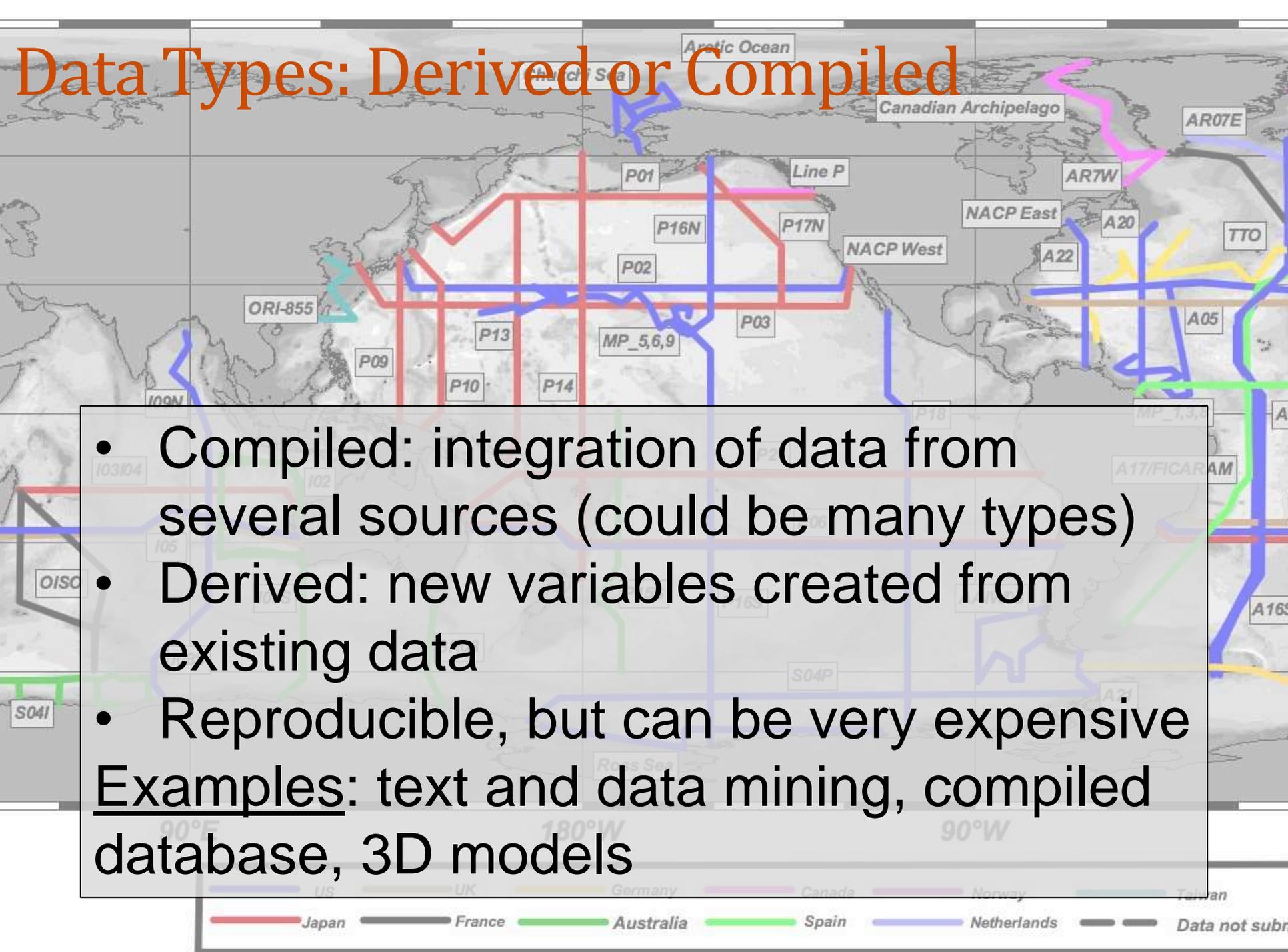
Data Types: Experimental

- 
- A woman with long brown hair and glasses is working in a laboratory. She is wearing a black top and is using a pipette to transfer liquid into a test tube. The background shows shelves with various laboratory equipment and supplies.
- Data collected under controlled conditions, *in situ* or laboratory-based
 - Should be reproducible, but can be expensive

Examples: gene sequences, chromatograms, spectroscopy, microscopy, cell counts

Data Types: Derived or Compiled

- Compiled: integration of data from several sources (could be many types)
 - Derived: new variables created from existing data
 - Reproducible, but can be very expensive
- Examples: text and data mining, compiled database, 3D models



Data Types: Simulation

Hawaii

Hawaii

March 2013

March 2014

March 2015

March 2016

- Results from using a model to study the behavior and performance of an actual or theoretical system
- Models and metadata, where the input can be more important than output data

Examples: climate models, economic models, biogeochemical models

Data storage & preservation



About backups...

Do them.

University of Southampton, School of Electronics and Computer Science
Southampton, UK, 2005

Plan for unexpected events



Metadata demonstration

