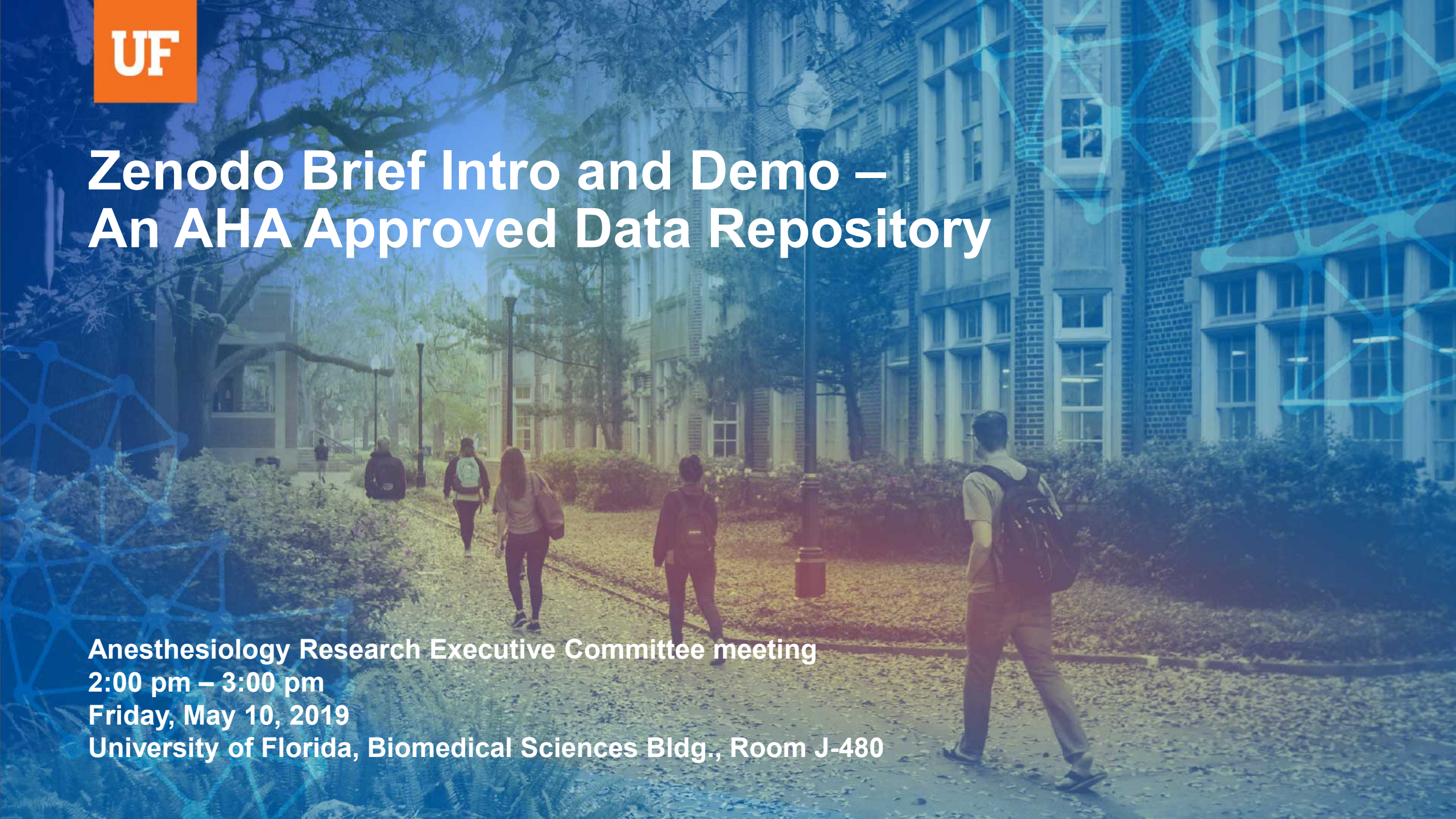

Zenodo Brief Intro and Demo – An AHA Approved Data Repository

Anesthesiology Research Executive Committee meeting
2:00 pm – 3:00 pm
Friday, May 10, 2019
University of Florida, Biomedical Sciences Bldg., Room J-480

Table of Contents

American Heart Association (AHA) Open Data

"The AHA requires certain applicants to include a data sharing plan with the application. Any factual data that is needed for independent verification of research results must be made freely and publicly available in an AHA-approved repository within 12 months of the end of the funding period (and any no-cost extension)." – American Heart Association, Open Science Policy Statements for AHA Funded Research (https://bit.ly/2WxY1F7)

AHA Compliance Threshold*

Applicants will be prompted to answer each of the following questions when completing a data plan in the application:

1. What data outputs will the research generate?
2. When will the data be shared?
3. Where will the data be made available?
4. Are any limits to data sharing required?

