Research Data Management Essentials

Kerry Sewell, MSLS

Research Librarian for the Health Sciences

This work is licensed under a Attribution-NonCommercial-ShareAlike 3.0 United States

Objectives

- Describe the current research data management climate
- Outline good data management practices and standards
- Provide resources that can assist in meeting data management needs and requirements

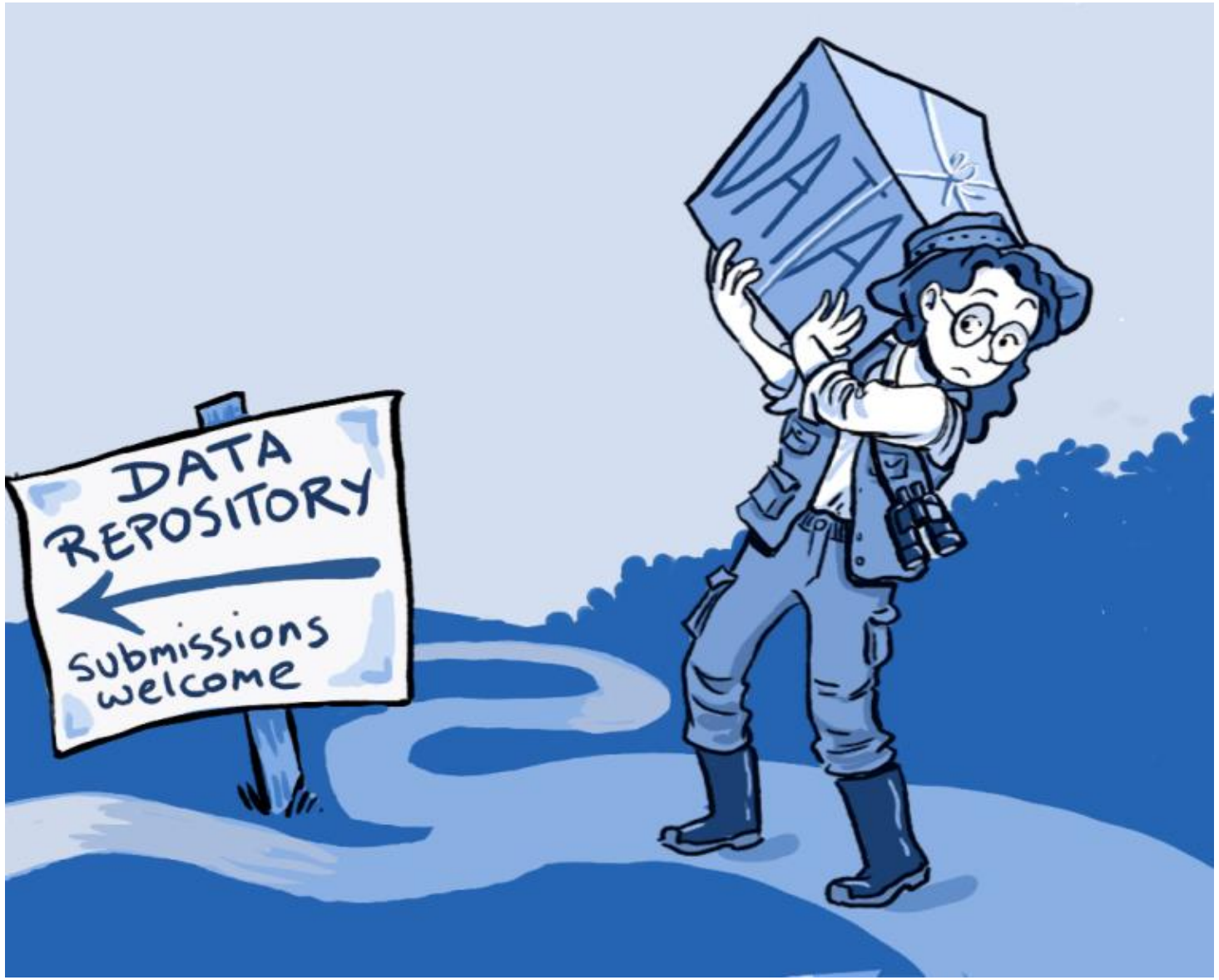
“a scholar’s positive contribution is measured by the sum of the original data that he contributes. Hypotheses come and go but data remain. Theories desert us, while data defend us. They are our true resources, our real estate, and our best pedigree. In the eternal shifting of things, only they will save us from the ravages of time and from the forgetfulness or injustice of men.”

(Santiago Ramón y Cajal, 1897)

y Cajal, S. R. (2004). *Advice for a young investigator*. MIT Press.

Current Research Data Management Climate

- Federal and foundation funders
- Publishers
- Focus on Rigor and Reproducibility



NIH and Data Sharing

NIH Data Sharing Policy (2003)

Helsinki Open Transport Data Manifesto. Created by Epsi Platform. CC-BY 2.0

NIH and Data Sharing

OSTP Memo, Feb. 23, 2013:

"Each agency's public access plan shall:

Ensure that all extramural researchers receiving Federal grants and contracts for scientific research and intramural researchers develop data management plans, as appropriate, describing how they will provide for long-term preservation of, and access to, scientific data in digital formats resulting from federally funded research, or explaining why long-term preservation and access cannot be justified"

Helsinki Open Transport Data Manifesto. Created by Epsi Platform. CC-BY 2.0

NIH Genomic Data Sharing Policy

Large scale data (2014):

GWAS, SNPs, genome sequence, transcriptomic, metagenomic, epigenomic, gene expression

Requirements:

- Genomic data sharing plan
- Shared no later than date of publication

Helsinki Open Transport Data Manifesto. Created by Epsi Platform. CC-BY 2.0

NIH and Data Sharing

DRAFT NIH Policy for Data Management and Sharing released (2019)

* The draft guidance is anticipated to be fully implemented by 2022

Helsinki Open Transport Data Manifesto. Created by Epsi Platform. CC-BY 2.0

NIH and Data Sharing

- All grant applications, regardless of amount of direct costs requested, must include a Data Management and Sharing Plan
- The individual Institutes, Centers, and Offices will fully assess the Data Management and Sharing Plan for its appropriateness and completeness as part of the review process
- Compliance with the Data Management Plan becomes part of the Terms and Conditions of the Award and failure to comply with the Data Management Plan may result in enforcement actions

Helsinki Open Transport Data Manifesto. Created by Epsi Platform. CC-BY 2.0

National Institutes of Health

NSF and Data Sharing

- Required Data Management Plan for full proposals (2011)

Helsinki Open Transport Data Manifesto. Created by Epsi Platform. CC-BY 2.0

Data Sharing and Other Funders

Publisher Requirements

*"Refusal to share data…in accordance with this policy will be grounds for rejection…**must specify that data are deposited publicly** and list the name(s) of repositories along with **digital object identifiers or accession numbers**"*

Publisher Requirements

- *"All data necessary to **understand**, **assess**, and **extend** the conclusions of the manuscript must be available*

http://www.sciencemag.org/site/feature/contribinfo/prep/gen_info.xhtml

Publisher Requirements

The NEW ENGLAND JOURNAL of MEDICINE

HOME | ARTICLES & MULTIMEDIA | ISSUES | SPECIALTIES & TOPICS | FOR AUTHORS | CME

EDITORIAL

Sharing Clinical Trial Data — A Proposal from the International Committee of Medical Journal Editors

Darren B. Taichman, M.D., Ph.D.Joyce Backus, M.S.L.S.Christopher Baethge, M.D.Howard Bauchner, M.D.Peter W. de Leeuw, M.D.Jeffrey M. Drazen, M.D.John Fletcher, M.B., B.Chir., M.P.H.Frank A. Frizelle, M.B., Ch.B., F.R.A.C.S.Trish Groves, M.B., B.S., M.R.C.Psych.Abraham Haileamlak, M.D.Astrid James, M.B., B.S.Christine Laine, M.D., M.P.H.Larry Peiperl, M.D.Anja Pinborg, M.D.Peush Sahni, M.B., B.S., M.S., Ph.D.Sinan Wu, M.D.

N Engl J Med 2016; 374:384-386 | January 28, 2016 | DOI: 10.1056/NEJMe1515172

Share:

Article | References | Citing Articles (3) | Metrics

The International Committee of Medical Journal Editors (ICMJE) believes that there is an ethical obligation to responsibly share data generated by interventional clinical trials because participants have put themselves at risk. In a growing consensus, many funders around the world — foundations, government agencies, and industry — now mandate data sharing. Here we outline the ICMJE's proposed requirements to help meet this obligation. We encourage feedback on the proposed requirements. Anyone can provide feedback at www.icmje.org by 18 April 2016.

The ICMJE defines a clinical trial as any research project that prospectively assigns people or a group of people to an intervention, with or without concurrent comparison or control groups, to study the cause-and-effect relationship between a health-related intervention and a health outcome. Further details may be found in the *Recommendations for the Conduct, Reporting, Editing and Publication of Scholarly Work in Medical Journals* at www.icmje.org.

As a condition of consideration for publication of a clinical trial report in our member journals, the ICMJE proposes to require authors to share with others the deidentified individual-patient data (IPD) underlying the results presented in the article (including tables, figures, and appendices or supplementary material) no later than 6 months after publication. The data underlying the results are defined as the IPD required to reproduce the article's findings, including necessary metadata. This requirement will go into effect for clinical trials that begin to enroll participants beginning 1 year after the ICMJE adopts its data-sharing requirements. (The ICMJE plans to adopt data-sharing requirements after considering feedback received to the proposals made here.)

http://www.nejm.org/doi/full/10.1056/NEJMe1515172#t=article

Publisher Requirements

The NEW ENGLAND JOURNAL of MEDICINE

HOME | ARTICLES & MULTIMEDIA | ISSUES | SPECIALTIES & TOPICS | FOR AUTHORS | CME

EDITORIAL

Sharing Clinical Trial Data — A Proposal from the International Committee of Medical Journal Editors

Darren B. Taichman, M.D., Ph.D.Joyce Backus, M.S.L.S.Christopher Baethge, M.D.Howard Bauchner, M.D.Peter W. de Leeuw, M.D.Jeffrey M. Drazen, M.D.John Fletcher, M.B., B.Chir., M.P.H.Frank A. Frizelle, M.B., Ch.B., F.R.A.C.S.Trish Groves, M.B., B.S., M.R.C.Psych.Abraham Haileamlak, M.D.Astrid James, M.B., B.S.Christine Laine, M.D., M.P.H.Larry Peiperl, M.D.Anja Pinborg, M.D.Peush Sahni, M.B., B.S., M.S., Ph.D.Sinan Wu, M.D.
N Engl J Med 2016; 374:384-386 | January 28, 2016 | DOI: 10.1056/NEJMe1515172

Share:

Article | References | Citing Articles (3) | Metrics

The International Committee of Medical Journal Editors (ICMJE) believes that there is an ethical obligation to responsibly share data generated by interventional clinical trials because participants have put themselves at risk. In a growing consensus, many funders around the world — foundations, government agencies, and industry — now mandate data sharing. Here we outline the ICMJE's proposed requirements to help meet this obligation. We encourage feedback on the proposed requirements. Anyone can provide feedback at www.icmje.org by 18 April 2016.

The ICMJE defines a clinical trial as any research project that prospectively assigns people or a group of people to an intervention, with or without concurrent comparison or control groups, to study the cause-and-effect relationship between a health-related intervention and a health outcome. Further details may be found in the *Recommendations for the Conduct, Reporting, Editing and Publication of Scholarly Work in Medical Journals* at www.icmje.org.

As a condition of consideration for publication of a clinical trial report in our member journals, the ICMJE proposes to require authors to share with others the deidentified individual-patient data (IPD) underlying the results presented in the article (including tables, figures, and appendices or supplementary material) no later than 6 months after publication. The data underlying the results are defined as the IPD required to reproduce the article's findings, including necessary metadata. This requirement will go into effect for clinical trials that begin to enroll participants beginning 1 year after the ICMJE adopts its data-sharing requirements. (The ICMJE plans to adopt data-sharing requirements after considering feedback received to the proposals made here.)

The NEW ENGLAND JOURNAL of MEDICINE

HOME | ARTICLES & MULTIMEDIA | ISSUES | SPECIALTIES & TOPICS | FOR AUTHORS | CME

Perspective

Toward Fairness in Data Sharing

The International Consortium of Investigators for Fairness in Trial Data Sharing
N Engl J Med 2016; 375:405-407 | August 4, 2016 | DOI: 10.1056/NEJMp1605654

Comments open through August 10, 2016

Article | References | Citing Articles (1) | Comments (16) | Metrics

The International Committee of Medical Journal Editors (ICMJE) has proposed a plan for sharing data from randomized, controlled trials (RCTs) that will require, as a condition of acceptance of trial results for publication, that authors make publicly available the deidentified individual patient data underlying the analyses reported in an article.[1] Before any data-sharing policy is enacted, we believe there is a need for the ICMJE, trialists, and other stakeholders to discuss the potential benefits, risks, and opportunity costs, as well as whether the same goals can be achieved by simpler means. Although we believe there are potential benefits to sharing data (e.g., occasional new discoveries), we believe there are also risks (e.g., misleading or inaccurate analyses and analyses aimed at unfairly discrediting or undermining the original publication) and opportunity costs (e.g., the ICMJE proposal would have enormous direct costs and would probably divert resources, both financial and human, from the actual conduct of trials). In 2010 alone, results of more than 27,000 RCTs were published.[2] We believe consideration needs to be given to whether it is worthwhile to undertake data sharing for all published trials or just for those whose results are under

http://www.nejm.org/doi/full/10.1056/NEJMe1515172#t=article

Publisher Requirements

THE WATCHDOGS

New science data-sharing rules are two scoops of disappointment

By ADAM MARCUS @armarcus *and* IVAN ORANSKY @ivanoransky / JUNE 6, 2017

APSTOCK

https://www.statnews.com/2017/06/06/data-sharing-rules-disappoint/

Publisher Requirements

ICMJE INTERNATIONAL COMMITTEE of MEDICAL JOURNAL EDITORS

- As of 1 July 2018 manuscripts submitted to ICMJE journals that report the results of clinical trials must contain a data sharing statement
- Clinical trials that begin enrolling participants on or after 1 January 2019 must include a data sharing plan in the trial's registration. If the data sharing plan changes after registration this should be reflected in the statement submitted and published with the manuscript, and updated in the registry record.

http://www.icmje.org/recommendations/browse/publishing-and-editorial-issues/clinical-trial-registration.html

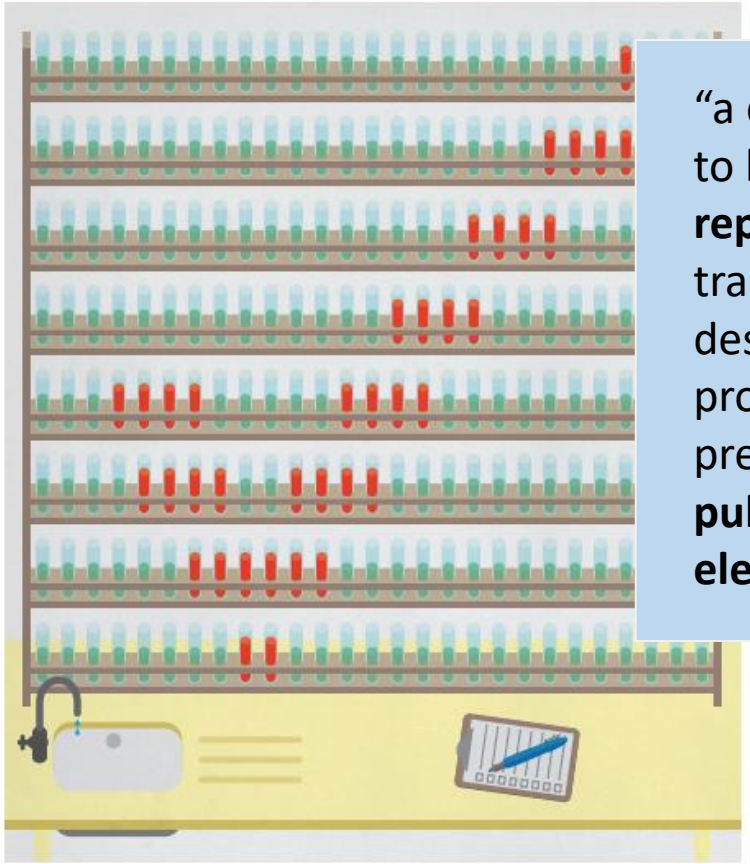
Rigor and Reproducibility

- **New guidelines for grants started January 25, 2016**
  - Scientific premise must describe strengths/weaknesses of prior research
  - Scientific rigor to ensure robust/unbiased experimental design, methodology, analysis, interpretation, reporting of results
  - Consideration of relevant biological variables
  - Authentication of key biological/chemical resources

Rigor and Reproducibility

- **New guidelines for grants started January 25, 2016**
  - Scientific premise must describe strengths/weaknesses of prior research
  - Scientific rigor to ensure robust/unbiased experimental design, methodology, analysis, interpretation, reporting of results

    Full transparency in reporting experimental details
  - Consideration of relevant biological variables
  - Authentication of key biological/chemical resources



NIH plans to enhance reproducibility

Francis S. Collins and **Lawrence A. Tabak** discuss initiatives that the US National Institutes of Health is exploring to restore the self-correcting nature of preclinical research.

A growing chorus of concern, from scientists and laypeople, contends that the complex system for ensuring

shorter term, however, the checks and balances that once ensured scientific fidelity have been hobbled. This has compromised

outnumbered by the hundreds of thousands published each year in good faith.

in such journals, including promotion and tenure, and in extreme circumstances, cash rewards[6].

Then there is the problem of what is not published. There are few venues for researchers to publish negative data or papers that point out scientific flaws in previously published work. Further compounding the problem is the difficulty of accessing unpublished data — and the failure of funding agencies to establish or enforce policies that insist on data access.

PRECLINICAL PROBLEMS

Reproducibility is potentially a problem in all scientific disciplines. However, human clinical trials seem to be less at risk because they are already governed by various regulations that stipulate rigorous design and independent oversight — including randomization, blinding, power estimates, pre-registration of outcome measures in standardized, public databases such as ClinicalTrials.gov and oversight by institutional review boards and data safety monitoring boards. Furthermore, the clinical trials community has taken important steps towards adopting standard reporting elements[7].

"a complex array of other factors seems to have contributed to the **lack of reproducibility**. Factors include poor training of researchers in experimental design; increased emphasis on making provocative statements rather than presenting technical details; and **publications that do not report basic elements of experimental design**"

http://www.nature.com/news/policy-nih-plans-to-enhance-reproducibility-1.14586

National Institutes of Health

ClinicalTrials.gov

First Received Date ICMJE	June 20, 2013
Last Updated Date	September 21, 2015
Start Date ICMJE	October 2013
Primary Completion Date	July 2015 (final data collection date for primary outcome measure)
Current Primary Outcome Measures ICMJE **(submitted: July 1, 2013)**	• For Phase I of the Study: Metrics Used to Understand Diabetes Control [Time Frame: 4 months] [Designated as safety issue: No] Identification of common factors patients use to understand their diabetes and diabetes control via a qualitative analysis of the patient interv • For Phase II of the Study: Change in Hemoglobin A1C [Time Frame: 6 months following enrollment] [Designated as safety issue: No] Change in A1C between enrollment and 6-months compared between study arms.
Original Primary Outcome Measures ICMJE **(submitted: June 24, 2013)**	Metrics Used to Understand Diabetes Control [Time Frame: 4 months] [Designated as safety issue: No] Identification of common factors patients use to understand their diabetes and diabetes control via a qualitative analysis of the patient intervie
Change History	Complete list of historical versions of study NCT01886170 on ClinicalTrials.gov Archive Site
Current Secondary Outcome Measures ICMJE **(submitted: July 1, 2013)**	• For Phase I of the study: Feedback on alternative formats [Time Frame: 4 months] [Designated as safety issue: No] qualitative and quantitative analysis of the feedback received on the alternative communication formats reviewed with participants during th • For Phase II of the Study: Understanding of diabetes control [Time Frame: At the time of enrollment] [Designated as safety issue: No] Accuracy of participant knowledge of level of current diabetes control
Original Secondary Outcome Measures ICMJE **(submitted: June 24, 2013)**	Feedback on alternative formats [Time Frame: 4 months] [Designated as safety issue: No] qualitative and quantitative analysis of the feedback received on the alternative communication formats reviewed with participants during the ir

The Final Rule

Data Elements Required in Final Rule	Provision No. in 42 CFR 11.48(a)	ClinicalTrials.gov PRS Pre-Final Rule Status		Comments
		Required	Optional	
Other measure(s)			X	Sub-element of Baseline Measure Information, (2)(iii). Any other measure(s) that were assessed at baseline and are used in the analysis of the primary outcome measure(s).
Name and Description of the Measure, including any categories that are used to submit Baseline Measure Data	(2)(iii)(A)	X		
Measure Type and Measure of Dispersion	(2)(iii)(B)	X		
Unit of Measure	(2)(iii)(C)	X		
Baseline Measure Data	(2)(iv)	X		
Number of Baseline Participants (and Units)	(2)(v)			If different from Overall Number of Baseline Participants or Overall Number of Units Analyzed
Outcomes and Statistical Analyses				
Outcome Measure Arm/Group Information	(3)(i)	X		
Analysis Population Information	(3)(ii)	X		
Number of Participants Analyzed	(3)(ii)(A)	X		
Number of Units Analyzed	(3)(ii)(B)	X		If the analysis is based on a unit other than participants, a description of the unit of analysis (e.g., eyes, lesions)
Analysis Population Description	(3)(ii)(C)		X	If Number of Participants Analyzed or Number of Units Analyzed differs from the number of human subjects or units assigned to the arm
Outcome Measure Information	(3)(iii)	X		
Name of the Specific Outcome Measure	(3)(iii)(A)	X		
Description of the Metric Used	(3)(iii)(B)		X	
Time Point(s) at which the Measurement was Assessed	(3)(iii)(C)	X		
Outcome Measure Type	(3)(iii)(D)	X		
Measure Type and Measure of Dispersion or Precision	(3)(iii)(E)	X		

Animal Research

OPEN ACCESS Freely available online

PLoS one

Survey of the Quality of Experimental Design, Statistical Analysis and Reporting of Research Using Animals

Carol Kilkenny[1]*, Nick Parsons[2], Ed Kadyszewski[3], Michael F. W. Festing[4], Innes C. Cuthill[5], Derek Fry[6], Jane Hutton[7], Douglas G. Altman[8]

1 The National Centre for the Replacement, Refinement and Reduction of Animals in Research, London, United Kingdom, 2 Warwick Medical School, University of Warwick, Coventry, United Kingdom, 3 Pfizer Global Research and Development, Groton, Connecticut, United States of America, 4 Animal Procedures Committee, London, United Kingdom, 5 School of Biological Sciences, University of Bristol, Bristol, United Kingdom, 6 Animals Scientific Procedures Inspectorate, Home Office, Shrewsbury, United Kingdom, 7 Department of Statistics, University of Warwick, Coventry, United Kingdom, 8 Centre for Statistics in Medicine, University of Oxford, Oxford, United Kingdom

Detailed information collected from 271 publications

- 59% stated hypothesis and number/characteristics of animals
- 13% used randomization
- 14% used blinding
- 30% of publications that used statistical methods did not describe methods

http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0007824

Reproducibility of Preclinical Research

- Scientists in haematology and oncology departments at Amgen tried to confirm findings from 53 “landmark” studies
- Findings confirmed in only 6 (11%) cases.

http://www.nature.com/nature/journal/v483/n7391/full/483531a.html

Rigor and Reproducibility

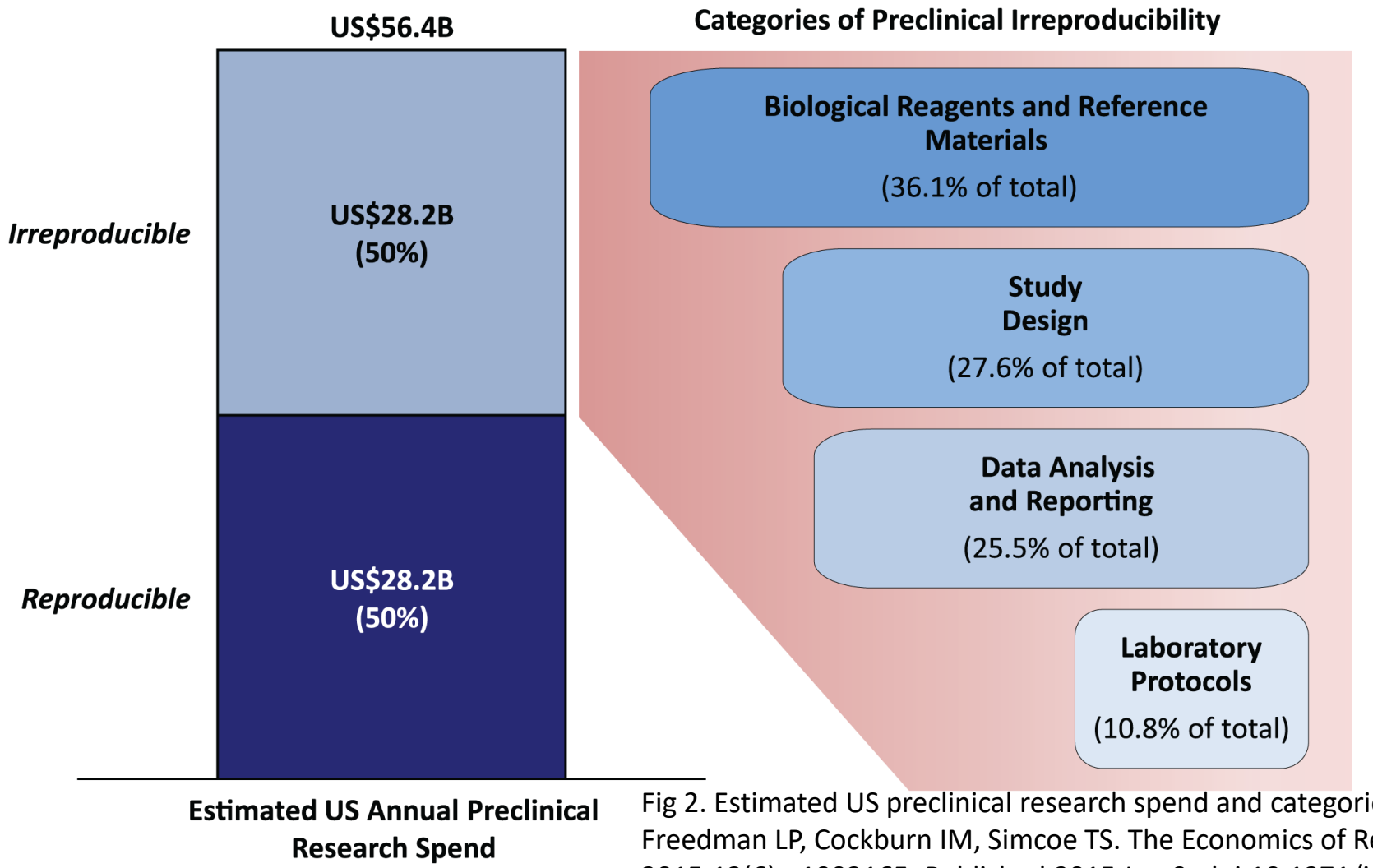


Fig 2. Estimated US preclinical research spend and categories of errors that contribute to irreproducibility. Freedman LP, Cockburn IM, Simcoe TS. The Economics of Reproducibility in Preclinical Research. *PLoS Biol*. 2015;13(6):e1002165. Published 2015 Jun 9. doi:10.1371/journal.pbio.1002165

Principles and Guidelines for Reporting Preclinical Research

Joint workshop June 2014: NIH, NPG, Science

Consensus from journal editors:

