JOHNS HOPKINS
MEDICINE

Data Acquisition and Management

Secure and Ethical Data Use

Richard L. Skolasky, Sc.D.
Lewis Cass Spencer Professor and Vice Chair of Research
Department of Orthopaedic Surgery
Johns Hopkins University School of Medicine

Overview

- Data Acquisition and Management
  - *General – data lifecycle*
  - *Data SOPs for every research project*
- Secure and Ethical Data Use
- Data Sharing and the IRB

Objectives

- Communicate plans for data lifecycle and provenance from creation to scholarship, sharing, and retention
- Identify data security practices fundamental to our goals for rigorous research, equity, and privacy

Researchers' Responsibilities

- Be able to find the data and trace its history
  - *Data not retained*
  - *Challenge of changes in location, personnel*
- Organize the data
  - *Poorly indexed*
  - *Duplicated, poorly-annotated changes*
- Link data with results, tables, & figures
  - *No idea which data were used*

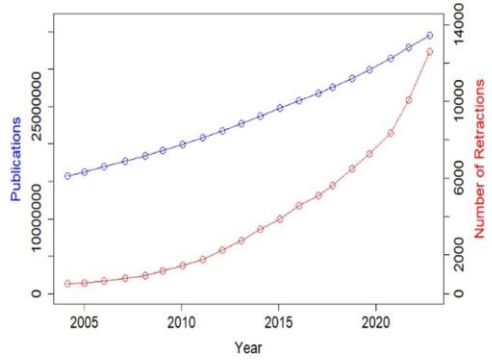
Retraction are on the rise!

Retractions of published results:

including (but not limited to) actual cases of scientific misconduct
cautionary tales from the real world

- https://ori.hhs.gov/case_summary - DHHS Office of Research Integrity - Case Summaries
- PubMed search for retractions - search term: *retraction of publication [pt]*

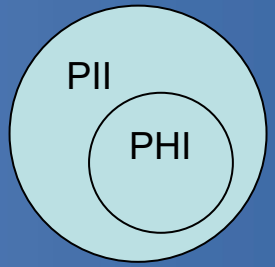
- 498 as of Feb 25, 2004
- 552 as of Jan 1, 2005
- 638 as of Jan 1, 2006
- 784 as of Feb 8, 2007
- 938 as of Feb 20, 2008
- 1170 as of Jan 14, 2009
- 1459 as of Jan 20, 2010
- 1787 as of Feb 2, 2011
- 2251 as of Feb 3, 2012
- 2757 as of Jan 23, 2013
- 3345 as of Jan 15, 2014
- 3859 as of Jan 21, 2015
- 4572 as of Jan 15, 2016
- 5107 as of Jan 15, 2017
- 5606 as of Sep 30, 2017
- 6513 as of Oct 2, 2018
- 7254 as of Sep 1, 2019
- 8347 as of Oct 1, 2020
- 10079 as of Sep 1, 2021
- 12583 as of Oct 1, 2022
- Browsing these is educational



Note the alarmingly **disproportionate** increase in retractions (in red) relative to total publications (blue)!

DATA HANDLING PRACTICES

Types of Data



PHI

- Any information about an individual's health, healthcare, or payment for healthcare that can be linked to their identity
  - *Name*
  - *Address*
  - *Dates (birth, admission, discharge)*

PII

- Any data that can be used to directly or indirectly identify an individual, and its sensitivity can vary, with some data requiring higher levels of protection
  - *Name*
  - *Address*
  - *Dates*
  - *Biometric data*
  - *Driver's license number*