Data Organization and Data Management Principles

- ❑ Define folder and file names and structure - and use them!
- ❑ Limit the depth of sub-folders to no more than 2
- ❑ Use meaningful names that include basic information (e.g. date, measurement, collection, etc.)
- ❑ Make the name unique
- ❑ Avoid Spaces
- ❑ Use ASCII Characters only
- ❑ Document, share, and evaluate
- ❑ Separate classes of products: raw data, derived data, graphics, code, documents, etc.
- ❑ Make Backups!

1. Define the contents of your data files
2. Define the variables
3. Use consistent data organization
4. Use stable file formats
5. Assign descriptive file names
6. Preserve processing information
7. Perform basic quality assurance
8. Provide documentation
9. Protect your data
10. Preserve your data
11. **De-identify your data (next two slides)**

Data Organization and Data Management Principles

18 HIPAA Identifiers	
Name	Account numbers
Geographic subdivisions smaller than a state (e.g. street address, city and ZIP code)	Certificate/license numbers
All dates that are related to an individual (e.g., date of birth, admission)	Vehicle identifiers and serial numbers, including license plate numbers
Telephone numbers	Device identifiers and serial numbers
Fax Numbers	Web universal locators (URLs)
Email addresses	IP address numbers
Social security numbers	Biometric identifiers such as fingerprints and voice prints
Medical record numbers	Full-face photographic images
Health plan beneficiary numbers	Other unique identifying numbers, characteristics or codes

Data Organization and Data Management Principles

De-identify Patient Data Defined

- "De-identified patient data is health information from a medical record that has been stripped of all 'direct identifiers – that is all information that can be used to identify the patient from whose medical record the health information was derived." – Laffel (2010)

There are 3 acceptable ways, according to HIPAA, to de-identify patient data.

1. "**Safe harbor**" option – all 18 identifiers are removed
2. "**Statistical**" option – retained statistician determines which of the 18 identifiers can be maintained without creating greater than "very small" risk of re-identification.
3. "**Limited data set**" option – organization removes 16 identifiers and protects what remains with special security precautions (Laffel, 2010)

AHA Acceptable Repositories

Discipline-specific data repositories

- ArrayExpress – The ArrayExpress Archive is a database of functional genomics experiments including gene expression where you can query and download data collected to MIAME and MINSEQE standards.
- BioModels – BioModels Database is a repository of computational models of biological processes
- CellML – The purpose of CellML is to store and exchange computer-based mathematical models.

General data repositories

1. Dataverse – The Dataverse Network is an open source application to publish, share, reference, extract and analyze research data.
2. Figshare – figshare allows users to upload any file format to be made visualisable in the browser so that figures, datasets…
3. Zenodo – ZENODO builds and operate a simple and innovative service that enables researchers, scientists, EU projects and institutions to share and showcase multidisciplinary research results (data and publications) that are not part of the existing institutional or subject-based repositories of the research communities.

Zenodo Brief Intro and Demo

Login to Zenodo

- Log in with GitHub
- Log in with ORCID (recommended)
  - Use Institutional account
- Log in with Zenodo
  - Sign up for a Zenodo account - https://zenodo.org/signup/

Develop community, deposit data sets, and FAQ

1. Show completed communities (e.g. Continental Journal of Biomedical Sciences (CJBM), INSC 590 spring 2019)
2. Demonstrate Zenodo via login to Zenodo sandbox for live in class demo https://sandbox.zenodo.org/login/?next=%2F
3. Describe the process to develop a community
4. Show how to develop a community
5. Demonstrate some features of Zenodo
6. Provide Zenodo FAQ - http://help.zenodo.org/

References

1. American Heart Association. (2019). Open Science Policy Statements for AHA Funded Research. Accessed online May 9, 2019 from https://bit.ly/2WxY1F7.
2. AHA Approved Data Repositories. (2019). Accessed online May 9, 2019 from https://bit.ly/2vOwRhG.
3. Laffel, G. (April 30, 2010). Using de-identified health information to improve care: What, how and why. Accessed online May 9, 2019 from https://bit.ly/2LwmDgz.
4. U.S. Department of Health & Human Services. (2005). Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule. Accessed online May 9, 2019 from https://bit.ly/2Dxkvfo.
5. Zenodo sandbox. (2019). Zenodo sandbox. Accessed online May 9, 2019 from https://sandbox.zenodo.org/.
6. Zenodo. (2019). https://zenodo.org/

Thank you

Questions/comments

Contact:
Plato Smith, Data Management Librarian, plato.smith@ufl.edu
Data Management and Curation Working Group
DATAMGMT-L@LISTS.UFL.EDU