- Rigorous statistical analysis
- **Transparency in reporting**
- **Data and material sharing**
- Consideration of refutations
- Consider establishing best practice guidelines for:
  - Image based data
  - Antibodies
  - Cell lines
  - Animals

http://www.nih.gov/about/reporting-preclinical-research.htm

National Institutes of Health

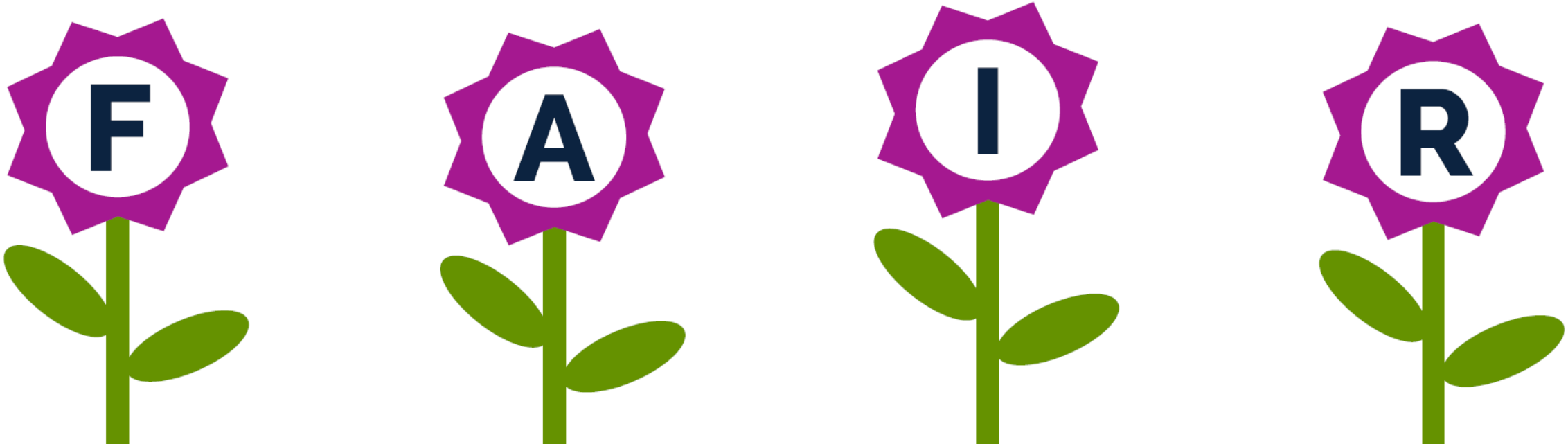
NIH Rigor and Reproducibility

- **New guidelines for grants started January 25, 2016**
  - Scientific premise must describe strengths/weaknesses of prior research
  - Scientific rigor to ensure robust/unbiased experimental design, methodology, analysis, interpretation, reporting of results

    Full transparency in reporting experimental details
  - Consideration of relevant biological variables
  - Authentication of key biological/chemical resources

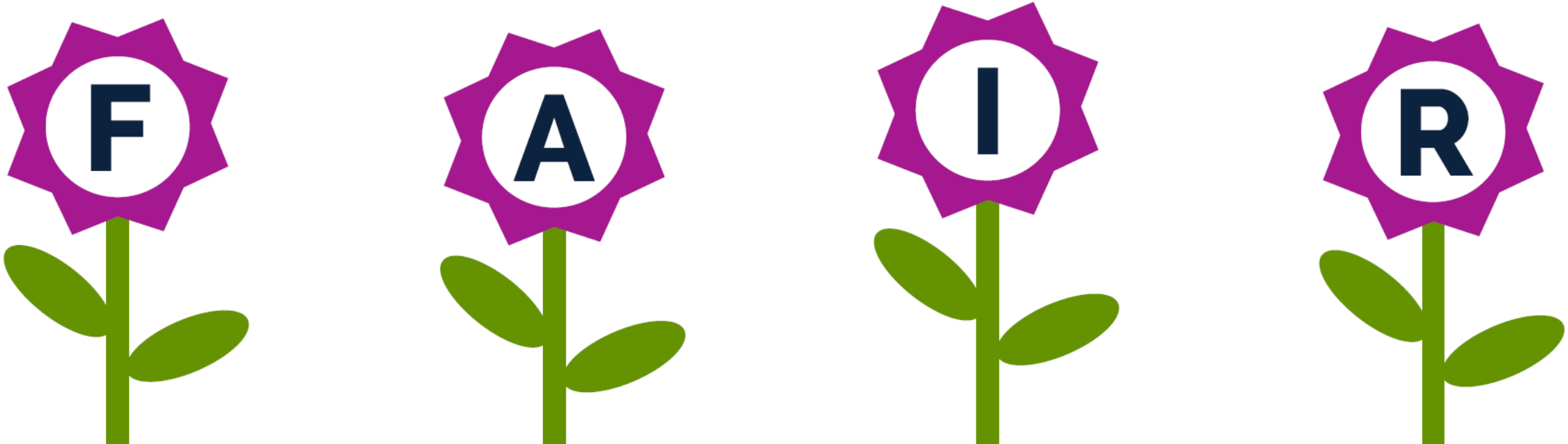
http://grants.nih.gov/reproducibility/index.htm

Findable **A**ccessible **I**nteroperable **R**eusable



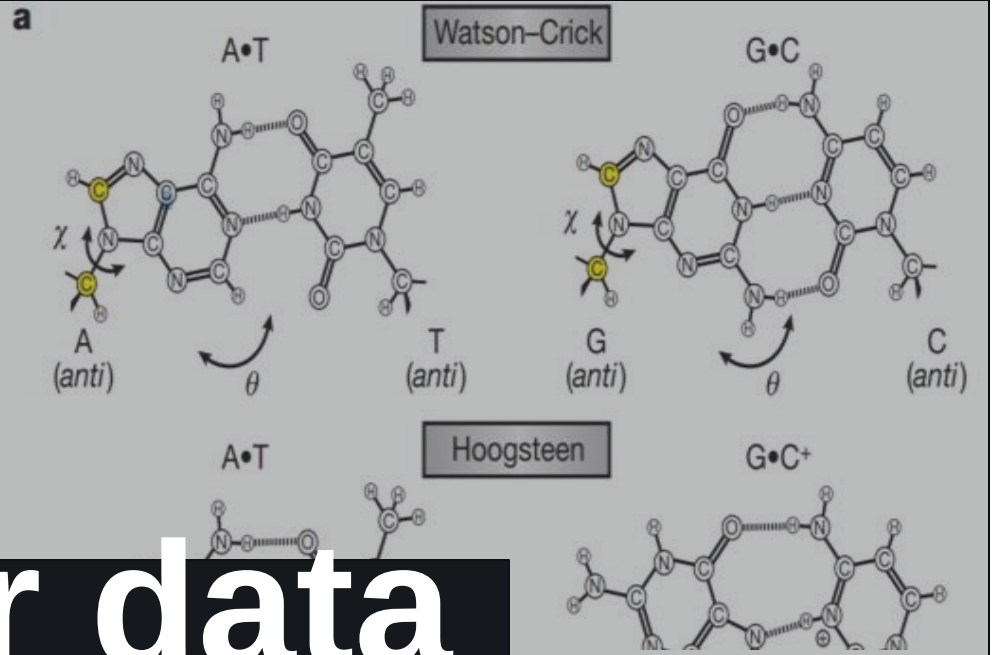
Documentation, documentation, documentation

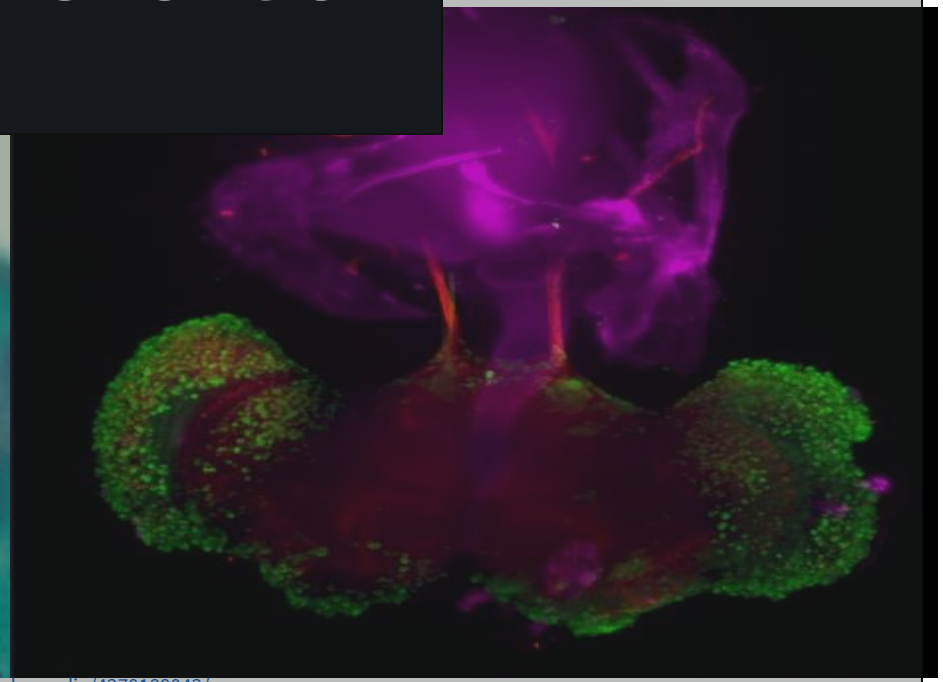
Findable **A**ccessible **I**nteroperable **R**eusable



Your data

	A	B	C	D	E
1	=QUERY('Institutional Table 2010-2011'!A1:K118,"Select C,D,F,H")			rospects	
2	Oxford	100	10.00	78.97	
3	Cambridge	97	9.46	82.66	
4	St Andrews	87	6.55	74.06	
5	Warwick	84	8.76	76.11	
6	London School of Economics	82.4	7.62	86.10	
7	UCL	81.5	8.34	78.70	
8	Edinburgh	78.3	9.26	74.25	
9	Imperial College	77.9	8.82	83.92	
10	Bath	75.6	5.54		
11	Loughborough	74.6	5.72		
12	York	74.4	7.48		



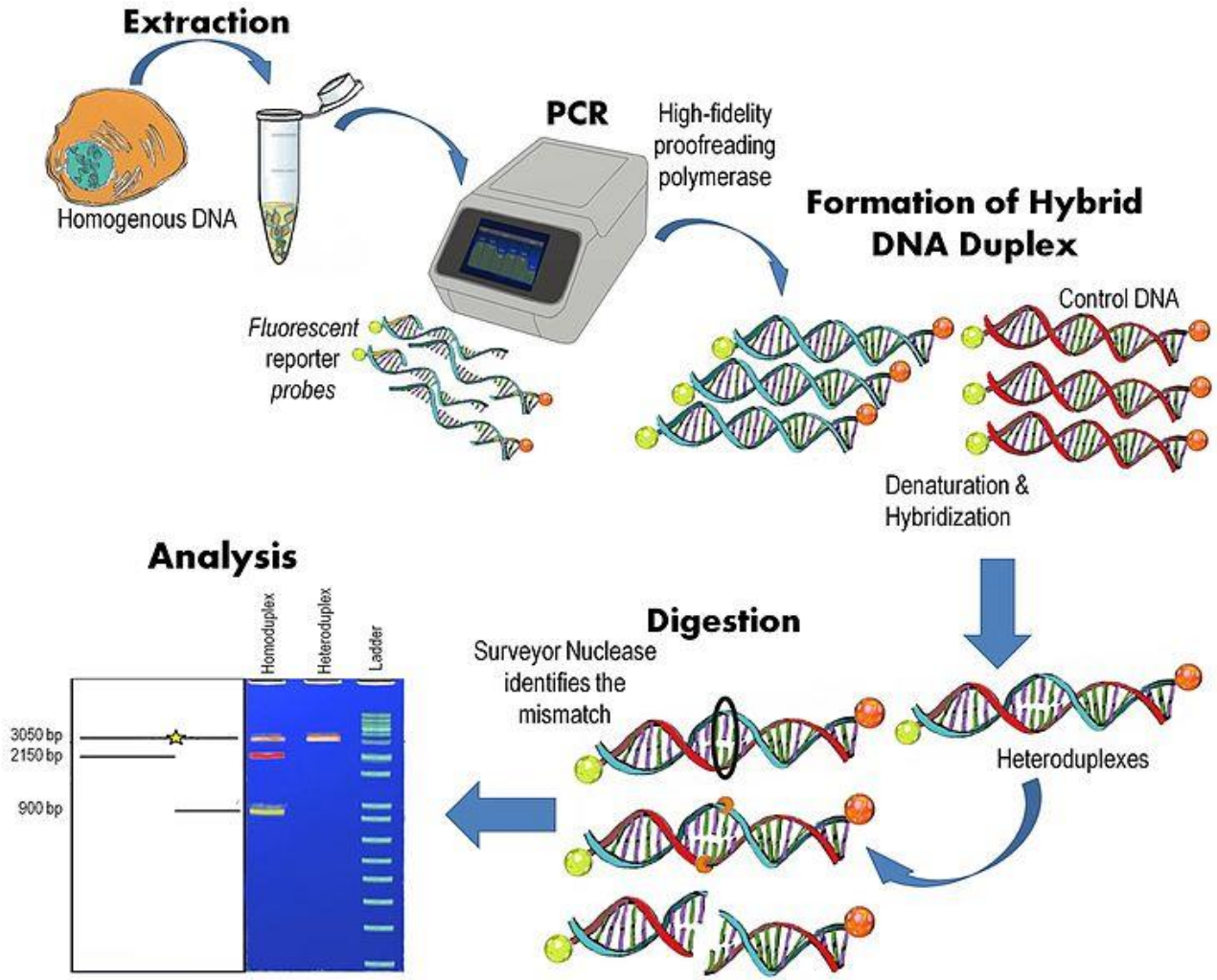


Tony Hirst. “Imprting cells with QUERY Select in google spreadsheet” 2010 https://www.flickr.com/photos/psychemedia/4270188042/
Wikipedia.org “Base Pairs”. 2001 http://upload.wikimedia.org/wikipedia/commons/5/5c/Hoogsteen_and_Watson%E2%80%93Crick_base_pairing.png
Carolina Biological Supply. “Lily pollen meiosis, anaphase II” 2011 https://www.flickr.com/photos/carolinabio/6240931651/
ZEISS Microscopy. “Drosophilia adult brain with proboscis” 2013 https://www.flickr.com/photos/zeissmicro/11404028293/

But what about…

Proteinbiochemist. “lab bench” 2009 https://www.flickr.com/photos/proteinbiochemist/3167660996/

and…



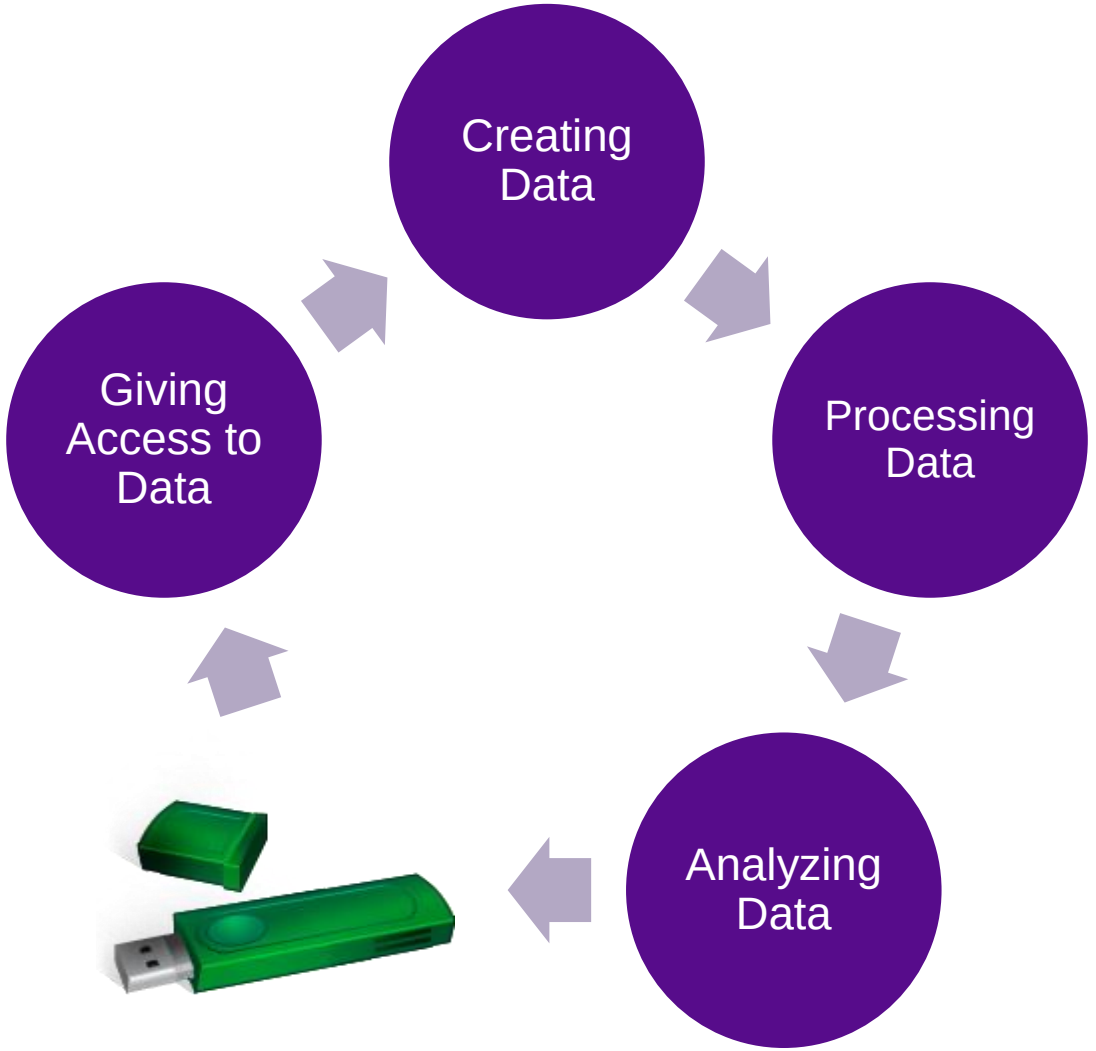
NaomiShomer. Surveyor Assay Workflow. Wikimedia 2016. https://commons.wikimedia.org/wiki/File:Surveyor_Assay_Workflow.jpg

and also…

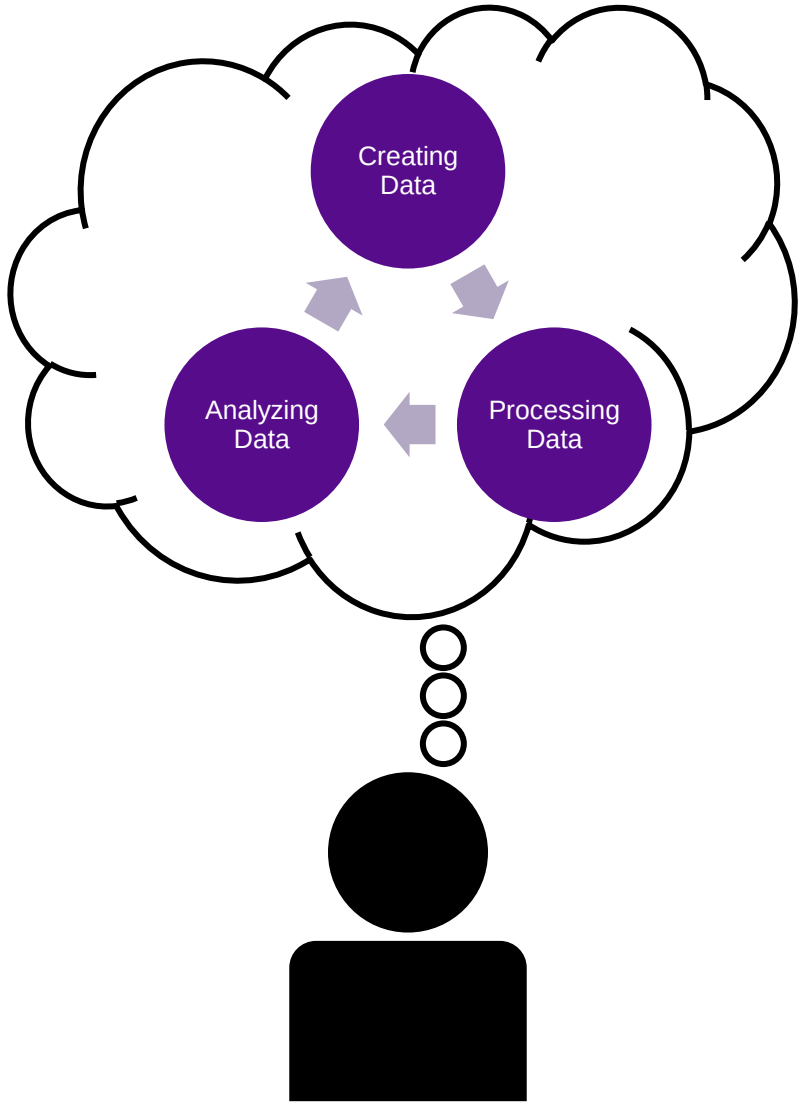
```
252 long fdosteps(long target, int mdelay_2) {              // The Movement Engine
253   long steps = 0;
254   char key = keypad.getKey();
255
256
257 steps =  fsoft( steps);
258   while (target >= steps) {
259     digitalWrite(pinclockplus, 1);
260     //   delayMicroseconds(mdelay_1);
261     digitalWrite(pinclockplus, 0);
262     delayMicroseconds(mdelay_2);
263     steps ++;
264   steps = fsoftstop(target, steps);
265     //   if (    digitalRead(STOP) == 1) {
266     //     fstop();
267     //     target = 0;
268     //     fscreen(0, "HIT STOP");
269     //     delay(1000);
270     //   }
271   }
272   return steps;
273   // steps = 0;
274 }
275 long fset_target( ) {            // Set the amount to be moved in mm
276   long temptarget = 0;
277   char key = '0';
278
279   String varskey = "";
280
281   fscreen(0, "Distance in mm");
```


Tudor Barker. “C++ code snippet” https://www.flickr.com/photos/tudedude/21904978214

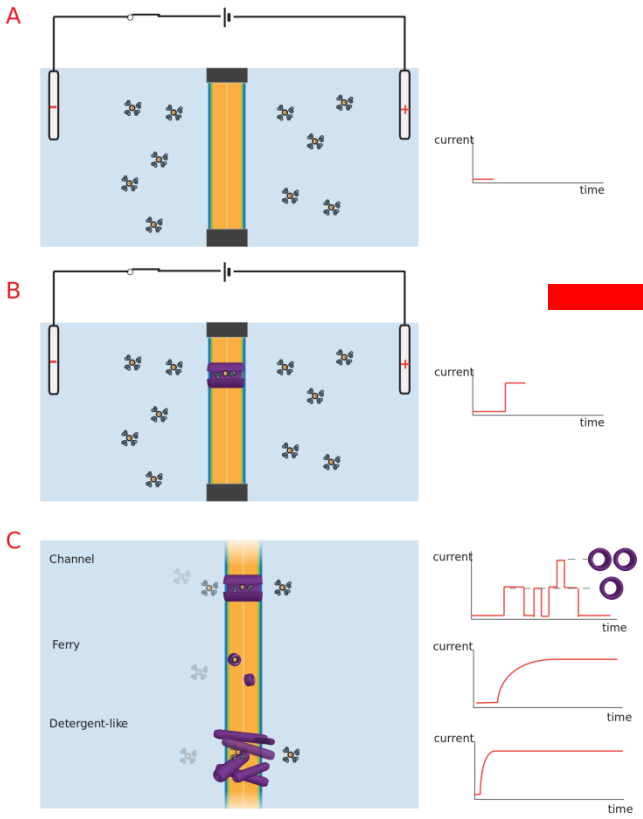
Data management lifecycle

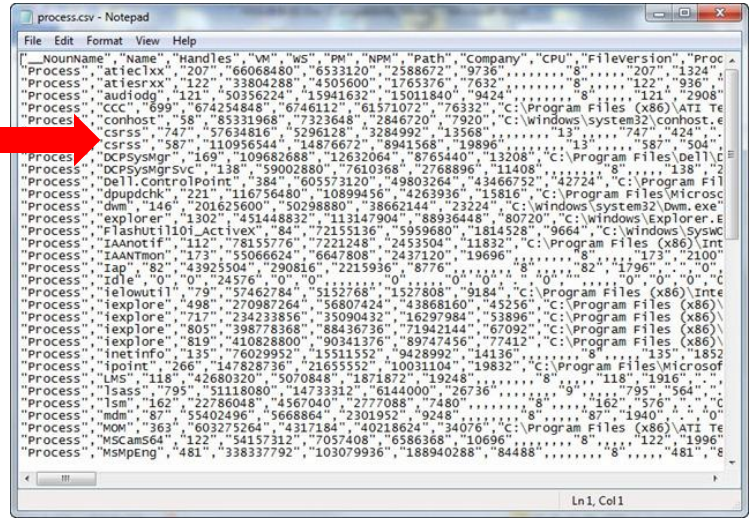


Think about your current workflow…



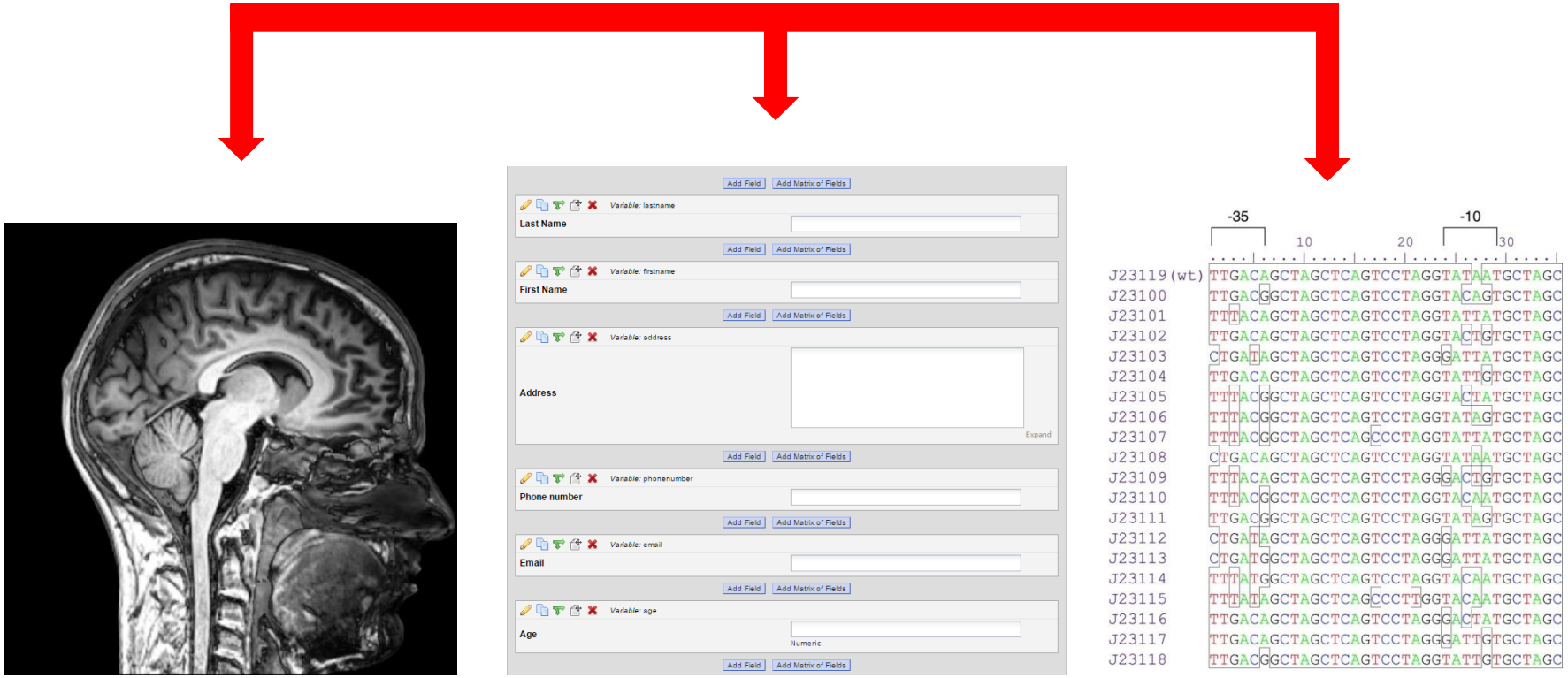
Can you easily locate raw data?





http://blogs.technet.com/cfs-file.ashx/__key/CommunityServer-Blogs-Components-WeblogFiles/00-00-00-76-18-metablogapi/2055.HSG09091001_5F00_71EB0EE6.jpg Figure from a paper

Can you connect different types of related data you collected?