Computers

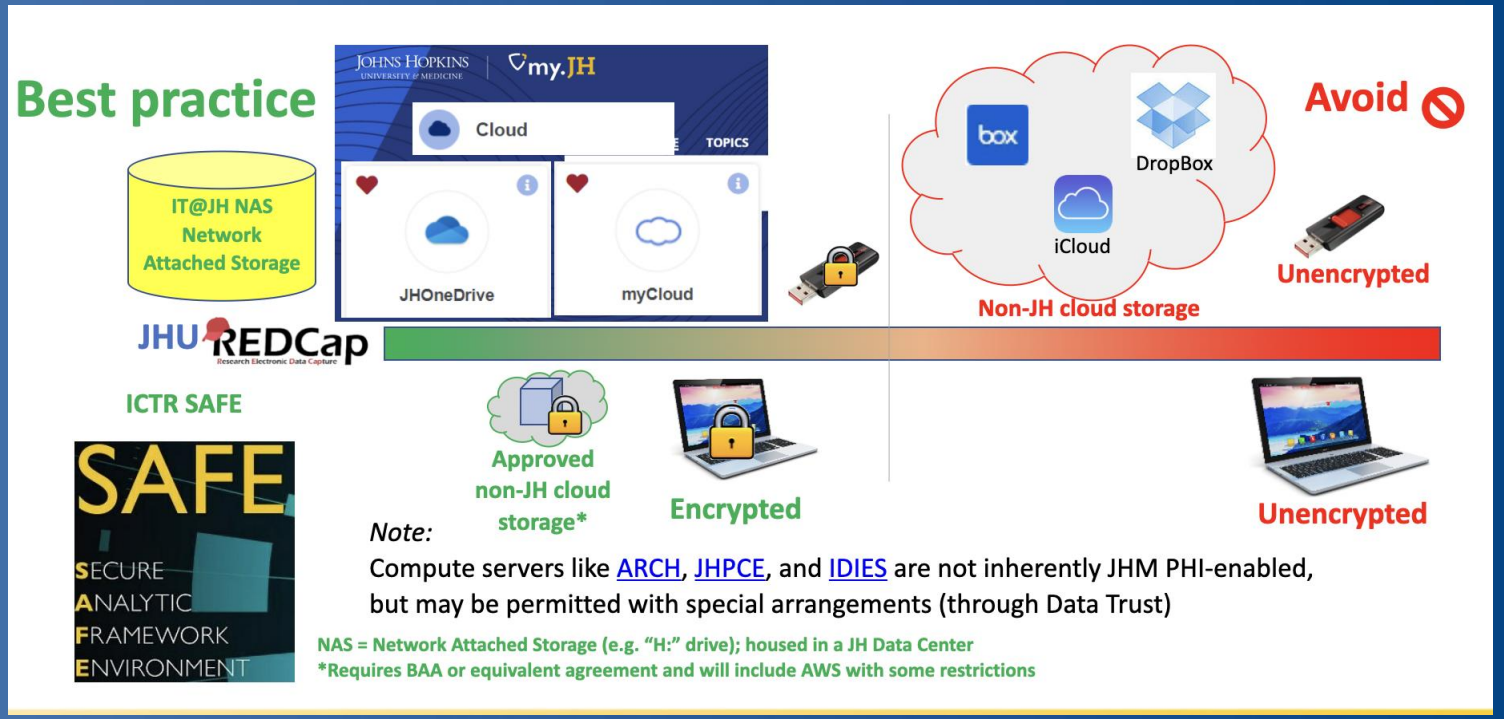
- Windows 10/11 and MacOS are supported
  - *Only JH-owned and managed computers for JH data handling*
  - *Don’t use your own for storage/analysis – you can access JH Cloud/SAFE desktop from any Chrome browser, even a home/personal computer*

Electronic messaging

- Use only Hopkins email for Hopkins work
  - You can create email alias at https://my.jh.edu/
- Hopkins email
  - No forwarding to other servers (e.g. Gmail)
  - Minimize PHI in body/attachments
  - Zero PHI in subject line
  - "[secure]" in Subject to help protect PHI/PII
  - Avoid long "To:" lists
    - Use bcc, listserv (lists.johnshopkins.edu), or distribution list
  - Recognize phishing and spear-phishing

- Manage all JH work data files on JH network folders
  - *JHOneDrive, JHTeams, JH SAFE "S:" drive, "H:" drive, JHU REDCap, or similar;*
- Minimize storing data on the hard drive of your JH-owned & managed computer/laptop
  - Avoid* storing JH data on:
    - USB drive
    - *DropBox, Box, Google Drive, iCloud, etc*