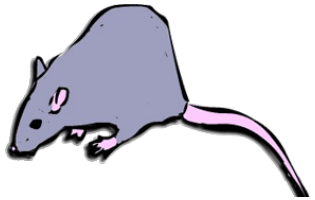
https://www.microscopyu.com/articles/superresolution/images/stormintrofigure4.jpg
http://parts.igem.org/wiki/images/0/05/BerkiGEM2006-Promoters.jpg
https://www.flickr.com/photos/mgdtgd/3507973704/in/photostream/

Are your naming conventions consistent with others on your team?

Rat 001

?

Rat A



Don’t end up here!

http://retractionwatch.com/2013/10/30/nejm-paper-on-sleep-apnea-retracted-when-original-data-cant-be-found/

Don’t end up here!

Multiple errors in table

http://retractionwatch.com/2013/10/30/nejm-paper-on-sleep-apnea-retracted-when-original-data-cant-be-found/

Don’t end up here!

Retraction Watch

NEJM paper on sleep apnea retracted when original data can’t be found

with 4 comments

The authors of a paper in the *New England Journal of Medicine* are retracting it, after being unable to find data supporting a table that required corrections.

The NEW ENGLAND JOURNAL of MEDICINE

Lost data!

Multiple errors in table

Did not alter conclusions in article

http://retractionwatch.com/2013/10/30/nejm-paper-on-sleep-apnea-retracted-when-original-data-cant-be-found/

Don’t end up here!

Multiple errors in table

Did not alter conclusions in article

BUT, could not locate primary data

http://retractionwatch.com/2013/10/30/nejm-paper-on-sleep-apnea-retracted-when-original-data-cant-be-found/

Don’t end up here!

Retraction Watch

NEJM paper on sleep apnea ret
found

with 4 comments

The authors of a paper i
Medicine are retracti
supporting a tab

ENGLAND
JRNAL of MEDICINE

M

RETRACTION

clusions in article

, could not locate primary data

http://retractionwatch.com/2013/10/30/nejm-paper-on-sleep-apnea-retracted-when-original-data-cant-be-found/

Or here!

Retraction Watch

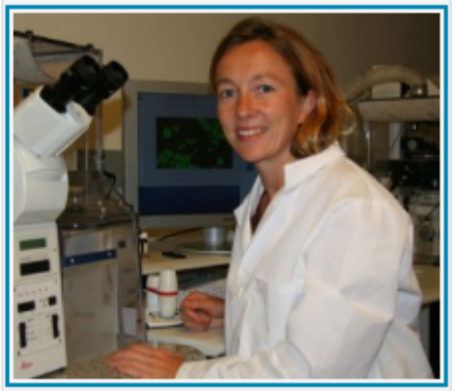
Tracking retractions as

Leading diabetes researcher acted negligently, probe concludes

with 2 comments

Several duplications in the work of a prominent diabetes researcher were the result of negligence, but there is not enough evidence to support charges of misconduct, according to an investigation at her university in Germany.

Recently, we've reported on several notices for papers co-authored by Kathrin Maedler, a researcher at the University of Bremen. So far, Maedler has one retraction, multiple corrections, and two expressions of concern to her name, after several of her papers were questioned on PubPeer. Previously, the University of Zurich in Switzerland — where Maedler completed her PhD in 2002 — determined there was a lack of evidence to support allegations of misconduct in papers that were part of her doctoral thesis.



Kathrin Maedler

"had published duplicate pictures in several cases and had repeatedly failed to exert due diligence in organising her area of study over a long period of time."

http://retractionwatch.com/2016/11/02/leading-diabetes-researcher-acted-negligently-probe-concludes/

Data management best practices

Files

File Names

sam_1262011.tif

File Names

sam_1262011.tif

12 June, 2011?
December 6, 2011?
January 26, 2011?

File Names

sam_1262011.tif

Scanning acoustic microscope?	12 June, 2011?
Systolic anterior motion?	December 6, 2011?
Sam the postdoc?	January 26, 2011?

File Names

sam_1262011.tif

Scanning acoustic microscope?	12 June, 2011?
Systolic anterior motion?	December 6, 2011?
Sam the postdoc?	January 26, 2011?

Unambiguous dates, the **ISO standard**:

- YYYYMMDD *or* YYYY-MM-DD
  - *e.g. 20120612 = June 6, 2012*
- YYYYMMDDTHH:MM:SS
  - *e.g. 20120612T14:03:12 = June 6, 2012 2:03:12 pm*

1 rat heart

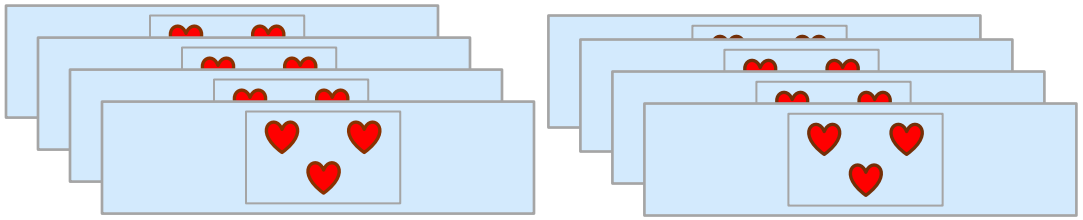
1 rat heart

100s of slices

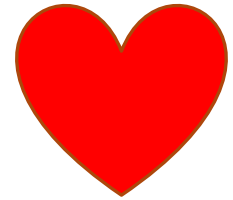
1 rat heart

100s of slices

100s of slides

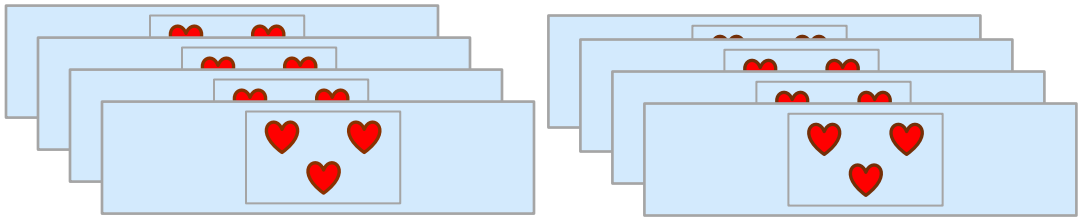


1 rat heart

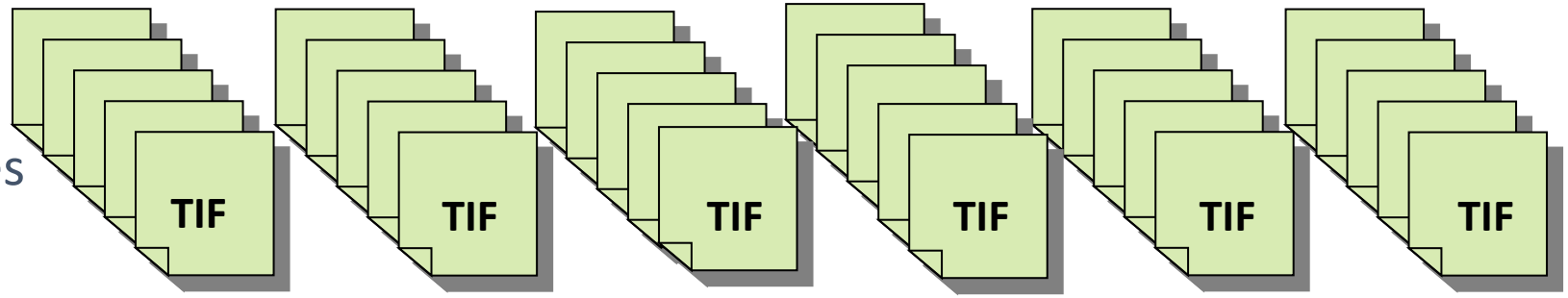


100s of slices

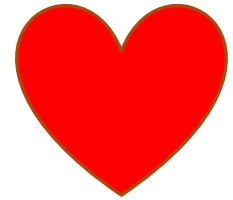
100s of slides



1000s of image files

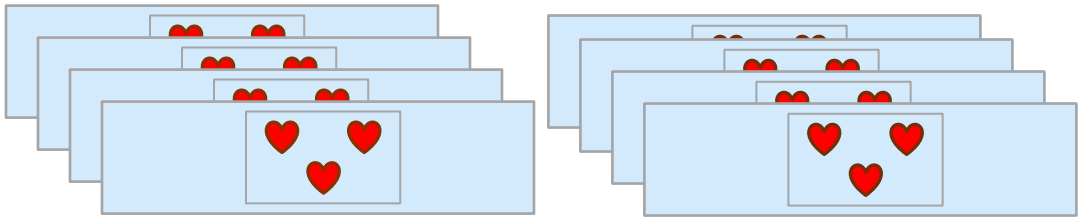


1 rat heart

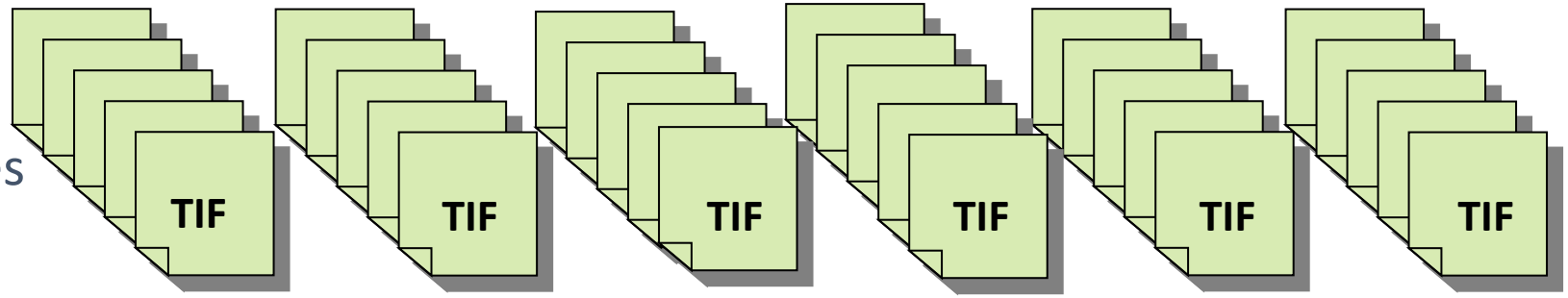


100s of slices

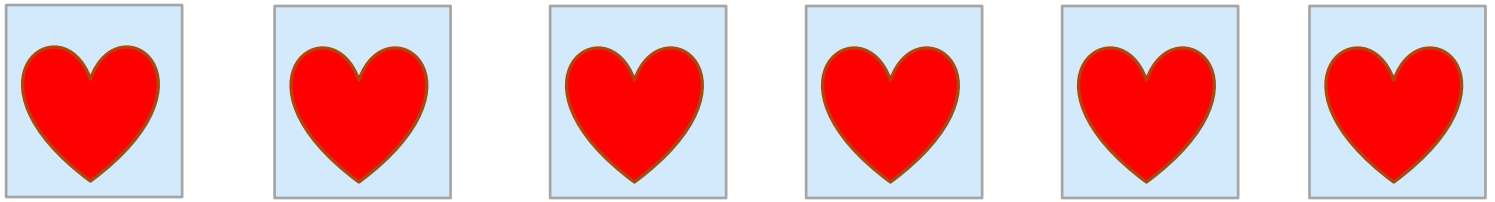
100s of slides



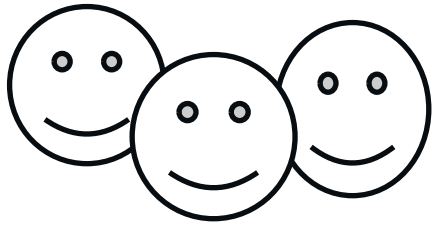
1000s of image files



100s of *huge* images



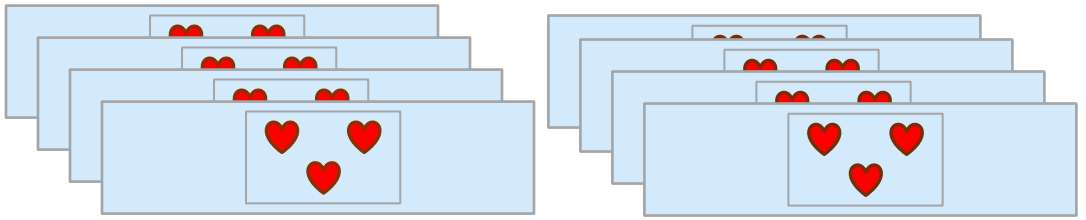
1 rat heart



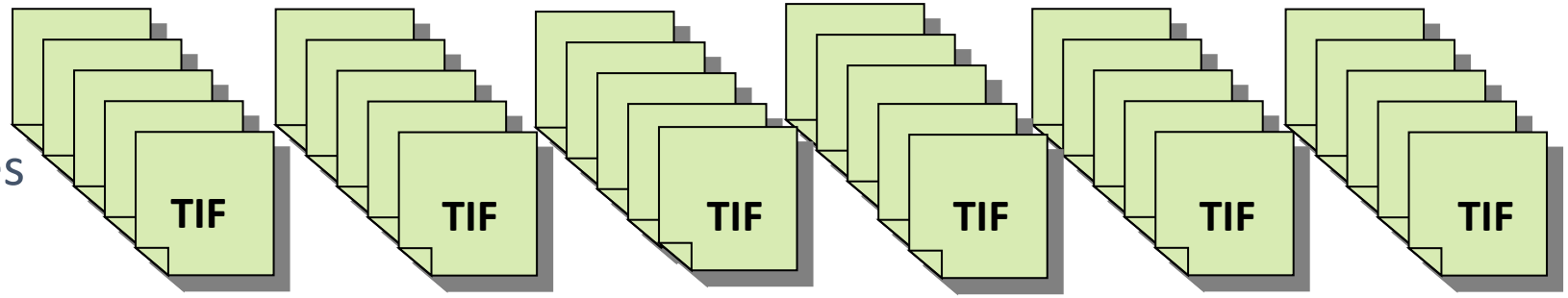
3 postdocs

100s of slices

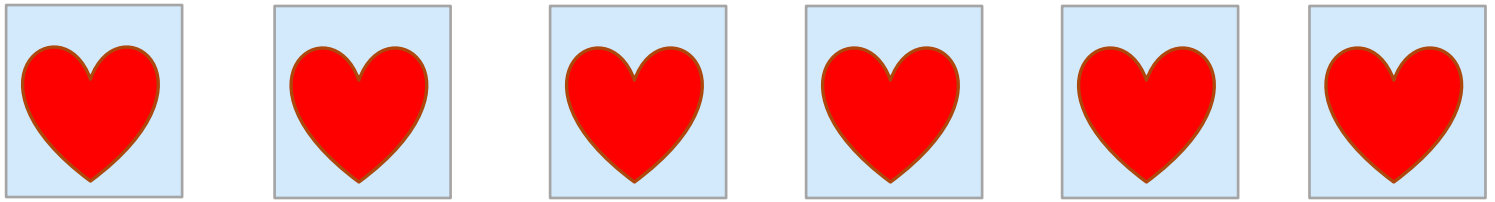
100s of slides



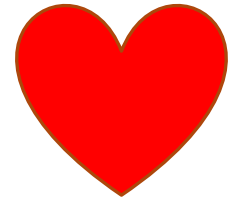
1000s of image files



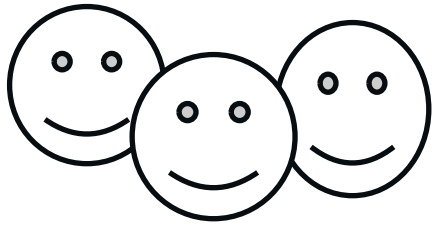
100s of *huge* images



1 rat heart



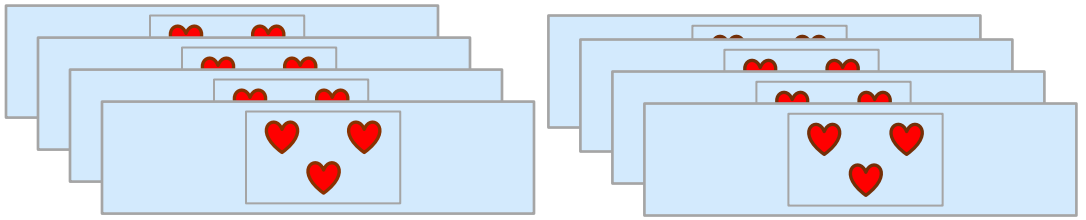
100s of slices



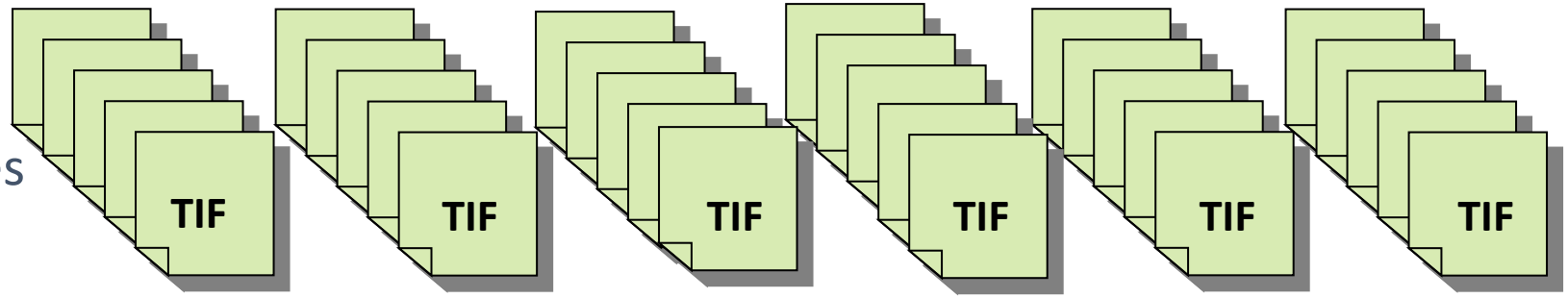
3 postdocs

5-7 experiments a week…

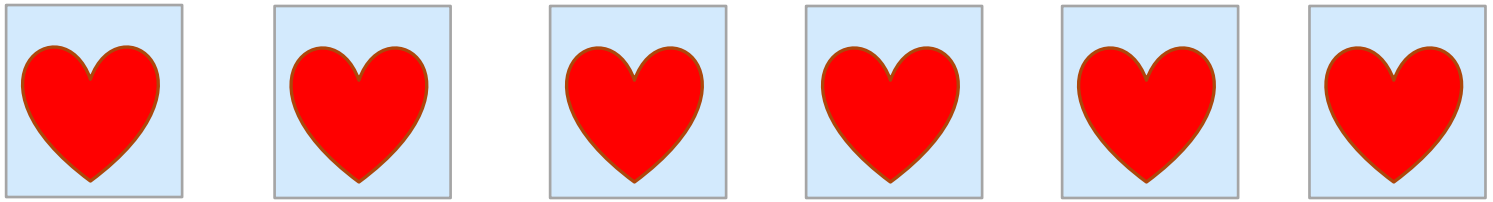
100s of slides



1000s of image files



100s of *huge* images



File names should…

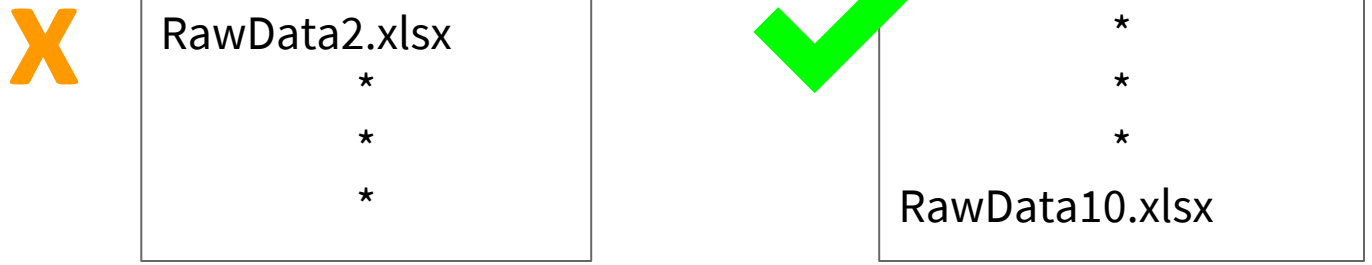
1. Embody their content, including major parameters
   AtherRat_ex012_ather_lipitor_128.tif

File names should…

1. Embody their content, including major parameters
   AtherRat_ex012_ather_lipitor_128.tif
2. Have non-cryptic/intuitive names where possible
   AtherRat_SOP_DataValidation_v01.docx

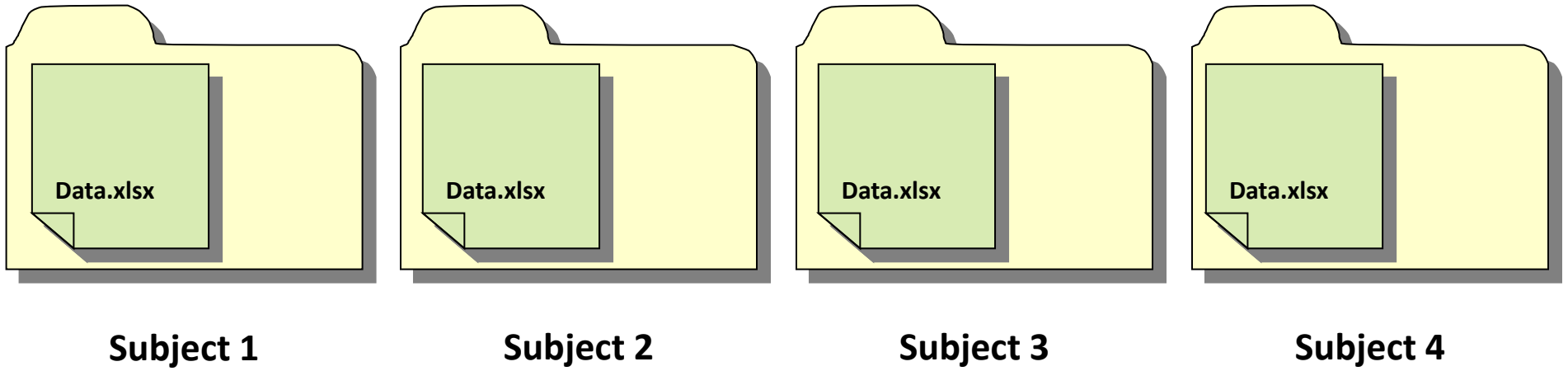
File names should…

3. Be extensible. “ex001” not “ex1”



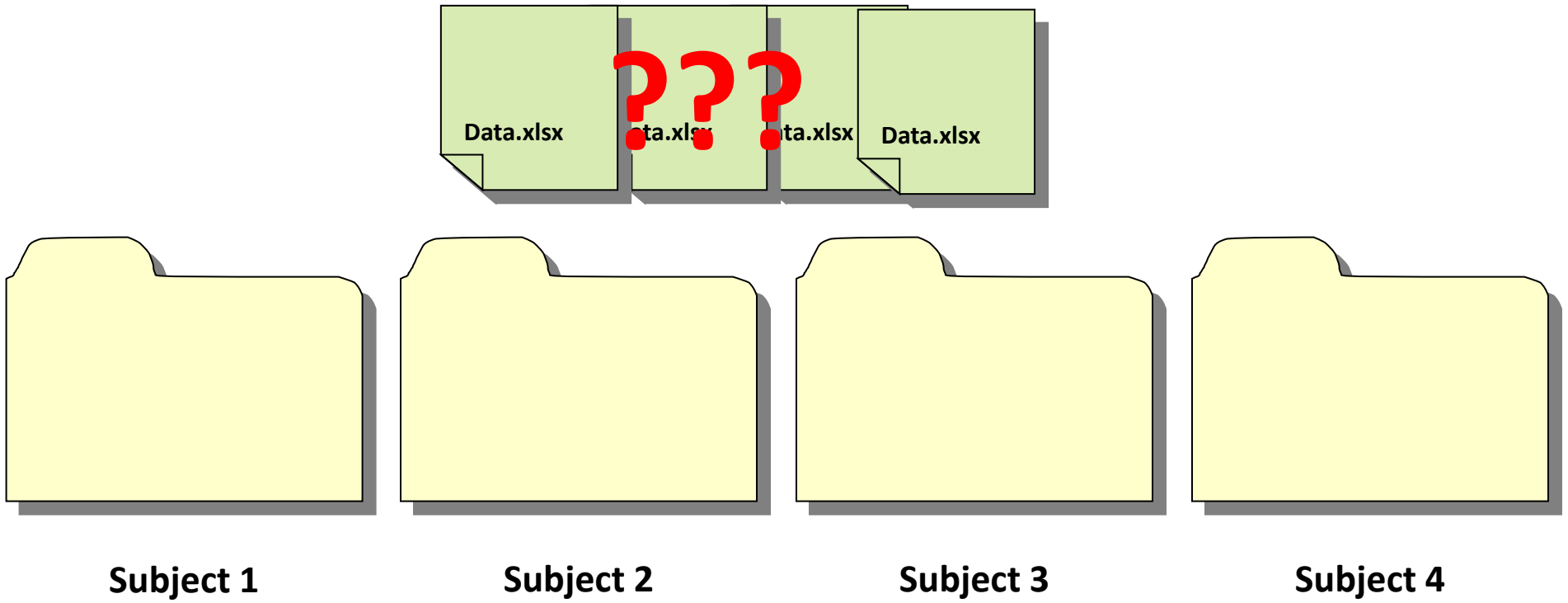
File names should…

4. Be unique, where possible and practical. Avoid 20 files called "data.xlsx" in different folders



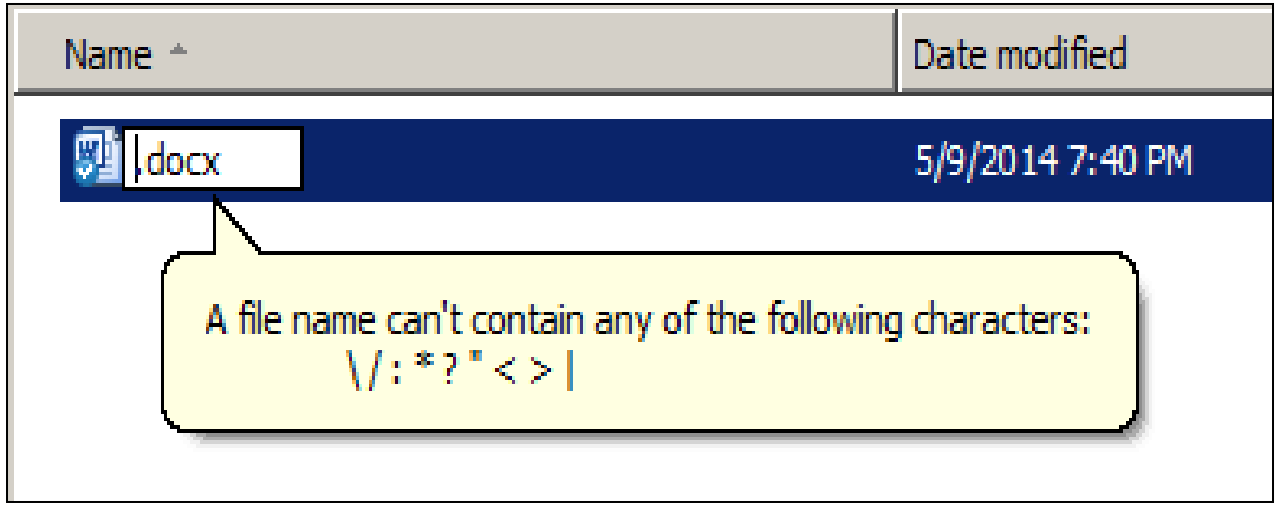
File names should…

4. Be unique, where possible and practical. Avoid 20 files called "data.xlsx" in different folders



File names should…

5. Do not use special characters – restrict file names to numbers, letters, and underscores



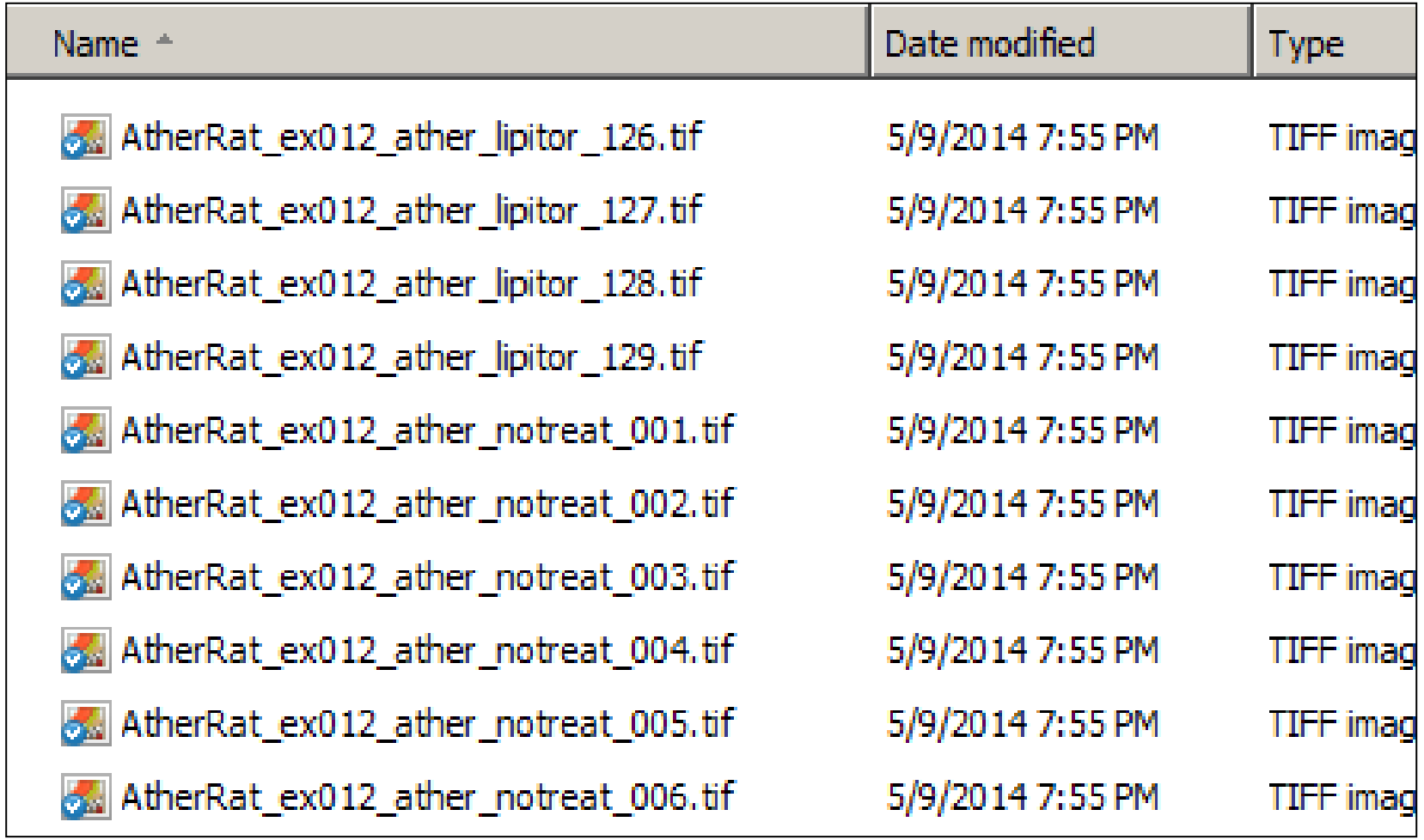
File names should…

6. Use consistent, documentable rules for naming files

AtherRat_012_056_mb_0423_raw.csv

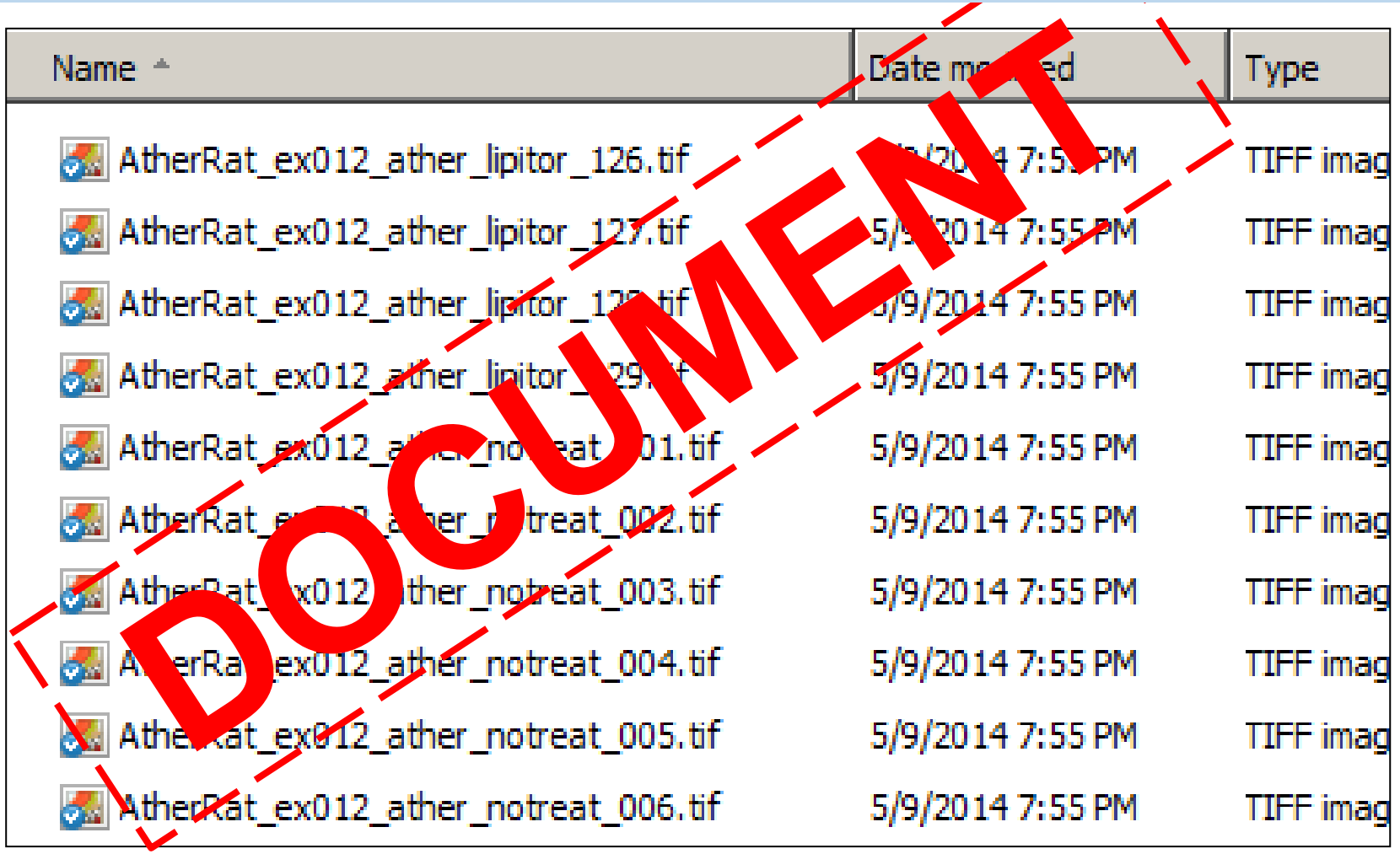
AtherRat = experiment name
012 = experiment number
056 = sample number
mb = stain used, methylene blue
0423 = 2-digit coordinates of image (4 across, 23 down)
Raw = data stage

File names should…



Name	Date modified	Type
AtherRat_ex012_ather_lipitor_126.tif	5/9/2014 7:55 PM	TIFF imag
AtherRat_ex012_ather_lipitor_127.tif	5/9/2014 7:55 PM	TIFF imag
AtherRat_ex012_ather_lipitor_128.tif	5/9/2014 7:55 PM	TIFF imag
AtherRat_ex012_ather_lipitor_129.tif	5/9/2014 7:55 PM	TIFF imag
AtherRat_ex012_ather_notreat_001.tif	5/9/2014 7:55 PM	TIFF imag
AtherRat_ex012_ather_notreat_002.tif	5/9/2014 7:55 PM	TIFF imag
AtherRat_ex012_ather_notreat_003.tif	5/9/2014 7:55 PM	TIFF imag
AtherRat_ex012_ather_notreat_004.tif	5/9/2014 7:55 PM	TIFF imag
AtherRat_ex012_ather_notreat_005.tif	5/9/2014 7:55 PM	TIFF imag
AtherRat_ex012_ather_notreat_006.tif	5/9/2014 7:55 PM	TIFF imag

File names should…



Data Collection

Your variables

	A	B	C	D	E
1	SID	wgt	smoking	name	sam
2	1	49	Y	Smith	13
3	2	252	2 packs	Sam Jones	37
4	3	28	N	Read, Kevin	A21
5	4	157	Never	Emma Banks	January
6					

Your variables

	A	B	C	D	E
1	SID	wgt	smoking	name	sam
2	1	49	Y	Smith	13
3	2	252	2 packs	Sam Jones	37
4	3	28	N	Read, Kevin	A21
5	4	157	Never	Emma Banks	January
6					

Your variables

	A	B	C	D	E
1	SID	wgt	smoking	name	sam
2	1	49	Y	Smith	13
3	2	252	2 packs	Sam Jones	37
4	3	28	N	Read, Kevin	A21
5	4	157	Never	Emma Banks	January
6					

Your variables

	A	B	C	D	E
1	SID	wgt	smoking	name	sam
2	1	49	Y	Smith	13
3	2	252	2 packs	Sam Jones	37
4	3	28	N	Read, Kevin	A21
5	4	157	Never	Emma Banks	January
6					

Your variables

- Intuitive / meaningful variable names e.g. study_id
- What do variable names mean?
- What does each variable contain?
- Are there a limited set of possible values?

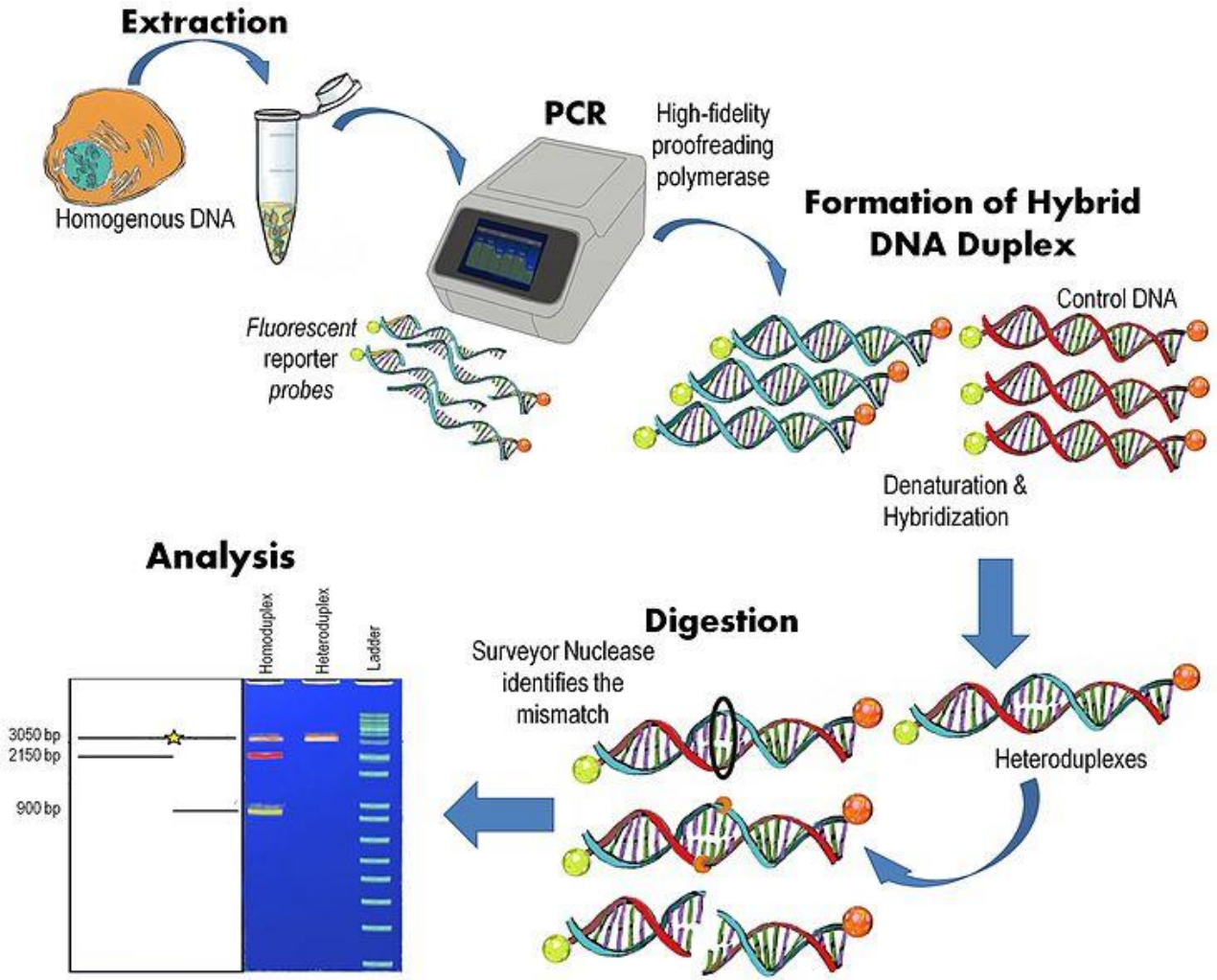
Name	Field Type	Description	Possible values	Units
study_id	text	Unique ID of study	8-digit number	
date_enrolled	date	Initial subject enrollment date	Date in format YYYY-MM-DD; All dates later than 2011-09-01	
weight	integer	Weight of subject		lbs

Your variables

	A	B	C	D	E	F	G
1	**Variable / Field Name**	**Form Name**	**Field Type**	**Field Label**	**Choices, Calculations, OR Slider Labels**	**Text Validation Min**	**Text Validation Max**
2	record_id	demographics	text	Record ID			
3	mrn	demographics	text	MRN			
4	last_name	demographics	text	Last name			
5	first_name	demographics	text	First name			
6	age	demographics	text	Age	1, <55 | 2, between 55 and 75 | 3, >75	21	105
7	gender	demographics	radio	Gender	1, Male | 2, Female		
8	race	demographics	radio	Race/Ethnicity	1, White | 2, Black | 3, Asian | 4, Hispanic/Latino | 5, Other		
9	describe_other	demographics	text	Describe			
10	education	demographics	radio	Highest Level of Education Completed	1, < highschool diploma | 2, highschool diploma | 3, associate degree | 4, bachelors degree | 5, masters degree | 6, graduate school or advanced degree		
11	yes	demographics	radio	Working	1, Yes | 2, No		
12	occupation	demographics	text	Occupation			
13	income	demographics	radio	Household Income	1, Household Income <30,000/year | 2, Household income between 30-50,000/year | 3, Household income 50-75,000/year | 4, Household income 75-100,000/year | 5, Household income 100-150,000/year | 6, Household income 150-250,000/year | 7, Household income >250,000/year		
14	htn	medical_history	radio	Hypertension	1, Yes | 2, No | 3, Unkown		
15	hld	medical_history	radio	Hyperlipidemia	1, Yes | 2, No | 3, Unkown		
16	dm	medical_history	radio	Diabetes	1, Yes | 2, No | 3, Unkown		
17	current_smoker	medical_history	radio	Current Smoker	1, Yes | 2, No | 3, Unkown		
18	former_smoker	medical_history	radio	Former Smoker	1, Yes | 2, No | 3, Unkown		
19	smoking_start_date	medical_history	text	Smoking start date			
20	smoking_quit_date	medical_history	text	Smoking Quit Date			
21	depression	medical_history	radio	Depression	1, Yes | 2, No | 3, Unknown		
22	anxiety	medical_history	radio	Anxiety	1, Yes | 2, No | 3, Unknown		
23	stress_cardiomyopathy	medical_history	radio	Stress Cardiomyopathy (TakoTsubo)	1, Yes | 2, No | 3, Unknown		
24	prior_mi	medical_history	radio	Prior MI	1, Yes | 2, No | 3, Unknown		
25	prior_stroke	medical_history	radio	Prior Stroke	1, Yes | 2, No | 3, Unknown		
26	prior_tia	medical_history	radio	Prior TIA	1, Yes | 2, No | 3, Unknown		
27	prior_hf	medical_history	radio	Prior HF	1, Yes | 2, No | 3, Unknown		
28	etoh_use	medical_history	radio	Alcohol Use	1, Yes | 2, No | 3, Unknown		
29	etoh_use_quantity	medical_history	radio	How much alcohol do you drink in a ty	1, 1-3 drinks | 2, 4-7 drinks | 3, 7-15 drinks | 4, greater than 15 drinks		
30	mj_use	medical_history	radio	Marijuana Use	1, Yes | 2, No | 3, Unknown		
31	age_at_menopause	medical_history	text	Age at Menopause			

Workflows

Workflows



NaomiShomer. Surveyor Assay Workflow. Wikimedia 2016. https://commons.wikimedia.org/wiki/File:Surveyor_Assay_Workflow.jpg

Workflows: Clinical Protocols

Current Primary Outcome Measures ICMJE **(submitted: November 24, 2008)**	HbA1c [Time Frame: 1 year] [Designated as safety issue: No]
Original Primary Outcome Measures ICMJE	*Same as current*
Change History	Complete list of historical versions of study NCT00797888 on ClinicalTrials.gov Archive Site
Current Secondary Outcome Measures ICMJE **(submitted: November 24, 2008)**	diabetes self-care activities [Time Frame: 1 year] [Designated as safety issue: No]
Study Design ICMJE	Allocation: Randomized Endpoint Classification: Efficacy Study Intervention Model: Parallel Assignment Masking: Open Label Primary Purpose: Supportive Care
Condition ICMJE	Type 2 Diabetes
Intervention ICMJE	Behavioral: telephonic Between 4-8 phone calls each year for health behavior counseling to improve HbA1c
Study Arm (s)	• Experimental: Telephonic Tailored telephonic intervention to improve HbA1c for participants in the diabetes registry Intervention: Behavioral: telephonic • Active Comparator: Standard registry People with diabetes who are in the A1c registry may receive letters from the DOHMH to promote im resources for healther foof and activites Intervention: Behavioral: telephonic

https://clinicaltrials.gov/ct2/show/record/NCT00797888?term=NCT00797888&rank=1

Workflows: In the weeds

- How, when and who will do the work?
- Will data be reviewed for quality?
- Who manages the entire process?

http://www.publicdomainpictures.net/pictures/20000/velka/field-of-

Workflows: In the weeds

- How, when and who will do the work?
- Will data be reviewed for quality?
- Who manages the entire process?

DOCUMENTED

http://www.publicdomainpictures.net/pictures/20000/velka/field-of-

Who is responsible for data management?

Who is responsible for data management?

Everyone!

(but everyone means no one)

Quality Control

Assign a person to be responsible for ensuring:

- ✓ Naming conventions adhered to
- ✓ Minimum documentation
- ✓ Version controls followed
- ✓ Data backed up

Storage

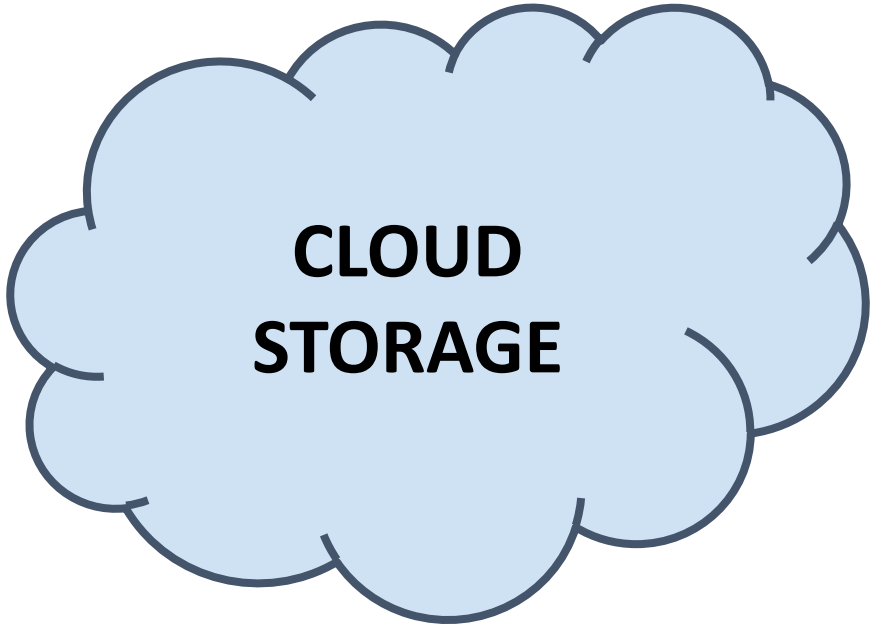
Storage @ ECU

- PirateDrive
- Dataverse—for final datasets
- For information about where to store sensitive data: https://datagovernance.ecu.edu/sensitive-data-storage-and-transmission/

Cloud Storage

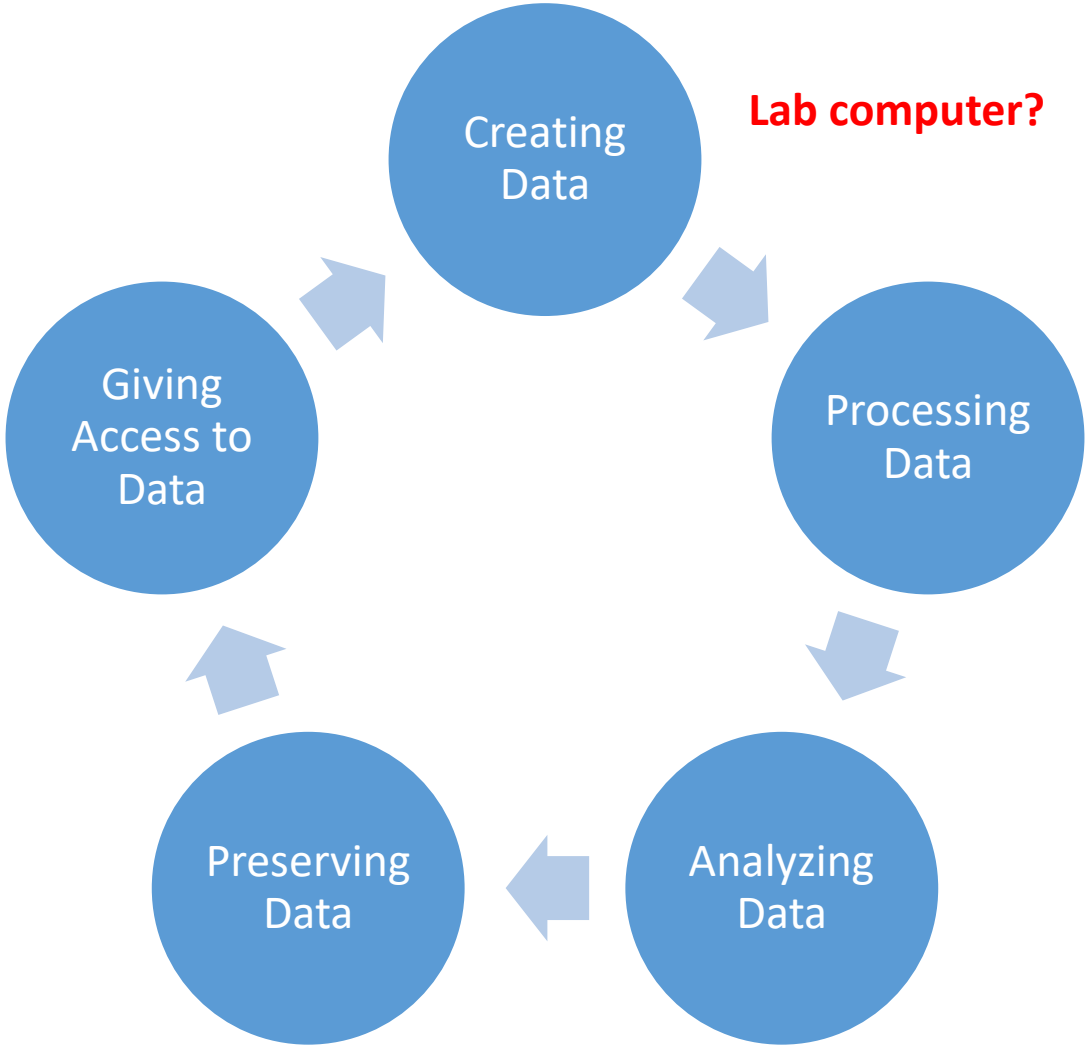
If you are going to use cloud storage:

- Check ownership policies
- Pick >1 provider

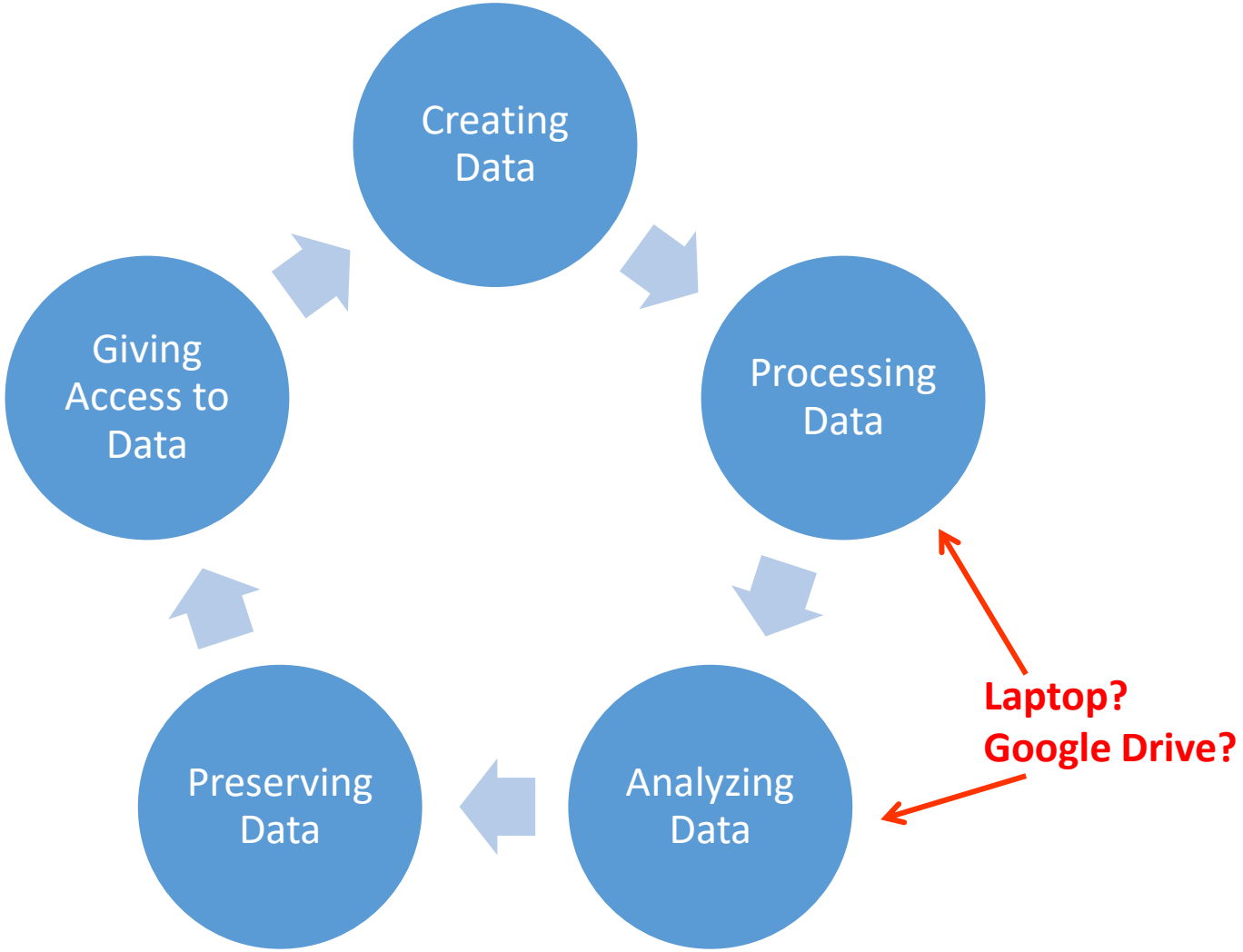


Where is data stored in different parts of the workflow?