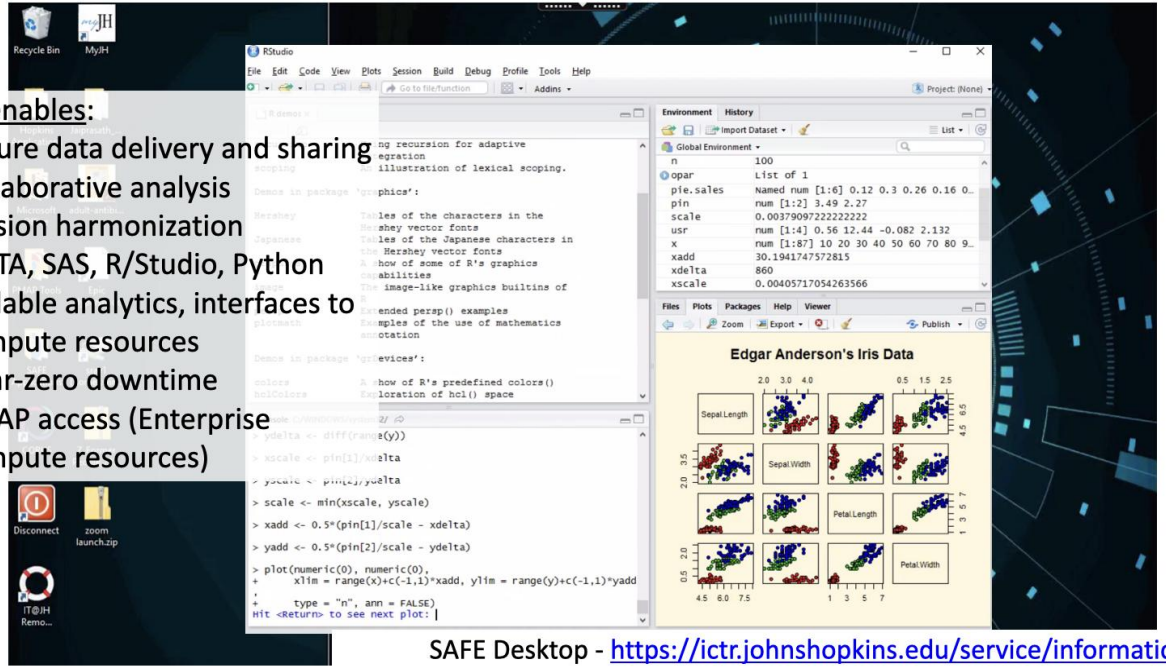
Data security landscape



JH SAFE Desktop

SAFE enables:

- Secure data delivery and sharing
- Collaborative analysis
- Version harmonization
- STATA, SAS, R/Studio, Python
- Scalable analytics, interfaces to compute resources
- Near-zero downtime
- PMAP access (Enterprise compute resources)



SAFE Desktop - https://ictr.johnshopkins.edu/service/informatics/safe-desktop/

Things to avoid

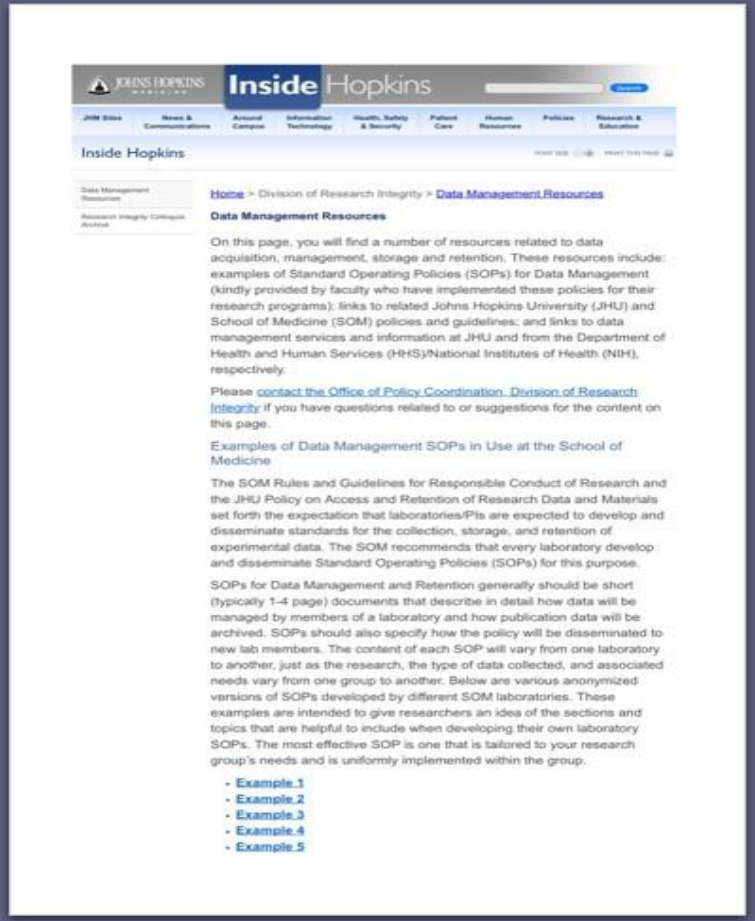
- Don’t: Image patients using native phone app
  - Do: Use Haiku (for record) or Corus (for sharing)
- Don’t: Store of work-related data using non-JH cloud services
  - No work use of DropBox, Box, Google Drive, iCloud
  - Allowed: JHOneDrive, JHTeams, JHSharepoint
- Don’t: Store data/presentations on unencrypted USB drives etc
  - Do: Use JHOneDrive and other best practice resources already listed