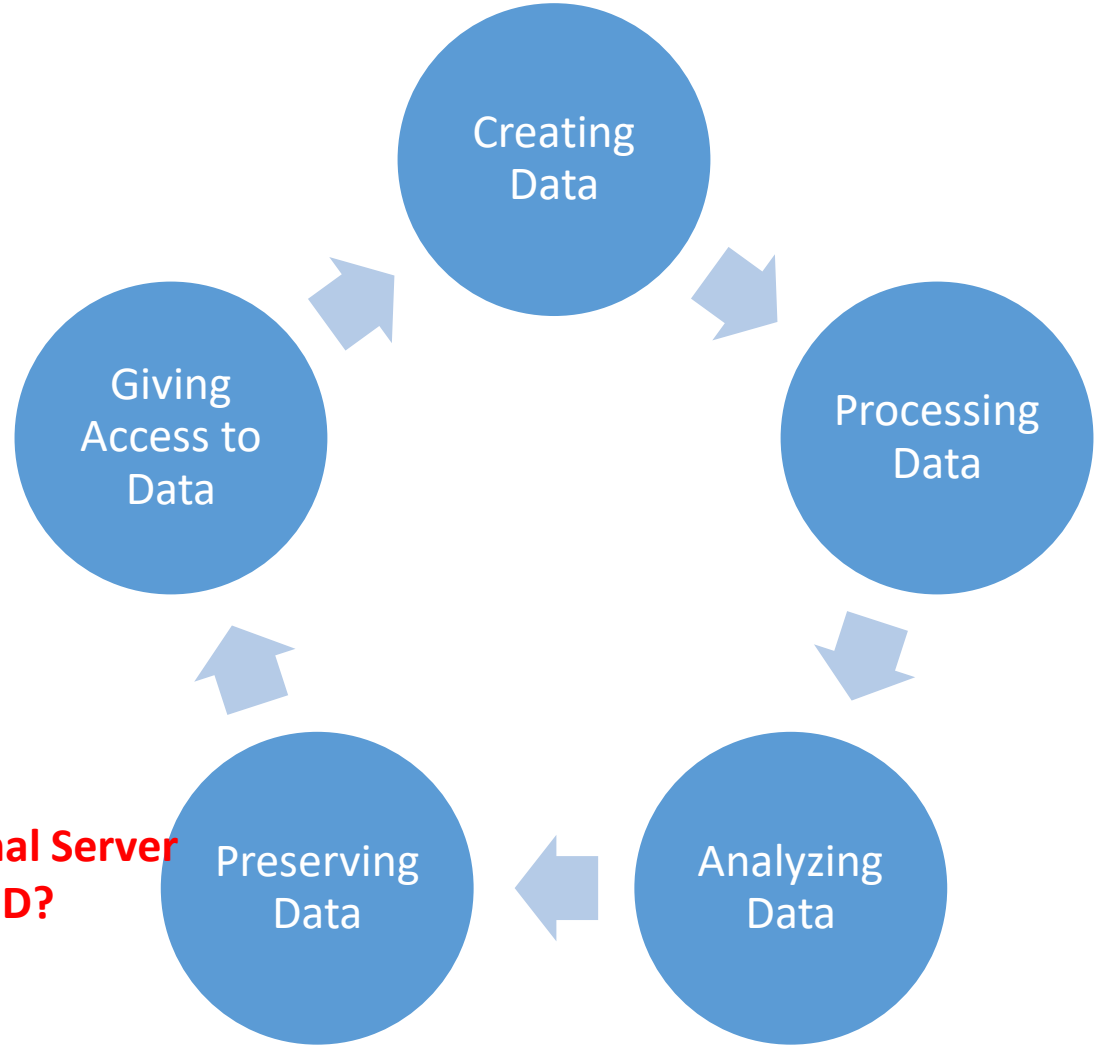
Data Lifecycle



Data Lifecycle



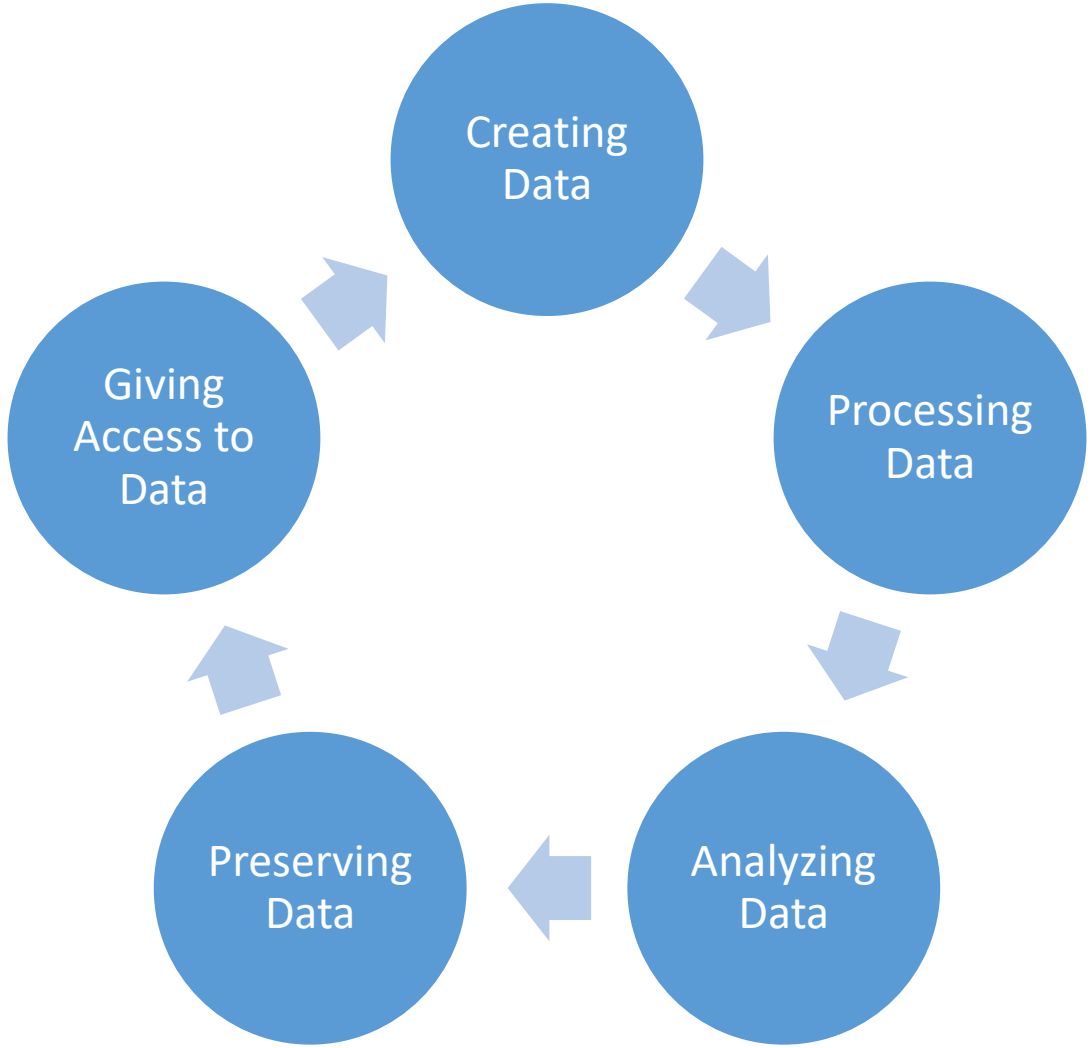
Data Lifecycle



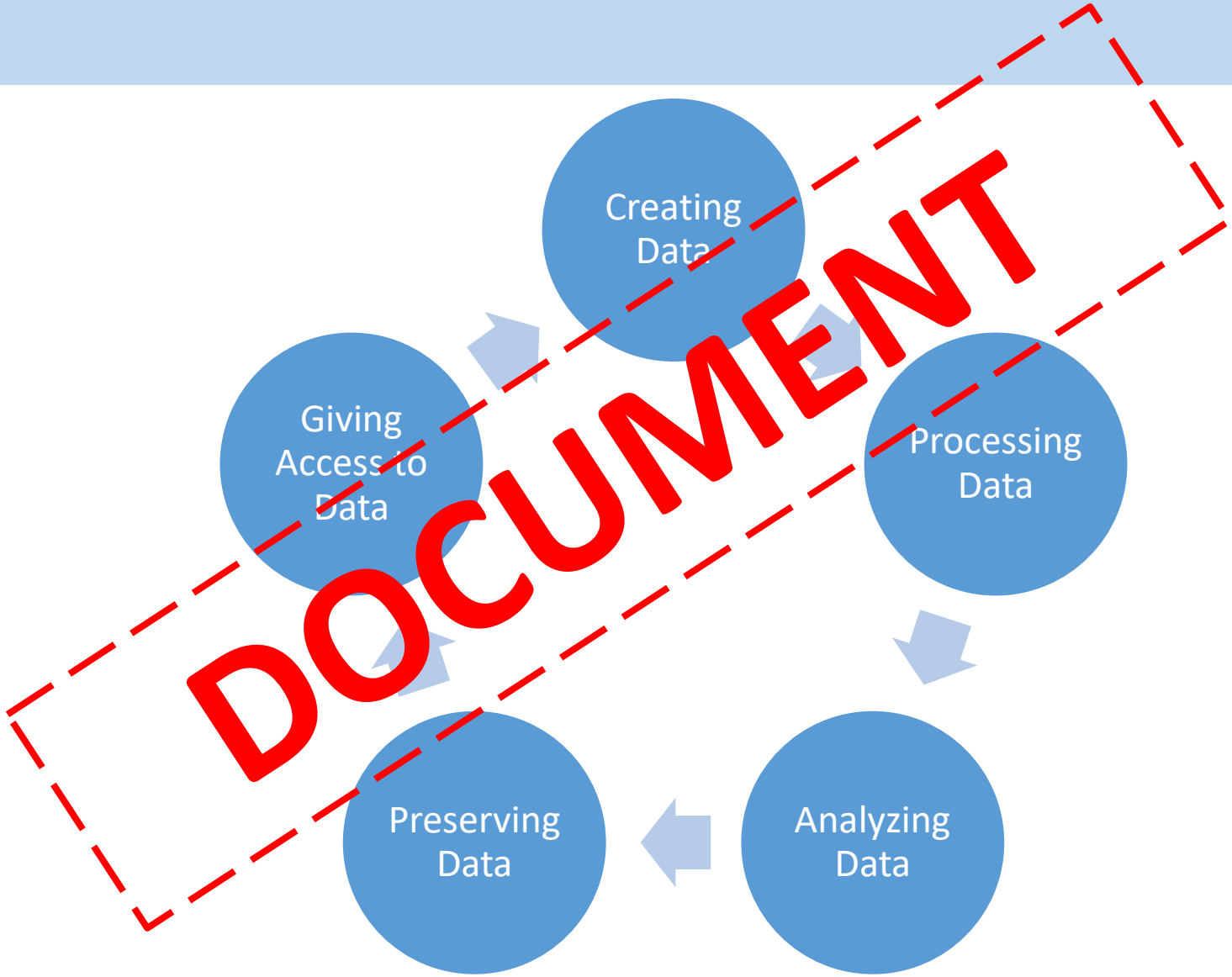
Institutional Server
External HD?

Data Lifecycle

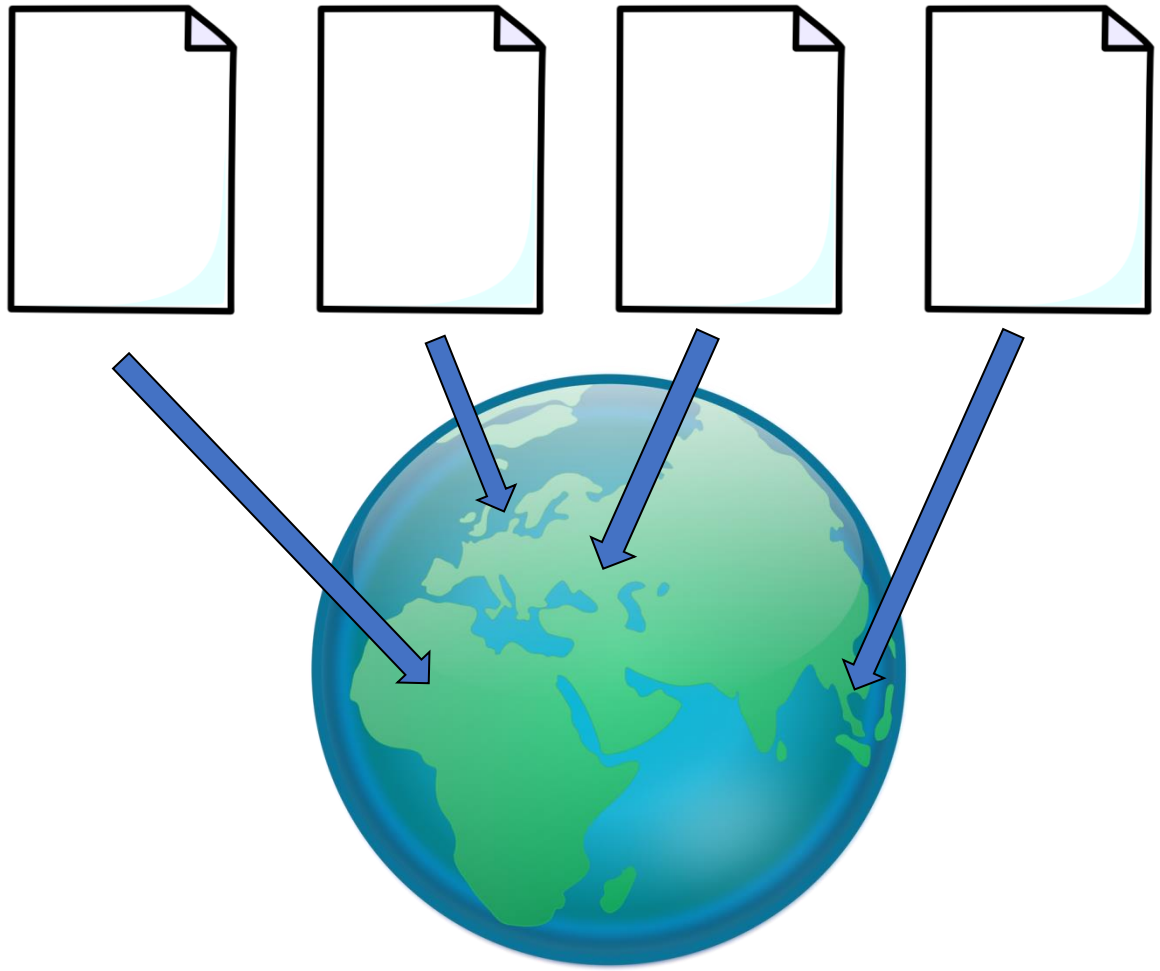
Repository



Data Lifecycle



Multiple locations

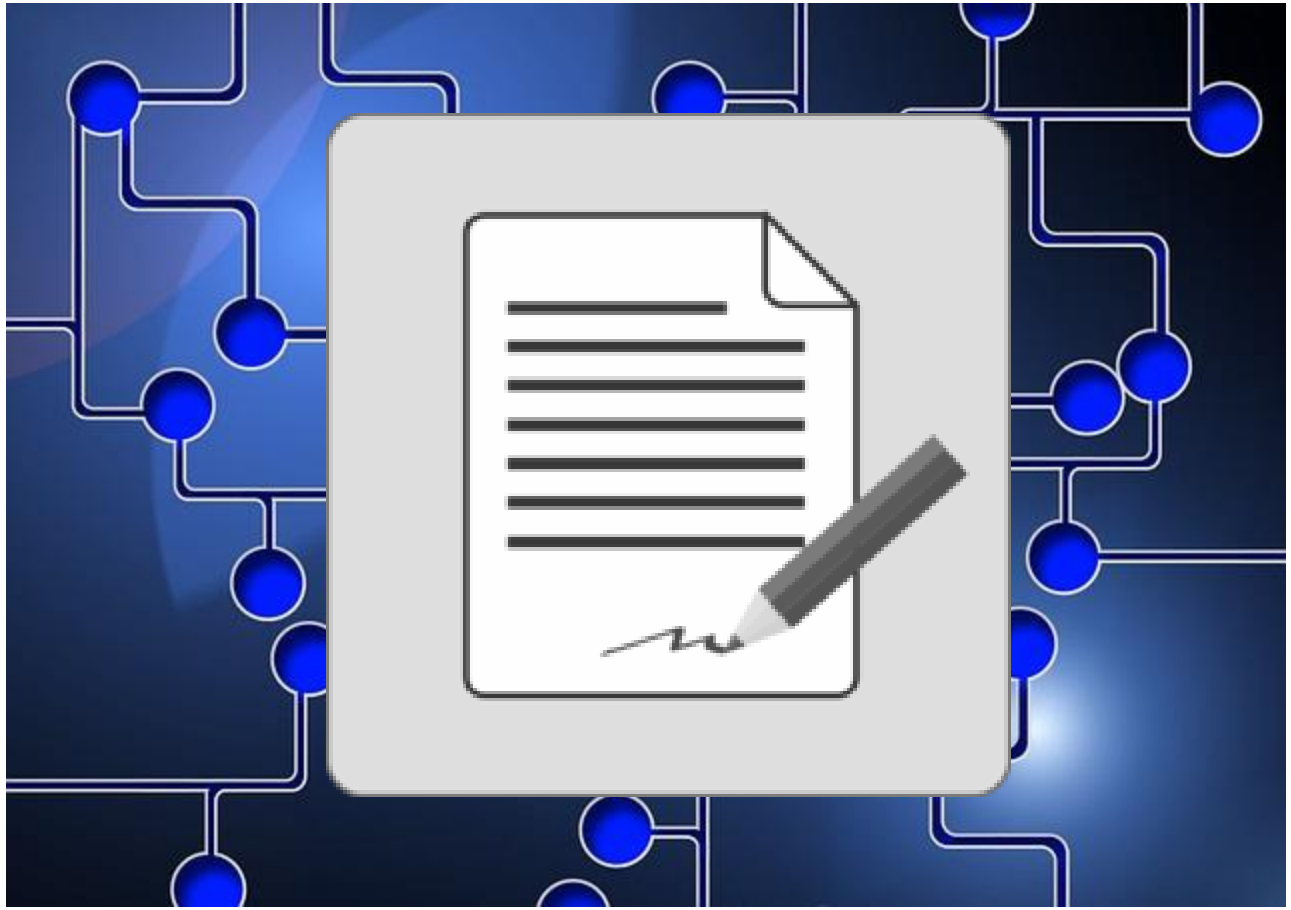


Security Considerations

Security Considerations

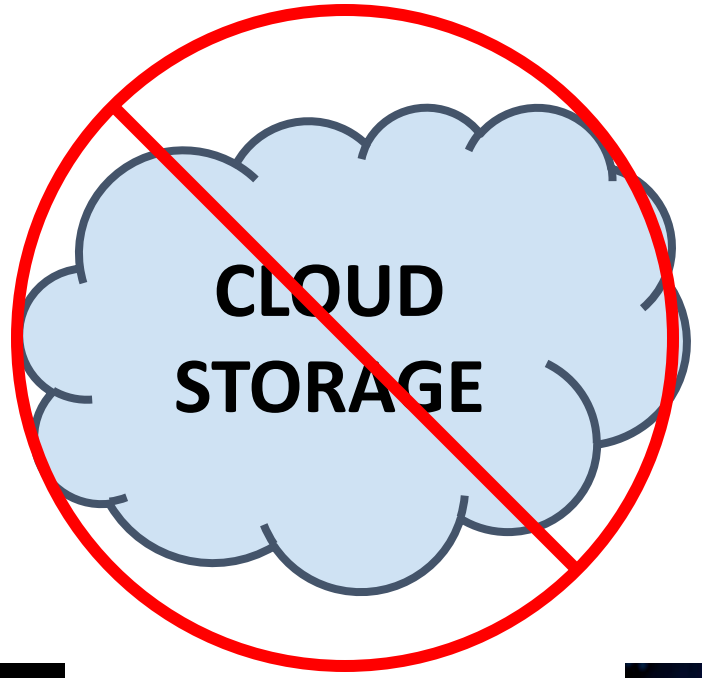
http://pixabay.com/p-538722/?no_redirect

Security Considerations



http://pixabay.com/p-538722/?no_redirect

Security Considerations: HIPAA



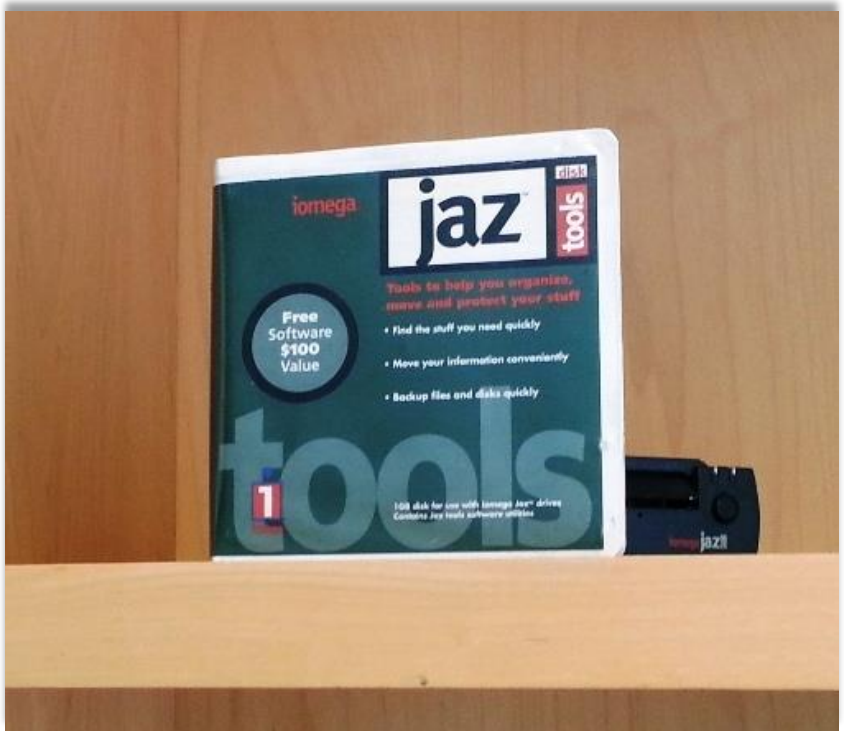
Security Considerations: HIPAA

Preservation

storage ≠ preservation

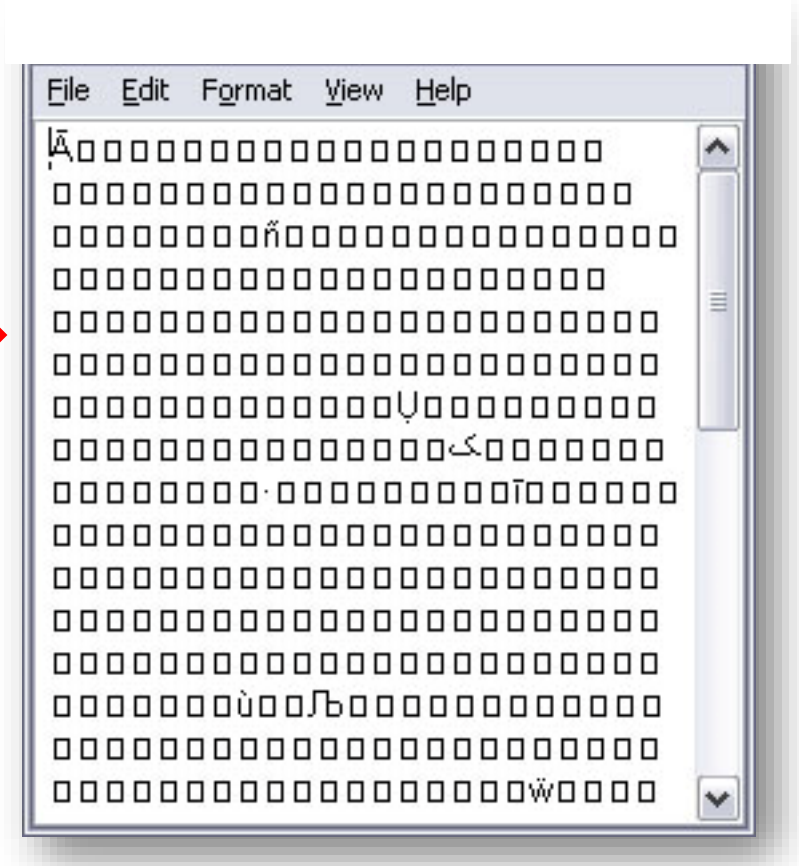
Preservation

Protects from: Hardware obsolescence



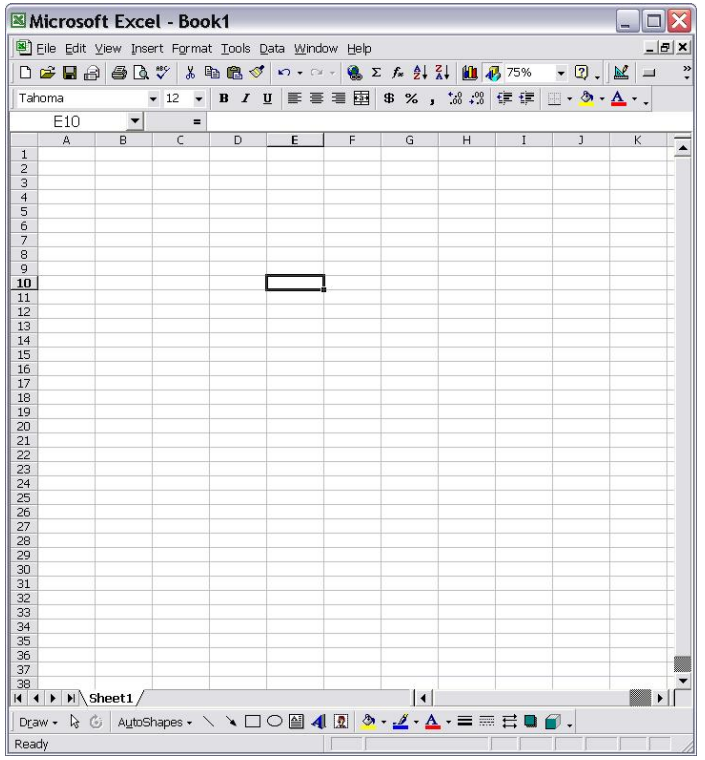
Preservation

Protects from: Software obsolescence

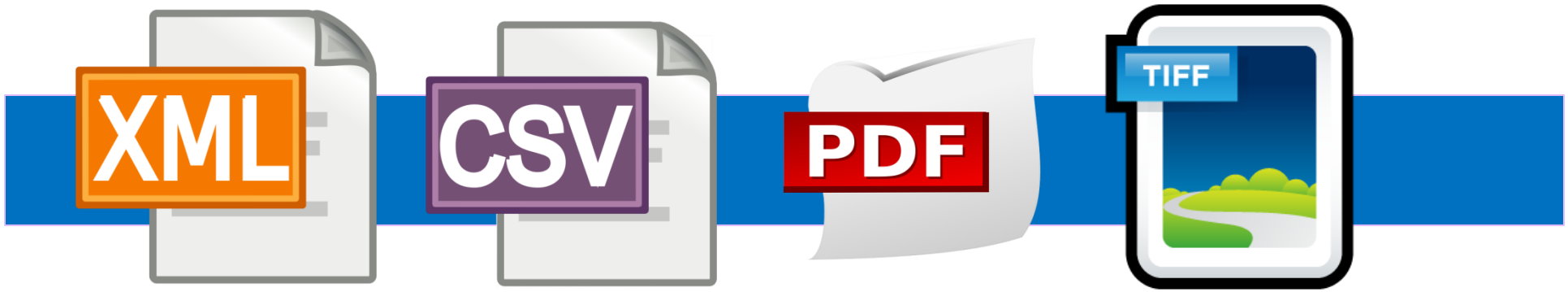


Preservation

Collection vs Dissemination

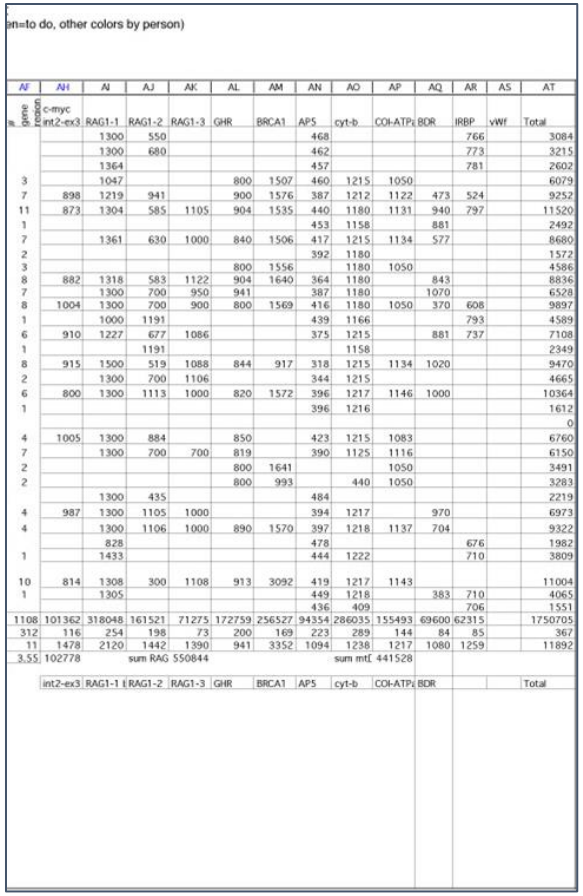


Preservation



Preservation

Microsoft Excel



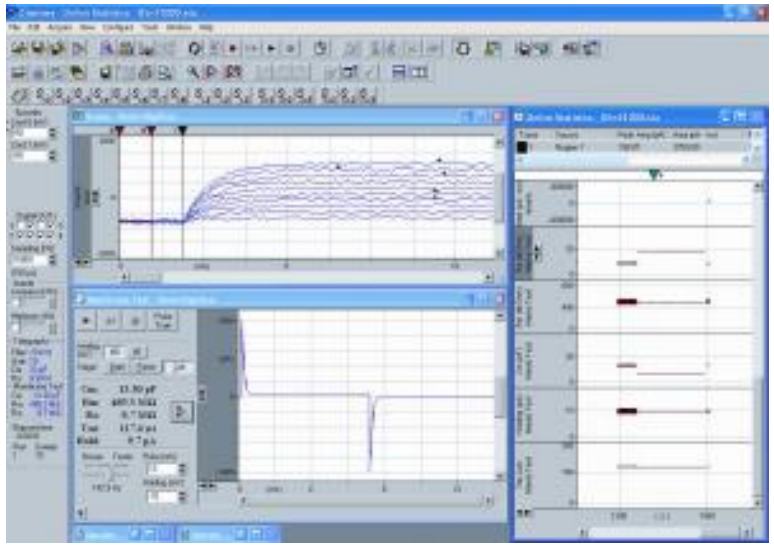
gn=to do, other colors by person)

# gene	person	c-myc int2-ex3	RAG1-1	RAG1-2	RAG1-3	GHR	BRCA1	APS	cyt-b	COI-ATPa	BDR	IRBP	vWf	Total
			1300	550				468				766		3084
			1300	680				462				773		3215
			1364					457				781		2602
3			1047			800	1507	460	1215	1050				6079
7		898	1219	941		900	1576	387	1212	1122	473	524		9252
11		873	1304	585	1105	904	1535	440	1180	1131	940	797		11520
1								453	1158		881			2492
7			1361	630	1000	840	1506	417	1215	1134	577			8680
2								392	1180					1572
3						800	1556		1180	1050				4586
8		882	1318	583	1122	904	1640	364	1180		843			8836
7			1300	700	950	941		387	1180		1070			6528
8		1004	1300	700	900	800	1569	416	1180	1050	370	608		9897
1			1000	1191				439	1166			793		4589
6		910	1227	677	1086			375	1215		881	737		7108
1				1191					1158					2349
8		915	1500	519	1088	844	917	318	1215	1134	1020			9470
2			1300	700	1106			344	1215					4665
6		800	1300	1113	1000	820	1572	396	1217	1146	1000			10364
1								396	1216					1612
														0
4		1005	1300	884		850		423	1215	1083				6760
7			1300	700	700	819		390	1125	1116				6150
2						800	1641			1050				3491
2						800	993		440	1050				3283
			1300	435				484						2219
4		987	1300	1105	1000			394	1217		970			6973
4			1300	1106	1000	890	1570	397	1218	1137	704			9322
			828					478				676		1982
1			1433					444	1222			710		3809
10		814	1308	300	1108	913	3092	419	1217	1143				11004
1			1305					449	1218		383	710		4065
								436	409			706		1551
1108		101362	318048	161521	71275	172759	256527	94354	286035	155493	69600	62315		1750705
312		116	254	198	73	200	169	223	289	144	84	85		367
11		1478	2120	1442	1390	941	3352	1094	1238	1217	1080	1259		11892
3.55		102778		sum RAG	550844				sum mt[	441528				

	int2-ex3	RAG1-1	RAG1-2	RAG1-3	GHR	BRCA1	APS	cyt-b	COI-ATPa	BDR			Total

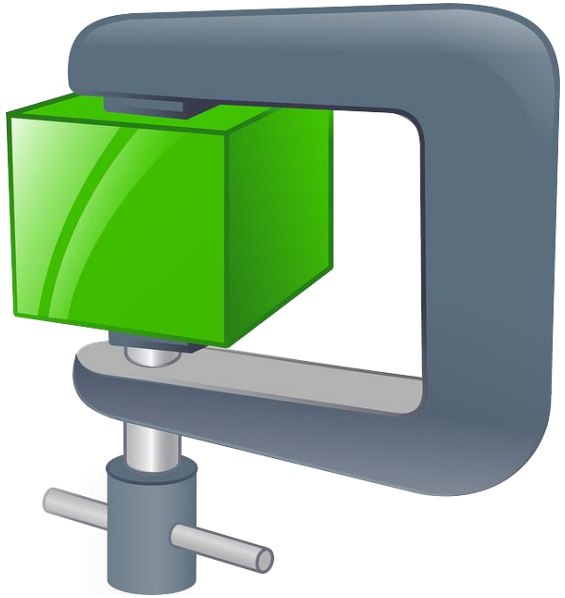
≠

Molecular Devices pClamp Software



CSV

Encryption and Compression



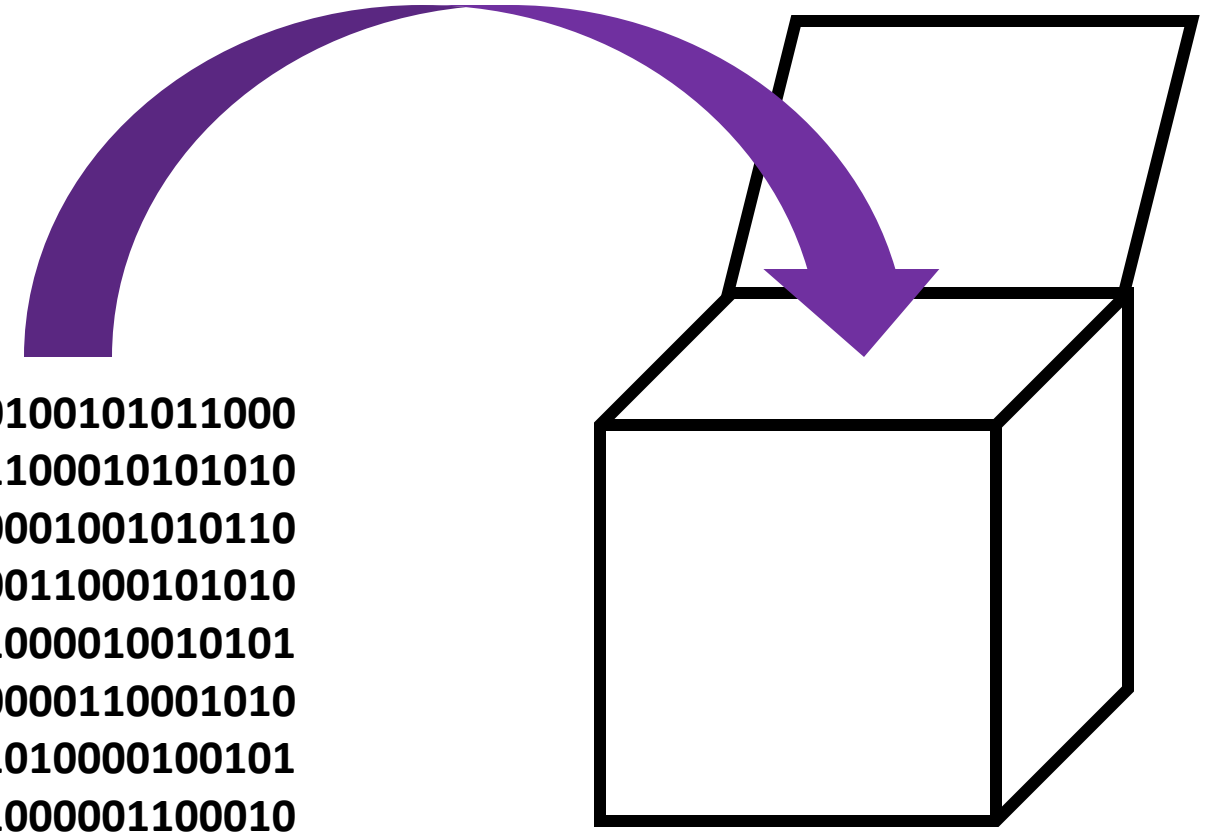
Data Ownership

Can’t assume you own your data
Check for:

- Funder policies on data ownership
- Institution policies on data ownership

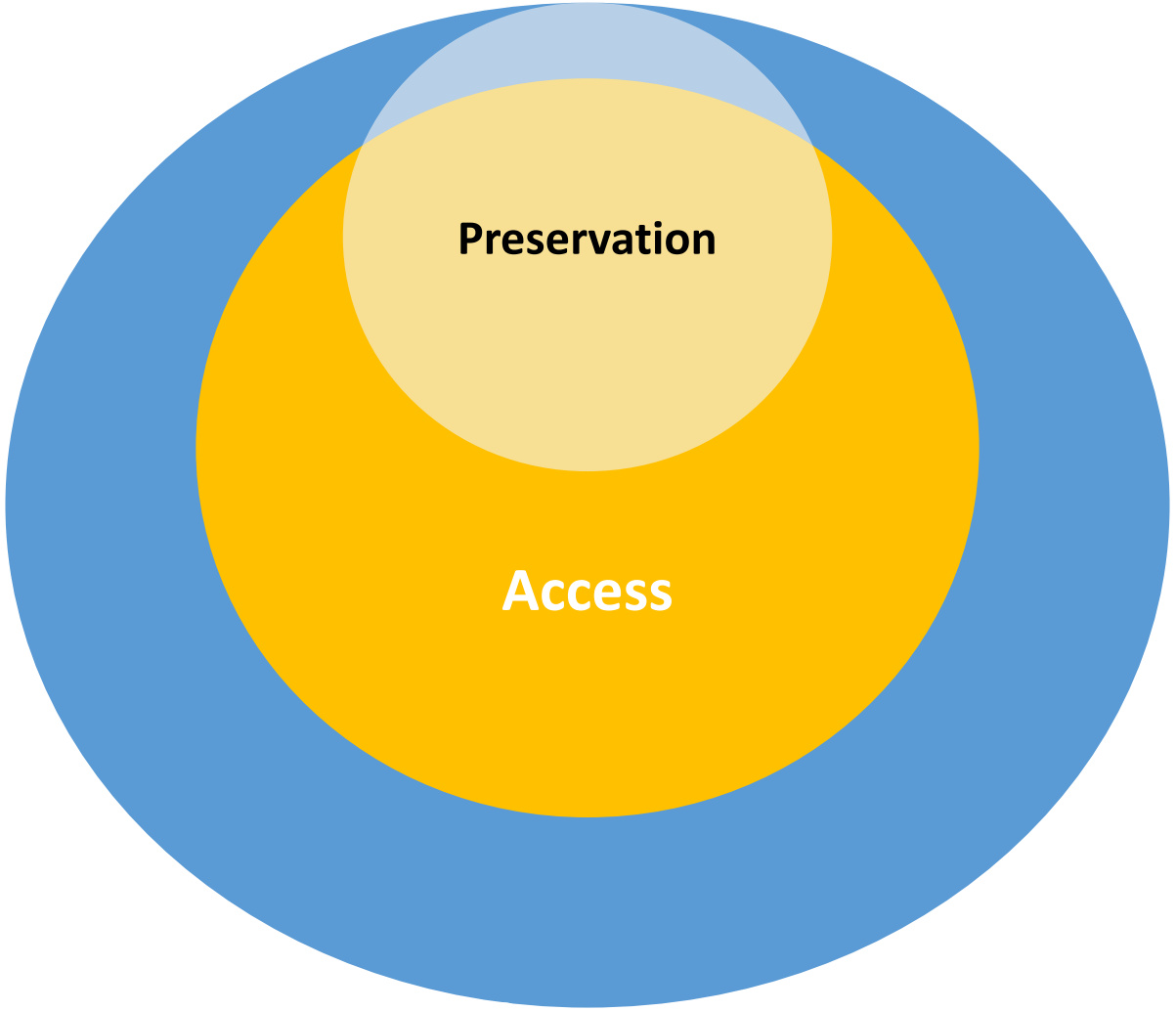
Providing Access

What is a repository?

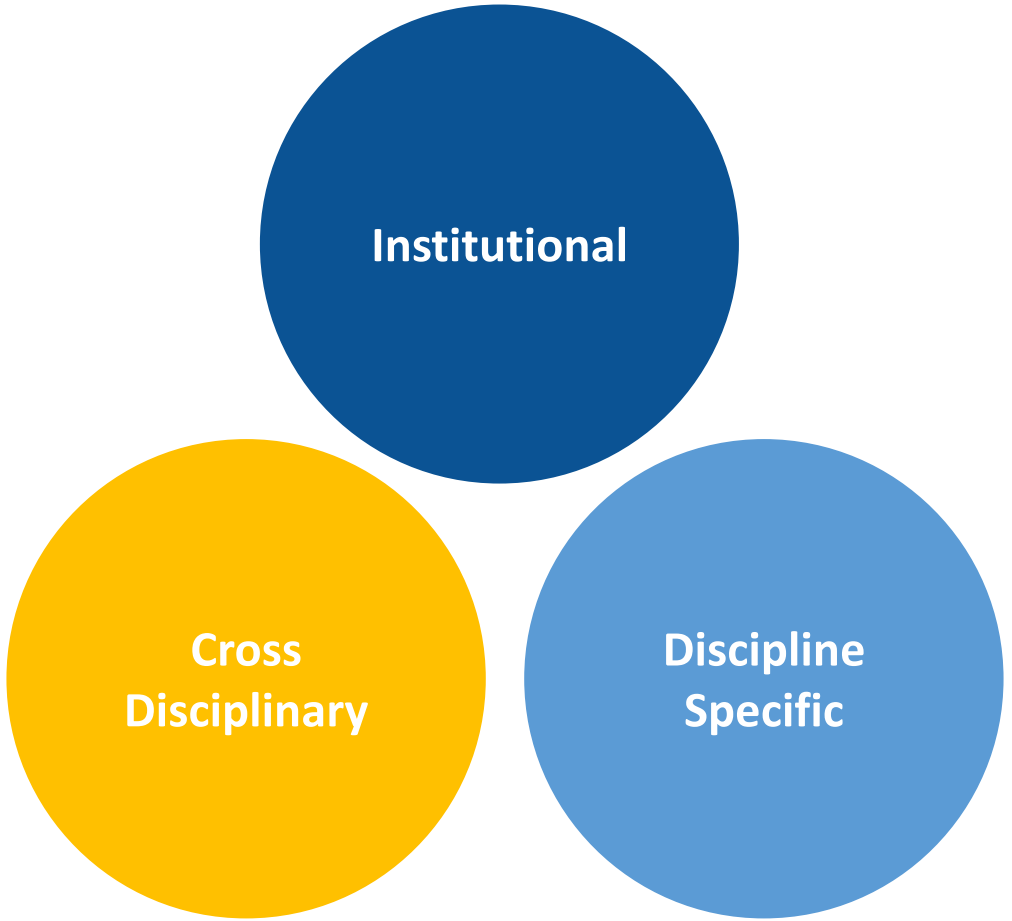


0001001010110000
0011000101010100
1000010010101100
0000110001010100
1010000100101010
1000001100010100
1010100001001010
1100000011000100
1010101001100000

Storage vs. Preservation vs. Access



Repository Types



NIH Data Sharing Repositories

NIH Data Sharing Repositories

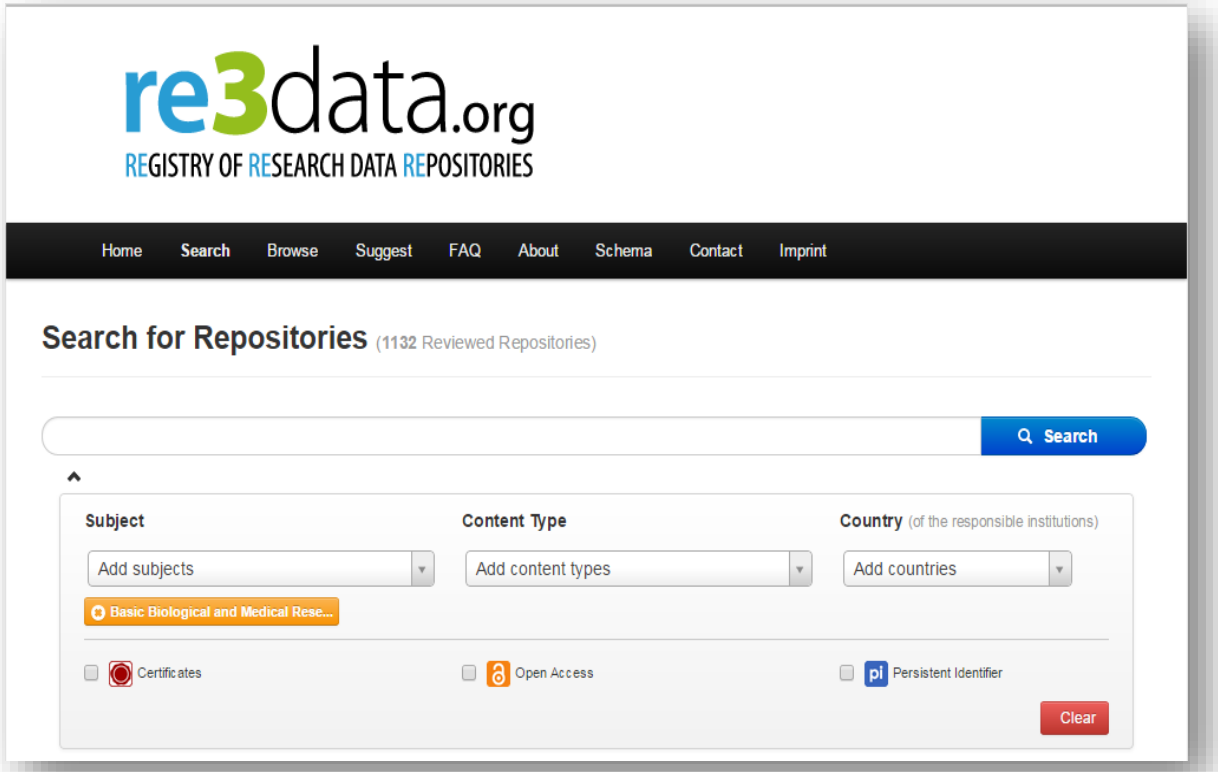
This table lists NIH-supported data repositories that accept submissions of appropriate data from NIH-funded investigators (and others). Also included are resources that aggregate information about biomedical data and information sharing systems. The table can be sorted according by name and by NIH Institute or Center and may be searched using keywords so that you can find repositories more relevant to your data. Links are provided to information about submitting data to and accessing data from the listed repositories. Additional information about the repositories and points-of-contact for further information or inquiries can be found on the websites of the individual repositories.

Show 50 entries Search:

IC	Repository Name	Repository Description	Data Submission Policy	Access to Data
NCI	The Cancer Imaging Archive (TCIA)	The Cancer Imaging Archive (TCIA) is a large archive of medical images of cancer accessible for public download. All images are stored in DICOM file format. The images are organized as "Collections", typically patients related by a common disease (e.g. lung cancer), image modality (MRI, CT, etc) or research focus.	How to Submit Data to TCIA	How to Access TCIA Data
NCI (NHGRI, NIGMS)	PeptideAtlas	PeptideAtlas is a multi-organism, publicly accessible compendium of peptides identified in a large set of tandem mass spectrometry proteomics experiments. Mass spectrometer output files are collected for human, mouse, yeast, and several other organisms, and searched using the latest search engines and protein sequences.	How to Submit Data to PeptideAtlas	How to Access PeptideAtlas Data
NHGRI	FlyBase: A Drosophila Genomic and Genetic Database	Drosophila Genomic and Genetic database that includes proteomics data, microarrays and Tiling BAC's.	How to Submit Data to FlyBase	How to Access FlyBase Data
NHGRI	The Zebrafish Model Organism Database (ZFIN)	ZFIN serves as the zebrafish model organism database. It aims to: a) be the community database resource for the laboratory use of zebrafish, b) develop and support integrated zebrafish genetic, genomic and developmental information, c) maintain the definitive reference data sets of zebrafish research information, d) to link this information extensively to corresponding data in other model organism and human databases, e) facilitate the use of zebrafish as a model for human biology, and f) serve the needs of the research community.	How to Submit Data to ZFIN	How to Access ZFIN Data
NHGRI	WormBase	WormBase is an international consortium of biologists and computer scientists dedicated to providing the research community with accurate, current, accessible information concerning the genetics, genomics and biology of C. elegans and related nematodes.	How to Submit Data to WormBase	How to Access WormBase Data

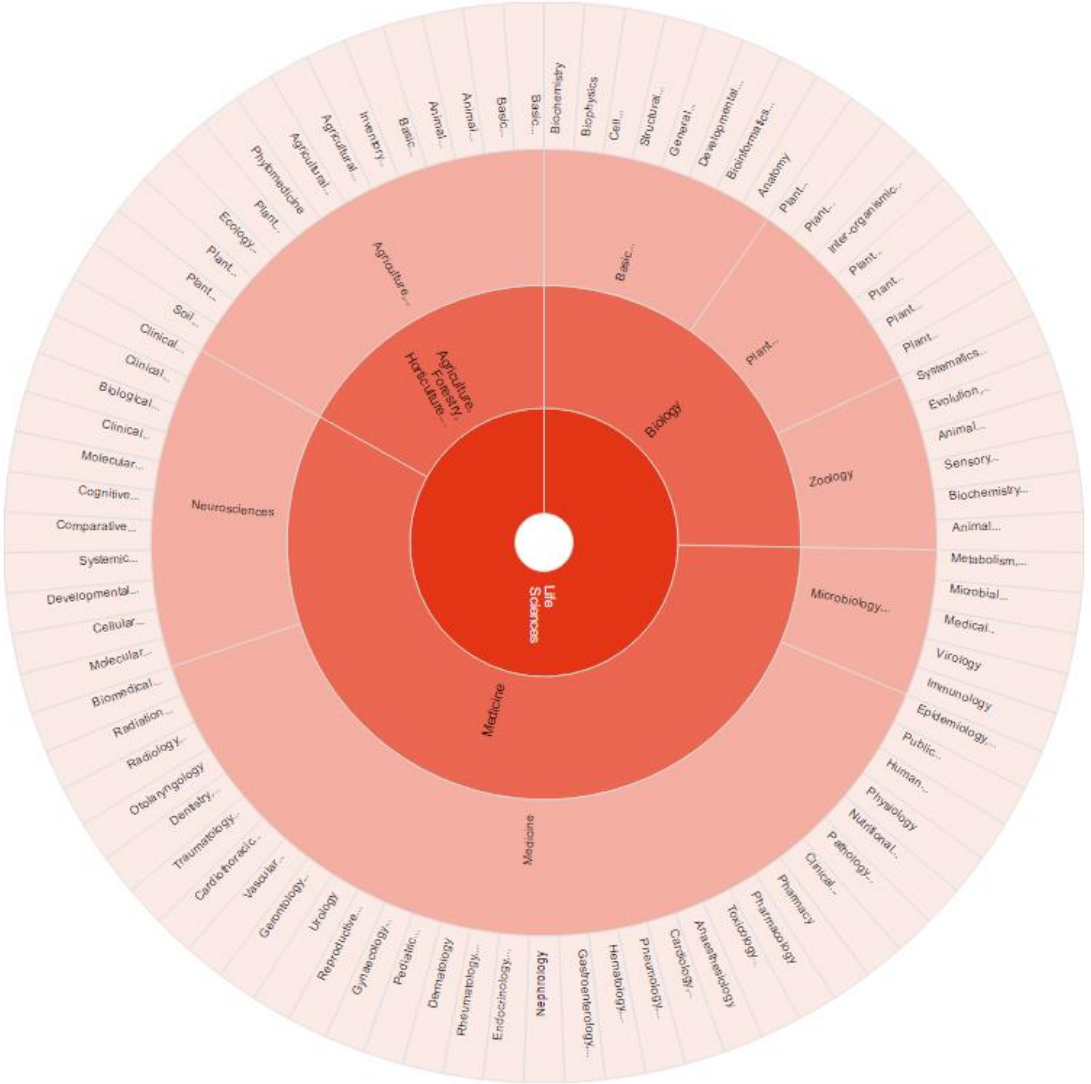
http://www.nlm.nih.gov/NIHbmic/nih_data_sharing_repositories.html