- **Never: Share JHED credentials with others**

SCHOLARSHIP

Potential for rigor

- Data driven research has great potential to be:
  - Replicable
  - *Reproducible*
  - Reusable



Data hygiene is foundational

- Data ecosystem is increasingly complex
- The more templated our systems are, the easier it is for faculty and trainees to do the right thing
- Helpful resources
  - https://intranet.insidehopkinsmedicine.org/division-of-research-integrity/data-management-resources/index.html

Slide courtesy of Antony Rosen

Research reproducibility recommendations

- Identify and share “best practices”
- Expand educational opportunities and accessibility
- Create an archive resource (archive@JHUSOM) that will serve as a repository for primary data that supports all publications

Research reproducibility recommendations

- Assess data storage capacities and capabilities through departmental reviews and provide assistance in design and implementation of practical, adaptive and cost-effective solutions

- Explore approaches for providing JHUSOM investigators with current information regarding the utility of experimental reagents

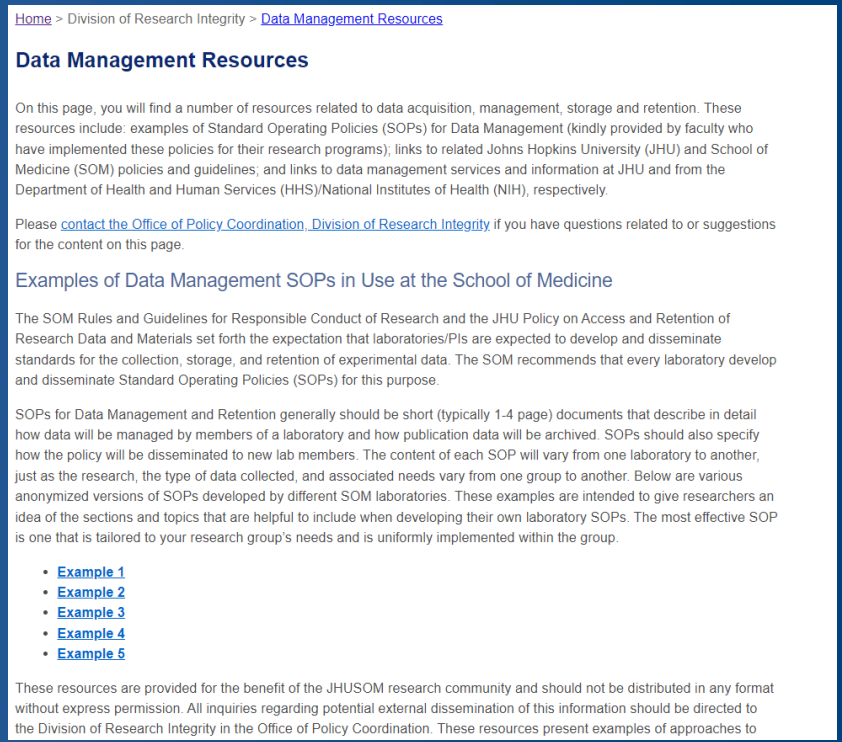
Expectations

- Data provenance
  - History of the data from creation to present, who/when/what changed
- Research notebooks and data that can be:
  - Linked to published works
  - Annotated by all involved parties
- Data storage SOPs

Data management SOPs

- Making it easier

https://intranet.insidehopkinsmedicine.org/division-of-research-integrity/data-management-resources/index.html



Home > Division of Research Integrity > Data Management Resources

Data Management Resources

On this page, you will find a number of resources related to data acquisition, management, storage and retention. These resources include: examples of Standard Operating Policies (SOPs) for Data Management (kindly provided by faculty who have implemented these policies for their research programs); links to related Johns Hopkins University (JHU) and School of Medicine (SOM) policies and guidelines; and links to data management services and information at JHU and from the Department of Health and Human Services (HHS)/National Institutes of Health (NIH), respectively.

Please contact the Office of Policy Coordination, Division of Research Integrity if you have questions related to or suggestions for the content on this page.

Examples of Data Management SOPs in Use at the School of Medicine

The SOM Rules and Guidelines for Responsible Conduct of Research and the JHU Policy on Access and Retention of Research Data and Materials set forth the expectation that laboratories/PIs are expected to develop and disseminate standards for the collection, storage, and retention of experimental data. The SOM recommends that every laboratory develop and disseminate Standard Operating Policies (SOPs) for this purpose.

SOPs for Data Management and Retention generally should be short (typically 1-4 page) documents that describe in detail how data will be managed by members of a laboratory and how publication data will be archived. SOPs should also specify how the policy will be disseminated to new lab members. The content of each SOP will vary from one laboratory to another, just as the research, the type of data collected, and associated needs vary from one group to another. Below are various anonymized versions of SOPs developed by different SOM laboratories. These examples are intended to give researchers an idea of the sections and topics that are helpful to include when developing their own laboratory SOPs. The most effective SOP is one that is tailored to your research group's needs and is uniformly implemented within the group.

- Example 1
- Example 2
- Example 3
- Example 4
- Example 5

These resources are provided for the benefit of the JHUSOM research community and should not be distributed in any format without express permission. All inquiries regarding potential external dissemination of this information should be directed to the Division of Research Integrity in the Office of Policy Coordination. These resources present examples of approaches to

Data management SOPs

1. Determine the Research Sponsor Requirements
2. Identify the Data to Be Collected
3. Define How the Data Will Be Organized
4. Explain How the Data Will Be Documented
5. Describe How Data Quality Will Be Assured
6. Present a Sound Data Storage and Preservation Strategy
7. Define the Project's Data Policies
8. Describe How the Data Will Be Disseminated
9. Assign Roles and Responsibilities
10. Prepare a Realistic Budget

Continuity of data

- When faculty/staff/trainees leave Johns Hopkins, these things must remain at JHU (unless arrangements are made in advance):
  - research records
  - (if applicable) clinical data registries
  - computers/other equipment using JHU funds

Continuity of data

- After faculty/staff/trainees leave Johns Hopkins, how will data be accessed to complete unfinished work?
- For IRB-governed research, changes in employment/enrollment status must be reflected in IRB Change in Research
- Enabling view of data on a JH server (e.g. JH OneDrive) with someone outside JH is still sharing – even if the data doesn't "physically" move.

Awesome responsibility

- Authorship
  - Recognize, appropriately, contributions to science
- Data stewardship
  - Protect our research participants' rights
  - Shepherd the data life cycle from collection and analysis to retention and sharing
  - Communicate the plan so that it is a shared understanding

Awesome responsibility

- Ethical use
  - "Data is the new oil" – a mantra in our world of big data
    - Implications of this fraught metaphor
    - Risks of repeating Henrietta Lacks' experience if we're not intentional about consent and equity
  - Sharing data to the extent allowable

More information

- Expanding menu of resources & navigation at https://researchit.jh.edu
- Reach out to datatrust@jhmi.edu with questions, pre-review of submission documents, etc.

More Information

Expanding menu of resources & navigation at https://researchit.jh.edu

Reach out to datatrust@jhmi.edu or Stuart Ray sray@jhmi.edu with questions, pre-review of submission documents, etc.

For additional information about the Responsible Conduct of Research (RCR) program or any other questions, please contact the Office of Research Integrity at integrity@jhmi.edu or phone at (667) 208-8948.

For more information about training requirements for JHU faculty, trainees and staff engaged in research at the School