Research Data Repositories

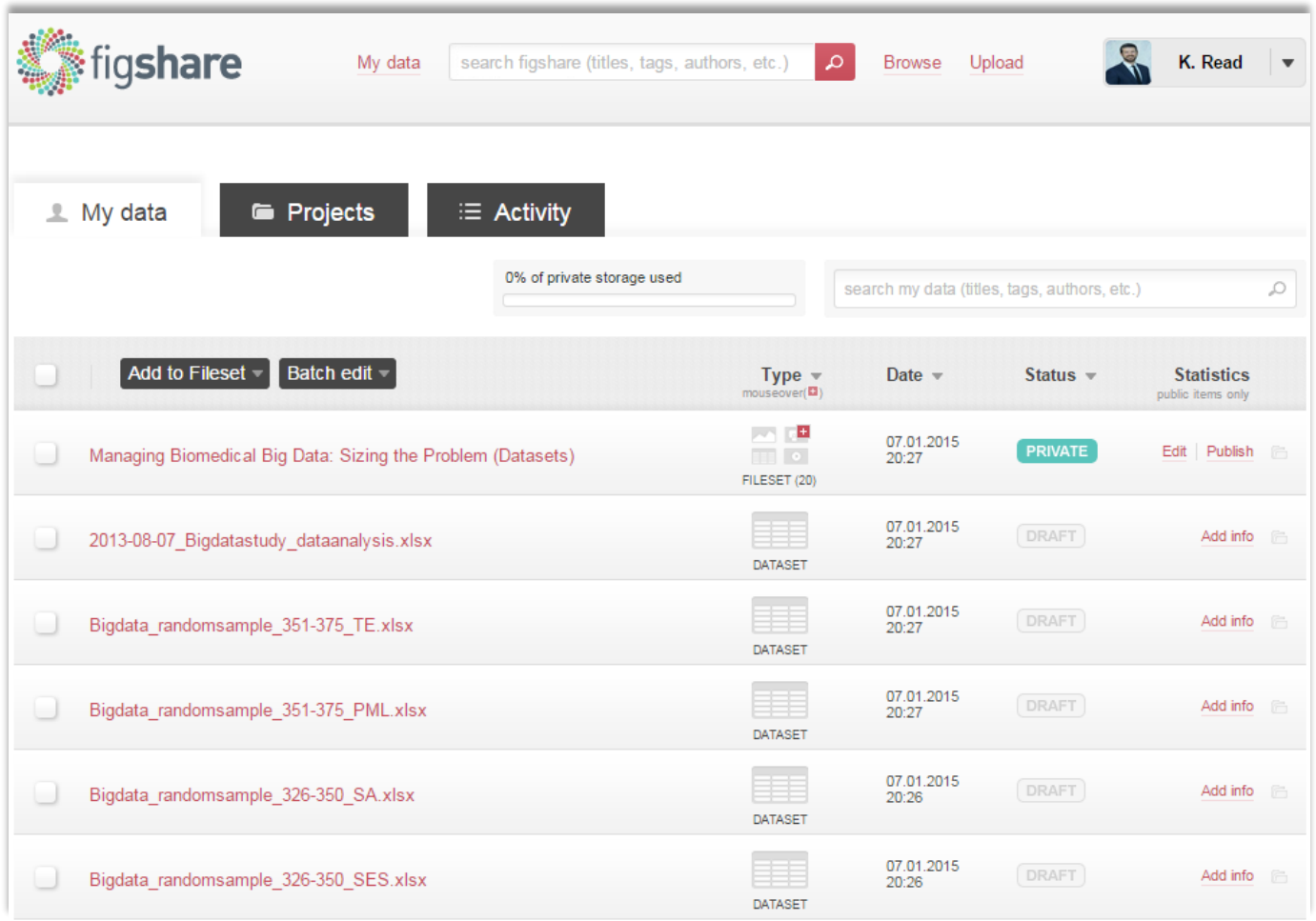


http://www.re3data.org/

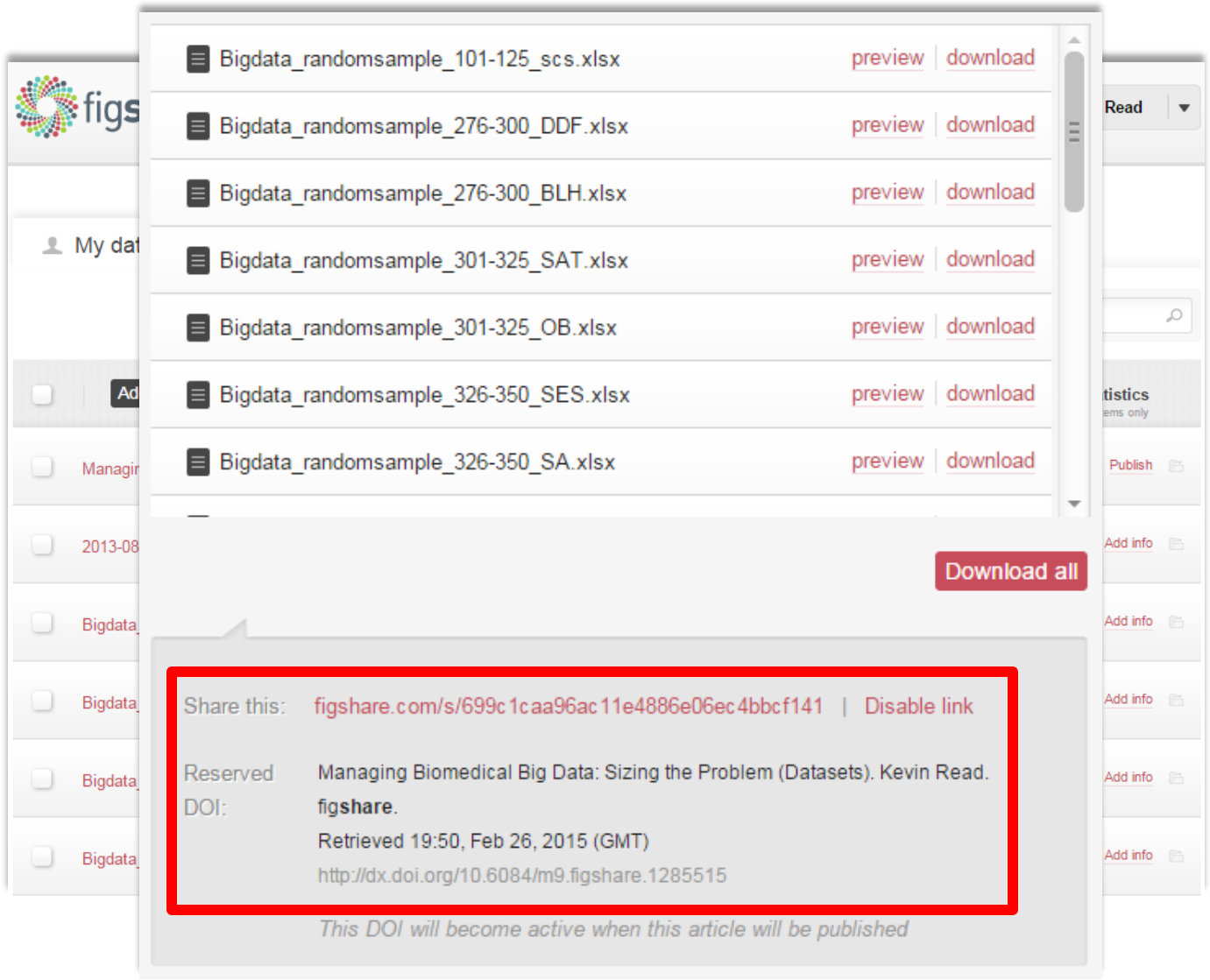
Research Data Repositories



Figshare

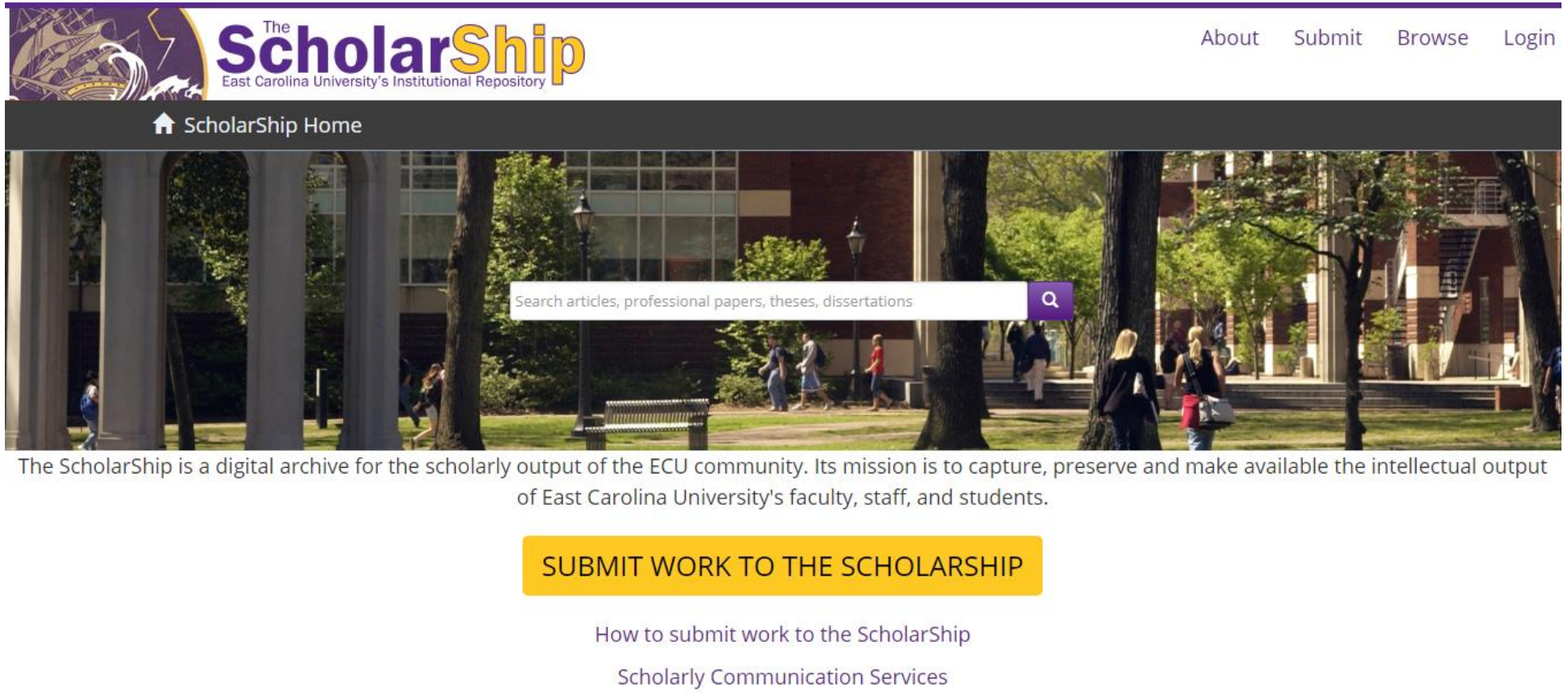

figshare
My data
search figshare (titles, tags, authors, etc.)
Browse
Upload
K. Read
My data
Projects
Activity
0% of private storage used
search my data (titles, tags, authors, etc.)
Add to Fileset
Batch edit
Type
mouseover
Date
Status
Statistics
public items only
Managing Biomedical Big Data: Sizing the Problem (Datasets)
FILESET (20)
07.01.2015 20:27
PRIVATE
Edit
Publish
2013-08-07_Bigdatastudy_dataanalysis.xlsx
DATASET
07.01.2015 20:27
DRAFT
Add info
Bigdata_randomsample_351-375_TE.xlsx
DATASET
07.01.2015 20:27
DRAFT
Add info
Bigdata_randomsample_351-375_PML.xlsx
DATASET
07.01.2015 20:27
DRAFT
Add info
Bigdata_randomsample_326-350_SA.xlsx
DATASET
07.01.2015 20:26
DRAFT
Add info
Bigdata_randomsample_326-350_SES.xlsx
DATASET
07.01.2015 20:26
DRAFT
Add info

Figshare


Bigdata_randomsample_101-125_scs.xlsx preview download
Bigdata_randomsample_276-300_DDF.xlsx preview download
Bigdata_randomsample_276-300_BLH.xlsx preview download
Bigdata_randomsample_301-325_SAT.xlsx preview download
Bigdata_randomsample_301-325_OB.xlsx preview download
Bigdata_randomsample_326-350_SES.xlsx preview download
Bigdata_randomsample_326-350_SA.xlsx preview download
Download all
Share this: figshare.com/s/699c1caa96ac11e4886e06ec4bbcf141 | Disable link
Reserved DOI: Managing Biomedical Big Data: Sizing the Problem (Datasets). Kevin Read. figshare.
Retrieved 19:50, Feb 26, 2015 (GMT)
http://dx.doi.org/10.6084/m9.figshare.1285515
This DOI will become active when this article will be published

Institutional Repositories

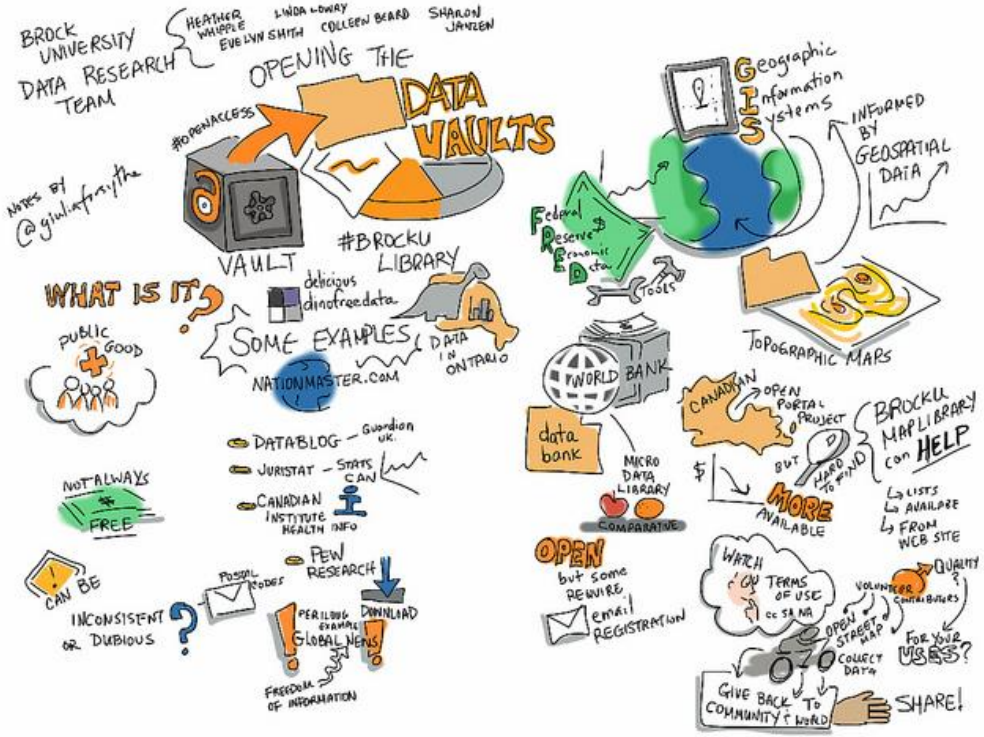
The ScholarShip http://thescholarship.ecu.edu/



Data Publishing

Providing access to your data

- Access vs. **meaningful** access
- Well-documented data
- Using standards



Avoiding Data Dumpsters

"One of the risks posed by these expanding repositories is the production of "**data dumpsters**": repositories of data without the metadata, data dictionaries, or documentation needed for **meaningful** or correct reanalysis."

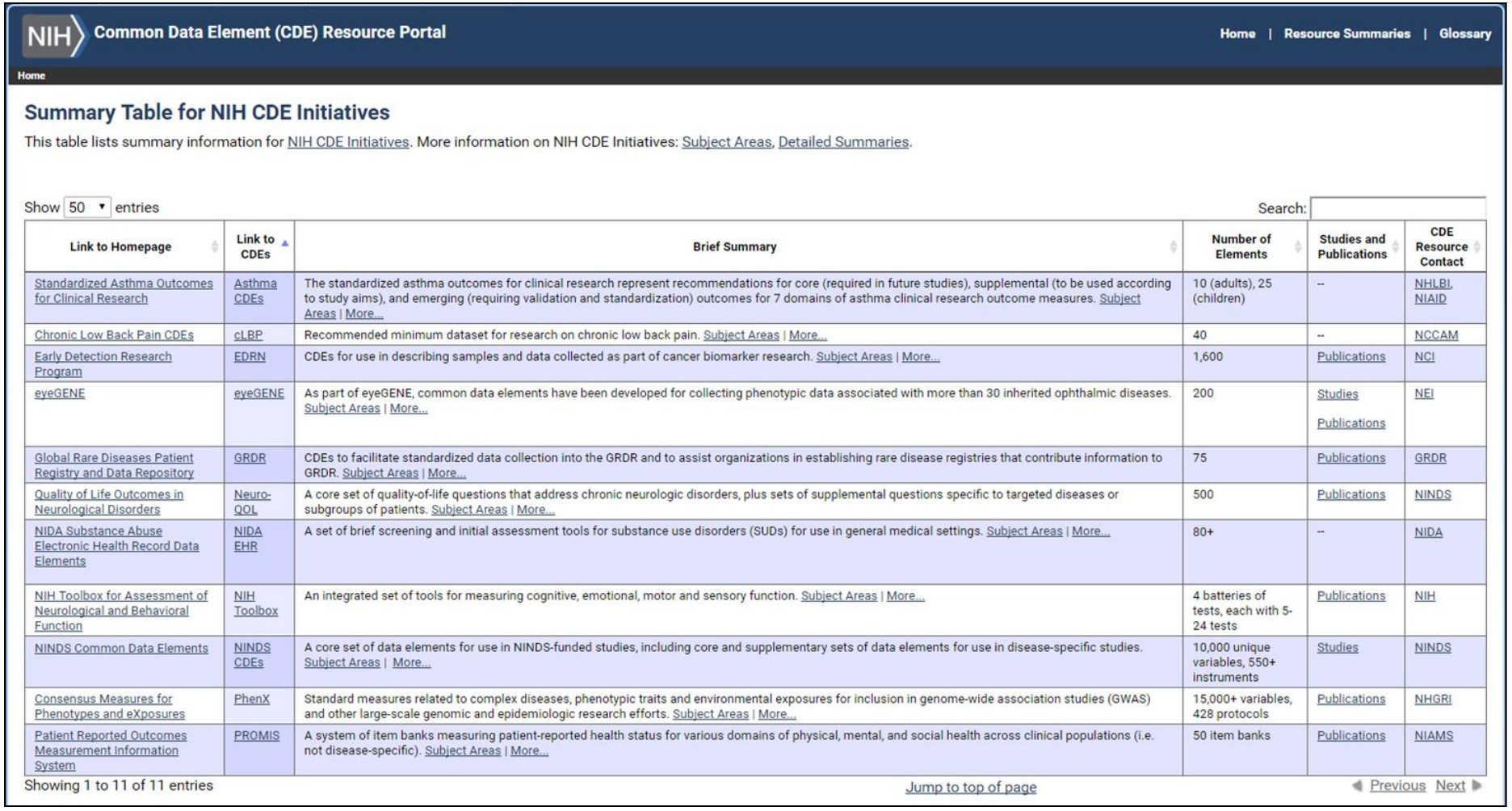
http://www.nejm.org/doi/full/10.1056/NEJMp1605148

What are data standards?

Data standards define:

- what to collect
- how to represent the data
- a vocabulary to use
- how to communicate the data

NIH Common Data Elements



NIH Common Data Element (CDE) Resource Portal

Home | Resource Summaries | Glossary

Home

Summary Table for NIH CDE Initiatives

This table lists summary information for NIH CDE Initiatives. More information on NIH CDE Initiatives: Subject Areas, Detailed Summaries.

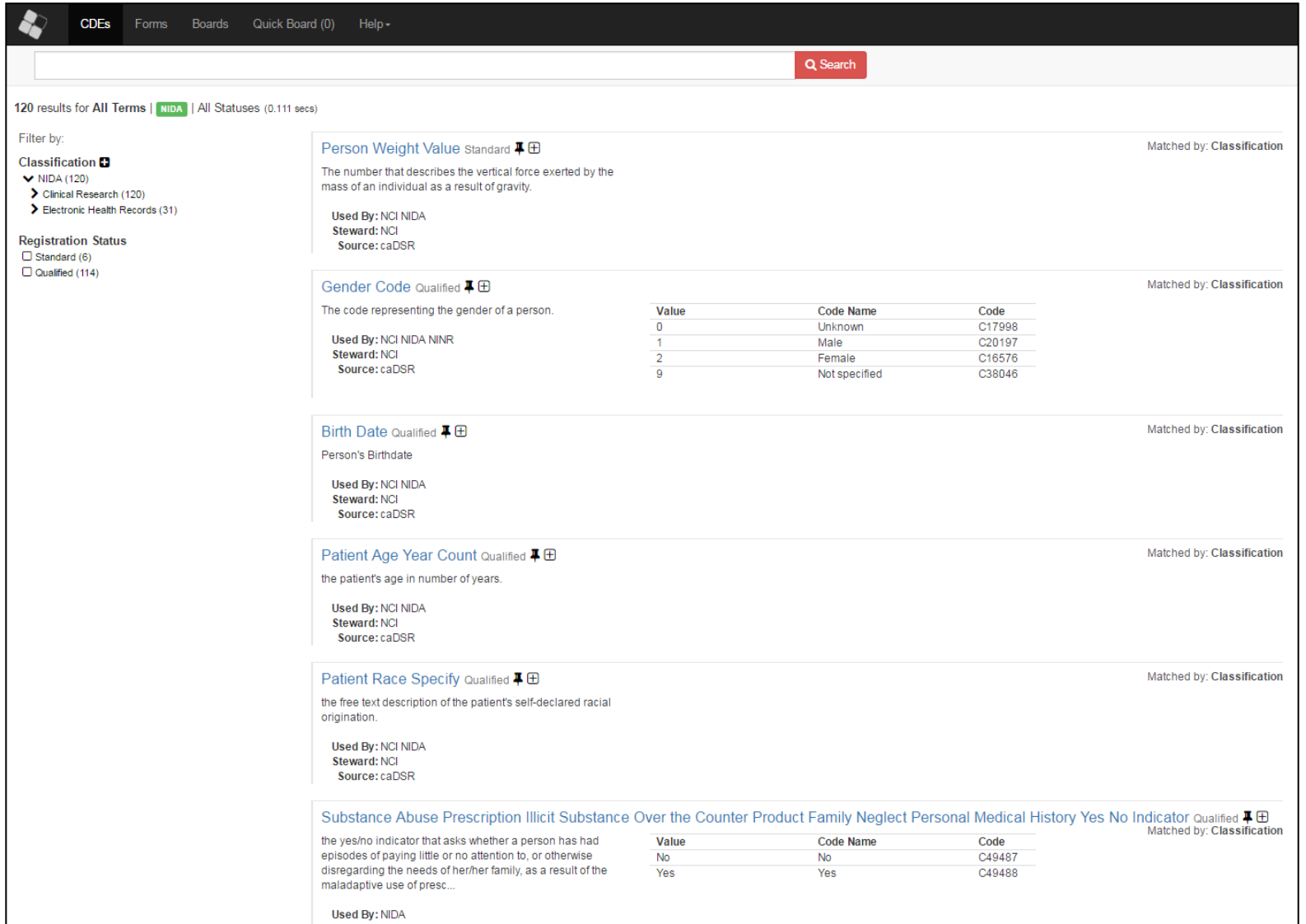
Show 50 entries Search:

Link to Homepage	Link to CDEs	Brief Summary	Number of Elements	Studies and Publications	CDE Resource Contact
Standardized Asthma Outcomes for Clinical Research	Asthma CDEs	The standardized asthma outcomes for clinical research represent recommendations for core (required in future studies), supplemental (to be used according to study aims), and emerging (requiring validation and standardization) outcomes for 7 domains of asthma clinical research outcome measures. Subject Areas | More...	10 (adults), 25 (children)	--	NHLBI, NIAID
Chronic Low Back Pain CDEs	cLBP	Recommended minimum dataset for research on chronic low back pain. Subject Areas | More...	40	--	NCCAM
Early Detection Research Program	EDRN	CDEs for use in describing samples and data collected as part of cancer biomarker research. Subject Areas | More...	1,600	Publications	NCI
eyeGENE	eyeGENE	As part of eyeGENE, common data elements have been developed for collecting phenotypic data associated with more than 30 inherited ophthalmic diseases. Subject Areas | More...	200	Studies Publications	NEI
Global Rare Diseases Patient Registry and Data Repository	GRDR	CDEs to facilitate standardized data collection into the GRDR and to assist organizations in establishing rare disease registries that contribute information to GRDR. Subject Areas | More...	75	Publications	GRDR
Quality of Life Outcomes in Neurological Disorders	Neuro-QOL	A core set of quality-of-life questions that address chronic neurologic disorders, plus sets of supplemental questions specific to targeted diseases or subgroups of patients. Subject Areas | More...	500	Publications	NINDS
NIDA Substance Abuse Electronic Health Record Data Elements	NIDA EHR	A set of brief screening and initial assessment tools for substance use disorders (SUDs) for use in general medical settings. Subject Areas | More...	80+	--	NIDA
NIH Toolbox for Assessment of Neurological and Behavioral Function	NIH Toolbox	An integrated set of tools for measuring cognitive, emotional, motor and sensory function. Subject Areas | More...	4 batteries of tests, each with 5-24 tests	Publications	NIH
NINDS Common Data Elements	NINDS CDEs	A core set of data elements for use in NINDS-funded studies, including core and supplementary sets of data elements for use in disease-specific studies. Subject Areas | More...	10,000 unique variables, 550+ instruments	Studies	NINDS
Consensus Measures for Phenotypes and eXposures	PhenX	Standard measures related to complex diseases, phenotypic traits and environmental exposures for inclusion in genome-wide association studies (GWAS) and other large-scale genomic and epidemiologic research efforts. Subject Areas | More...	15,000+ variables, 428 protocols	Publications	NHGRI
Patient Reported Outcomes Measurement Information System	PROMIS	A system of item banks measuring patient-reported health status for various domains of physical, mental, and social health across clinical populations (i.e. not disease-specific). Subject Areas | More...	50 item banks	Publications	NIAMS

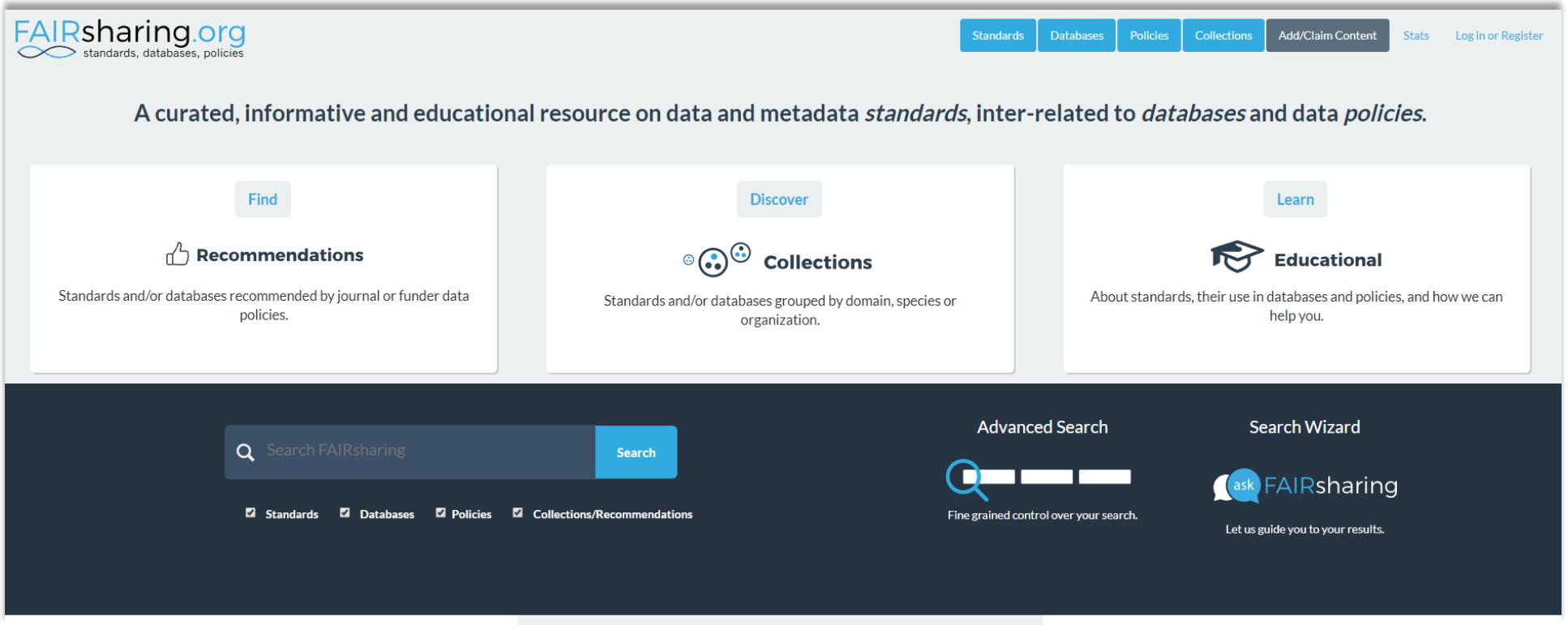
Showing 1 to 11 of 11 entries Jump to top of page Previous Next

https://www.nlm.nih.gov/cde/

NIH Common Data Elements



FAIRsharing.org



Clinical Data Acquisition Standards Harmonization (CDASH)

Table of Contents

Section **Page**

6.0 Change Control and the Process for Creating New CDASH Domains 162

7.0 Appendices 162

7.1 Appendix A: Commonly Used CDISC Controlled Terminology 162

7.2 Appendix B: Regulatory References 169

Appendix B1: Common Identifiers and Timing Variables 170

Appendix B2: Adverse Events (AE) 171

Appendix B3: Prior and Concomitant Medications (CM) 175

Appendix B4: Demographics (DM) 176

Appendix B5: Disposition (DS) 177

Appendix B6: Drug Accountability (DA) 178

Appendix B7: ECG Test Results (EG) 179

Appendix B8: Exposure (EX) 180

Appendix B9: Inclusion/Exclusion Criteria Not Met (IE) 181

Appendix B10: Laboratory Test Results (LB) 182

Appendix B11: Medical History (MH) 183

Appendix B12: Physical Examination (PE) 184

Appendix B13: Protocol Deviations (DV) 185

Appendix B14: Substance Use (SU) 186

Appendix B15: Vital Signs (VS) 187

7.3 Appendix C: CDASH Contributors 188

Appendix C1: CDASH Leadership Team 188

Appendix C2: CDASH v1.1 Team Contributors 188

Appendix C3: Participating Companies 189

7.4 Appendix D: List of Abbreviations and Glossary 191

7.5 Appendix E: CDASH Variable Naming Fragments 195

7.6 Appendix F: Acknowledgements 196

7.7 Appendix G: Changes from V1.0 to V1.1 197

7.8 Appendix H: Representation and Warranties, Limitations of Liability, and Disclaimers 202

Appendix H1: CDISC Patent Disclaimers 202

Appendix H2: Representations and Warranties 202

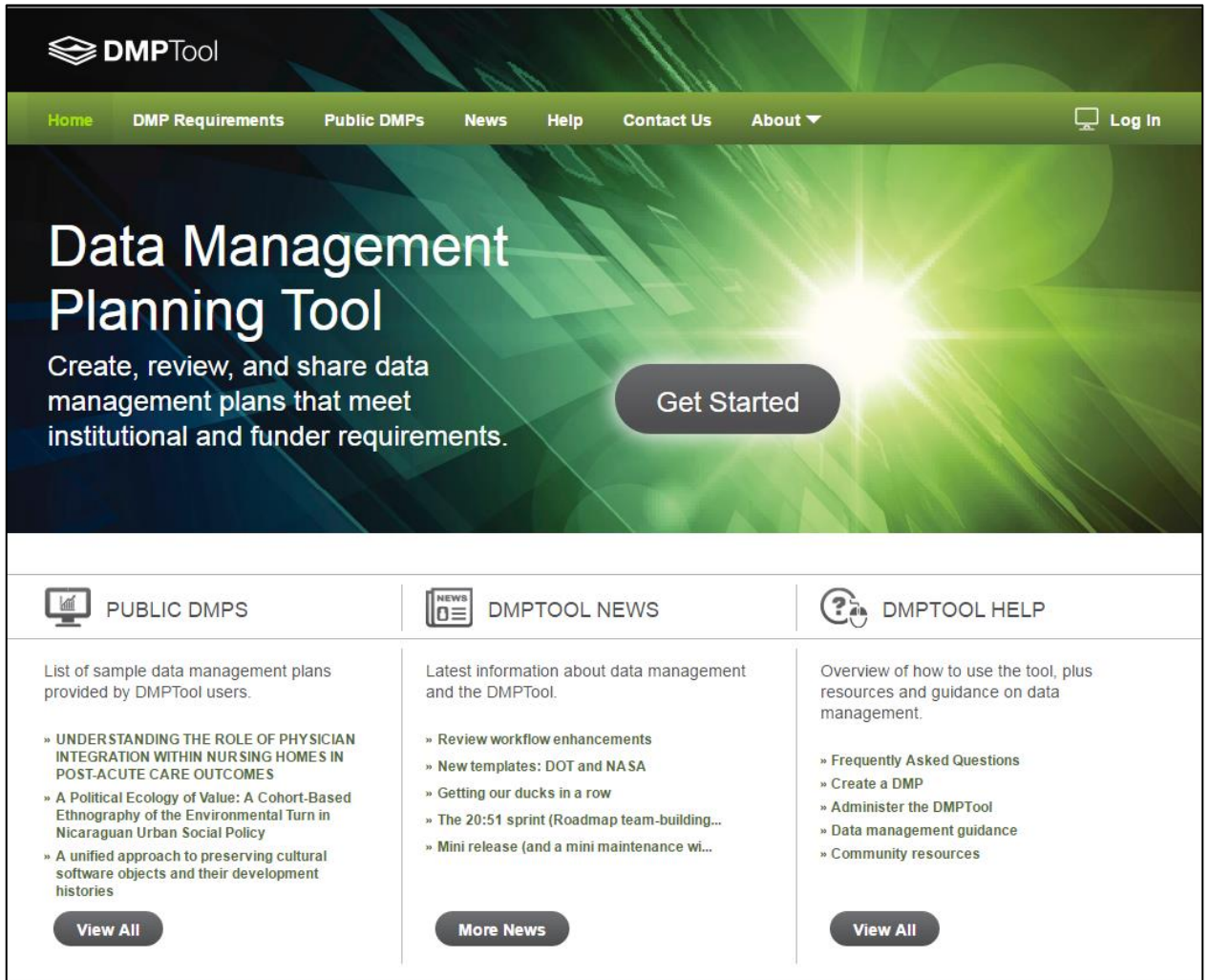
Appendix H3: No Other Warranties/Disclaimers 202

Appendix H4: Limitation of Liability 202

Tools for Data Management

Data Management Planning Tool

https://dmptool.org



National Institutes of Health

DMP Tool: Templates

DMP Tool

New template: NIH Genomic Data Sharing

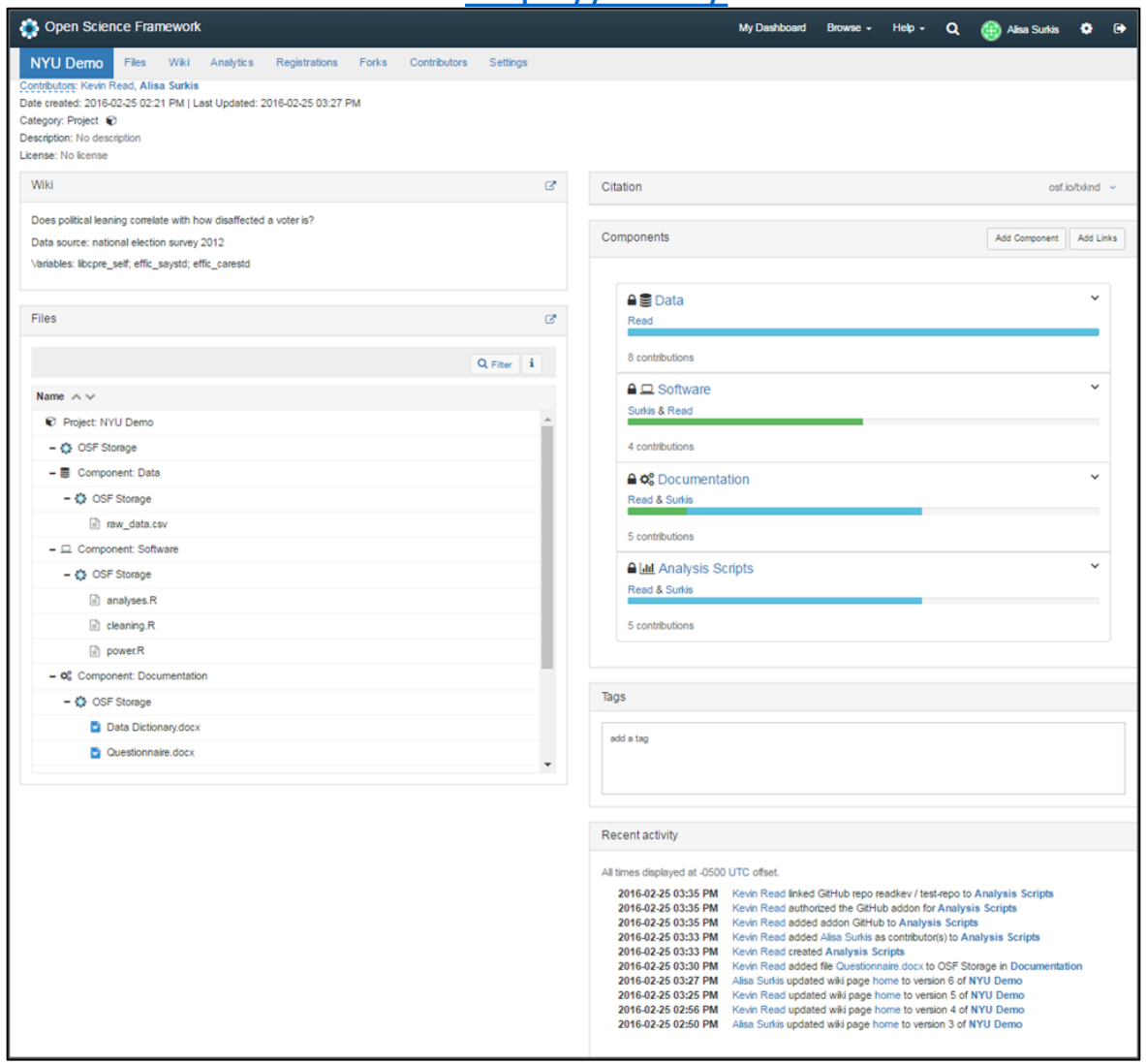
The National Institutes of Health issued new Guidance for Investigators in Developing Genomic Data Sharing Plans along with some helpful sample plans (dated July 14, 2015). The DMPTool team has been monitoring the responses to the OSTP memo by federal agencies, but this alert came to us via the DMP admin email list. Please continue to let us know when you hear anything at all (see the links from a previous post "How you Can Help")!

We added a National Institutes of Health drop-down list to the DMPTool that contains the new NIH-GDS: Genomic Data Sharing template in addition to the NIH-GEN: Generic template. The basic data management requirements for most NIH grants remain unchanged (pending further notice); researchers can continue to use the generic template for most grants.

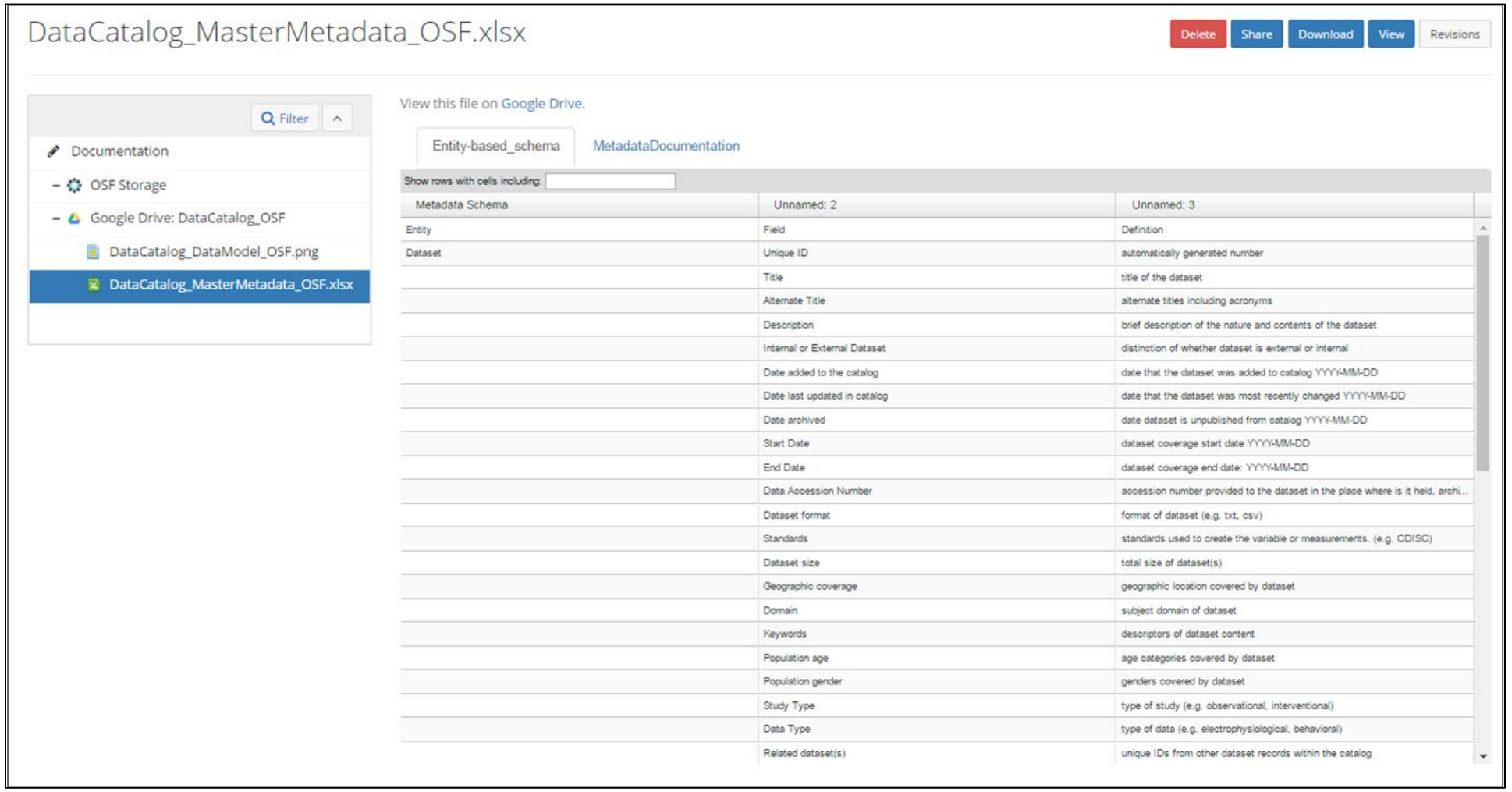
The new guidance pertains to those proposing research that will generate large-scale human and non-human genomic data. It describes the type of information that should be provided in a genomic data sharing plan and when the plan should be submitted, including instructions for IRB review, appropriate uses of the data, and suggested/required data repositories. The new guidance is an update to the existing NIH GDS Policy that became effective on January 25, 2015.

Open Science Framework

https://osf.io/

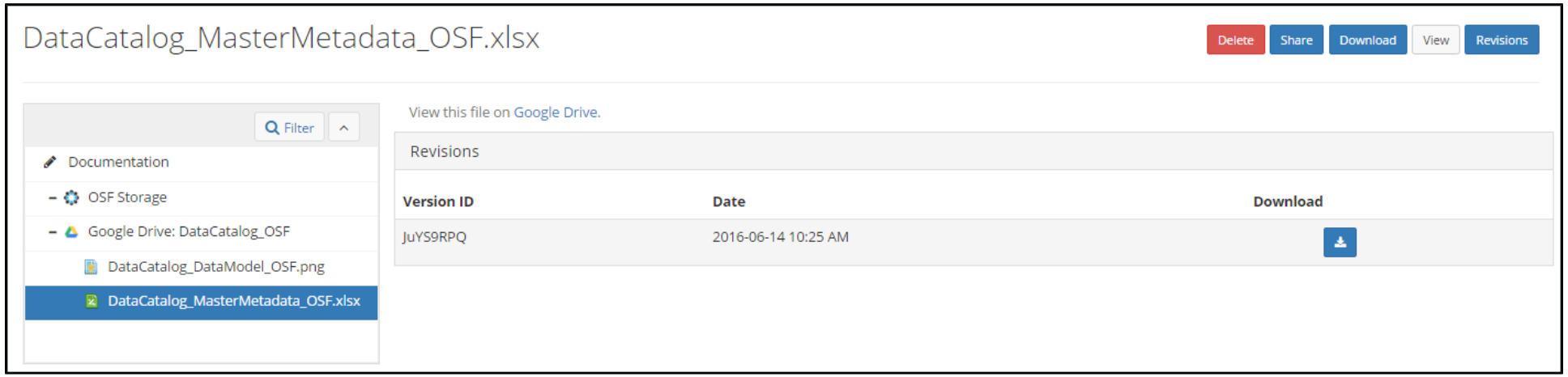


Open Science Framework

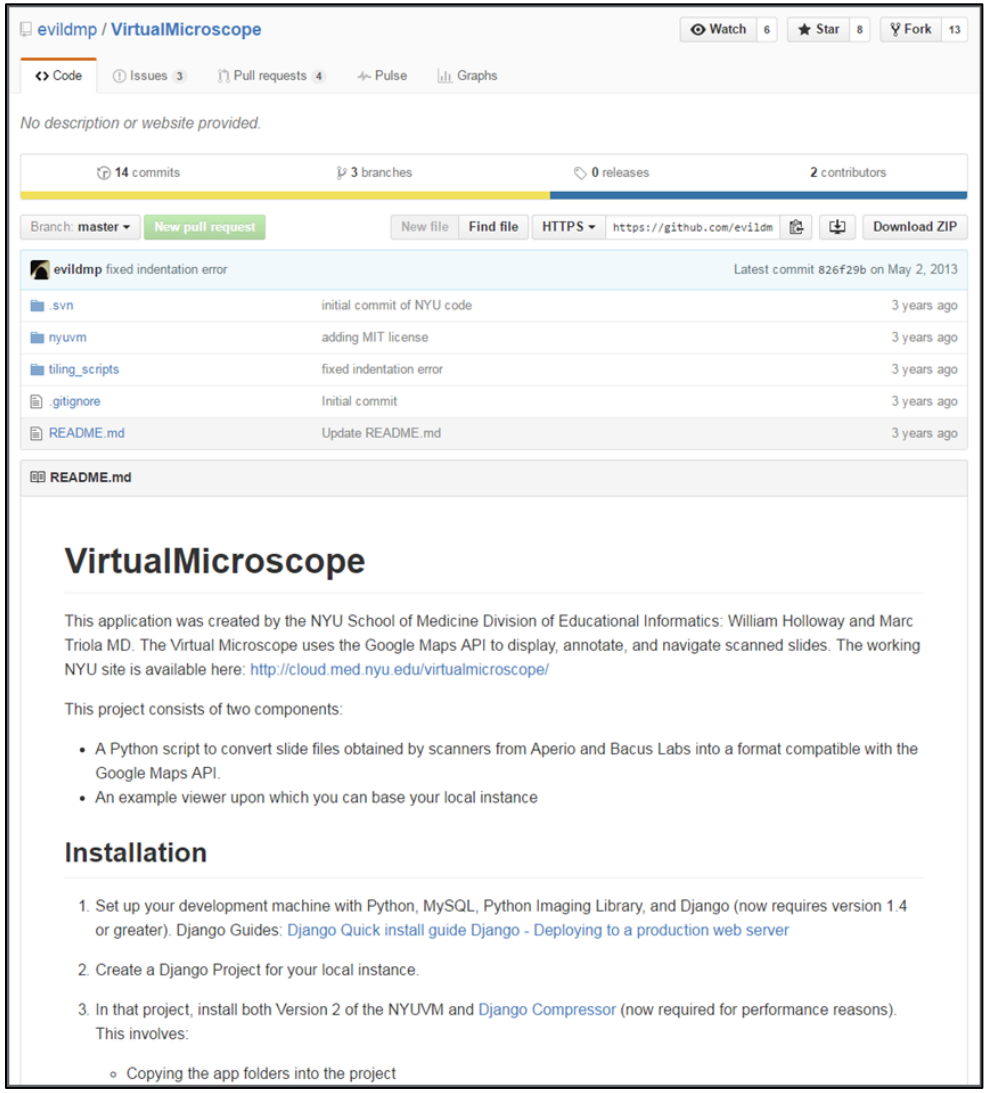


Metadata Schema	Unnamed: 2	Unnamed: 3
Entity	Field	Definition
Dataset	Unique ID	automatically generated number
	Title	title of the dataset
	Alternate Title	alternate titles including acronyms
	Description	brief description of the nature and contents of the dataset
	Internal or External Dataset	distinction of whether dataset is external or internal
	Date added to the catalog	date that the dataset was added to catalog YYYY-MM-DD
	Date last updated in catalog	date that the dataset was most recently changed YYYY-MM-DD
	Date archived	date dataset is unpublished from catalog YYYY-MM-DD
	Start Date	dataset coverage start date YYYY-MM-DD
	End Date	dataset coverage end date: YYYY-MM-DD
	Data Accession Number	accession number provided to the dataset in the place where is it held, archi...
	Dataset format	format of dataset (e.g. txt, csv)
	Standards	standards used to create the variable or measurements. (e.g. CDISC)
	Dataset size	total size of dataset(s)
	Geographic coverage	geographic location covered by dataset
	Domain	subject domain of dataset
	Keywords	descriptors of dataset content
	Population age	age categories covered by dataset
	Population gender	genders covered by dataset
	Study Type	type of study (e.g. observational, interventional)
	Data Type	type of data (e.g. electrophysiological, behavioral)
	Related dataset(s)	unique IDs from other dataset records within the catalog

Open Science Framework



GitHub



evildmp / VirtualMicroscope

Watch 6 Star 8 Fork 13

Code Issues 3 Pull requests 4 Pulse Graphs

No description or website provided.

14 commits 3 branches 0 releases 2 contributors

Branch: master New pull request New file Find file HTTPS https://github.com/evildm Download ZIP

evildmp fixed indentation error Latest commit 826f29b on May 2, 2013

.svn	initial commit of NYU code	3 years ago
nyuvm	adding MIT license	3 years ago
tiling_scripts	fixed indentation error	3 years ago
.gitignore	Initial commit	3 years ago
README.md	Update README.md	3 years ago

README.md

VirtualMicroscope

This application was created by the NYU School of Medicine Division of Educational Informatics: William Holloway and Marc Triola MD. The Virtual Microscope uses the Google Maps API to display, annotate, and navigate scanned slides. The working NYU site is available here: http://cloud.med.nyu.edu/virtualmicroscope/

This project consists of two components:

- A Python script to convert slide files obtained by scanners from Aperio and Bacus Labs into a format compatible with the Google Maps API.
- An example viewer upon which you can base your local instance

Installation

1. Set up your development machine with Python, MySQL, Python Imaging Library, and Django (now requires version 1.4 or greater). Django Guides: Django Quick install guide Django - Deploying to a production web server
2. Create a Django Project for your local instance.
3. In that project, install both Version 2 of the NYUVM and Django Compressor (now required for performance reasons). This involves:
    - Copying the app folders into the project

GitHub

RepGerryConnolly authored 22 days ago 1 parent a94b2bb commit d4c74352752eba6e4a01345c235b7569bc2c4f1e

Showing 1 changed file with 2 additions and 2 deletions. Unified Split

4 index.md View

@@ -317,9 +317,9 @@ BUDGET FORMULATION.—The Director of the Office of Management and Budget shall

```
317 317
318 318  ### Organization and Workforce
319 319
320     -##### CIO approves new bureau CIOs
    320 +##### CIO approves bureau CIOs
321 321
322     - - M1. *CIO Role/Responsibility*: CIO approves new bureau CIOs. The CIO shall be involved in the recruitment and shall approve the selection of any new bureau CIO (includes bureau leadership with CIO duties but not title—see definitions).
    322 + - M1. *CIO Role/Responsibility*: CIO approves bureau CIOs. The CIO shall approve the selection of any bureau CIO (includes bureau leadership with CIO duties but not title—see definitions). The CIO shall also be involved in the recruitment and shall approve the selection of any new bureau CIO.
323 323   - *Statutory Language*: PERSONNEL-RELATED AUTHORITY.—Notwithstanding any other provision of law, for each covered agency … the Chief Information Officer of the covered agency shall approve the appointment of any other employee with the title of Chief Information Officer, or who functions in the capacity of a Chief Information Officer, for any component organization within the covered agency. 40 U.S.C. 11319 (b)(2)
324 324
325 325  ##### CIO role in ongoing bureau CIOs' evaluations
```

Lab Notebooks

labarchives Chance Favors the Organized Lab	General Purpose
labguru	Inventory Management
Benchling	DNA tool integration

Conclusions

- Plan data management before starting research
- Can’t ignore the march toward research data management and sharing

Conclusions

- Plan data management before starting research
- Can't ignore the march toward research data management and sharing

DOCUMENT

Conclusions

- Plan data management before starting research
- Can't ignore the march toward research data management and sharing

Conclusions

- Plan data management before starting research
- Can't ignore the mandate toward research data management and sharing

Thank you!

Email: browderk@ecu.edu.

Photo references

- Mike. “Padlock” 2005 https://www.flickr.com/photos/zebble/6080622/
- Theen Moy. “SAHMRI Lab” 2014 https://www.flickr.com/photos/theenmoy/13969480280/
- Kevin Dooley. “Future tense” 2009 https://www.flickr.com/photos/pagedooley/4114167117/
- Wikipedia.org. “To Share or Not to Share” 2013 http://upload.wikimedia.org/wikipedia/commons/6/65/To_deposit_or_not_to_deposit,_that_is_the_question_-_journal.pbio.1001779.g001.png
- Opensource.com “What open data means – and what it doesn’t” 2010 https://www.flickr.com/photos/opensourceway/5265955179/
- Pixabay. “Test” 2013 http://pixabay.com/static/uploads/photo/2013/11/25/17/27/test-218181_640.jpg
- Pixabay. “Research” 2012 http://pixabay.com/p-66365/?no_redirect
- Draw&Research. “Versioncontrol” 2013 https://www.flickr.com/photos/dibujoscccd/8659059842/
- Wikimedia. “Backup Backup Backup” 2011 http://upload.wikimedia.org/wikipedia/commons/e/ee/Backup_Backup_Backup_-_And_Test_Restores.jpg
- Wikimedia. “Modern warehouse pallent rack storage system” 2000 http://upload.wikimedia.org/wikipedia/commons/a/a2/Modern_warehouse_with_pallet_rack_storage_system.jpg
- Mcability. “Padlock Unlock” 2014 http://pixabay.com/en/padlock-lock-security-unlock-322494/
- Gadini. “Padlocks Locks for Bags” 2015 http://pixabay.com/en/padlocks-locks-for-bags-597815/
- Simonatova. “Egyptian museum mummy antiquity” 2015 http://pixabay.com/en/egyptian-museum-mummy-antiquity-630297/
- Sarah. “Metadata is a love note to the future” 2011 https://www.flickr.com/photos/sarahseverson/6245395188/
- Jonas Ahrentorp. “Clutter” 2007 https://www.flickr.com/photos/neofob/753826914/
- Alan Levine. “Life is sharing” 2012 https://www.flickr.com/photos/cogdog/8188824613/
- Niklas Wilkstrom. “Sharing is caring” 2010 https://www.flickr.com/photos/niklaswikstrom/5214708665/
- Github. “Repositories images” 2013 https://pages.github.com/images/ghfm@2x.png
- Futureatlas.com. “Citation needed” 2010 https://www.flickr.com/photos/87913776@N00/5129607997/
- Avlxyz. “Authentic 5.25 inch floppy disks” 2014 http://www.flickr.com/photos/avlxyz/5766413625/
- Mattlogein. “Floppy disk” 2014. http://www.flickr.com/photos/49503210657@N01/127984947
- Normann. “Computer Model 1981” 2014 http://www.flickr.com/photos/26009408@N00/8600423939
- Weiler, Aron.” Encrypted file” 2014. http://www.flickr.com/photos/57523780@N00/52632856
- Mccapdevila. “Current clamp recording of neuron” 2012. http://en.wikipedia.org/wiki/File:Current_Clamp_recording_of_Neuron.GIF

With thanks to Kevin Read and Alisa Surkis, NYU Langone Health Sciences Library. These slides are adapted from their training program on Research Data Management