Johns Hopkins University Data Management Services

Dave Fearon, Betsy Gunia, and Tim DiLauro

datamanagement@jhu.edu

http://dmp.data.jhu.edu/

JHU Data Management Services

How we started:

- NSF's Data Management Planning requirement (2011)

  - NSF exploring broader requirements, closer reviewer evaluation

  - Other funders adopting data sharing and data management planning requirements

- JHU's (and Deans') interest in data management:

  - Competitive grants

  - Leveraging data resources for sharing, preservation & re-use

  - Compliance with JHU Data Retention Policy

DataConservancy

JOHNS HOPKINS UNIVERSITY   THE SHERIDAN LIBRARIES

JOHNS HOPKINS UNIVERSITY POLICY ON ACCESS AND RETENTION OF RESEARCH DATA AND MATERIALS
January 2, 2008

2. APPLICABILITY OF POLICY: This Policy on Access and Retention of Research Data and Materials shall apply to all Johns Hopkins University faculty, staff, postdoctoral fellows, students and any other personnel in the design, conduct or reporting of research the University.

3. OWNERSHIP OF RESEARCH DATA: The University generated by research projects conducted at or Hopkins University regardless of funding sources sponsorship, other agreements or University policy
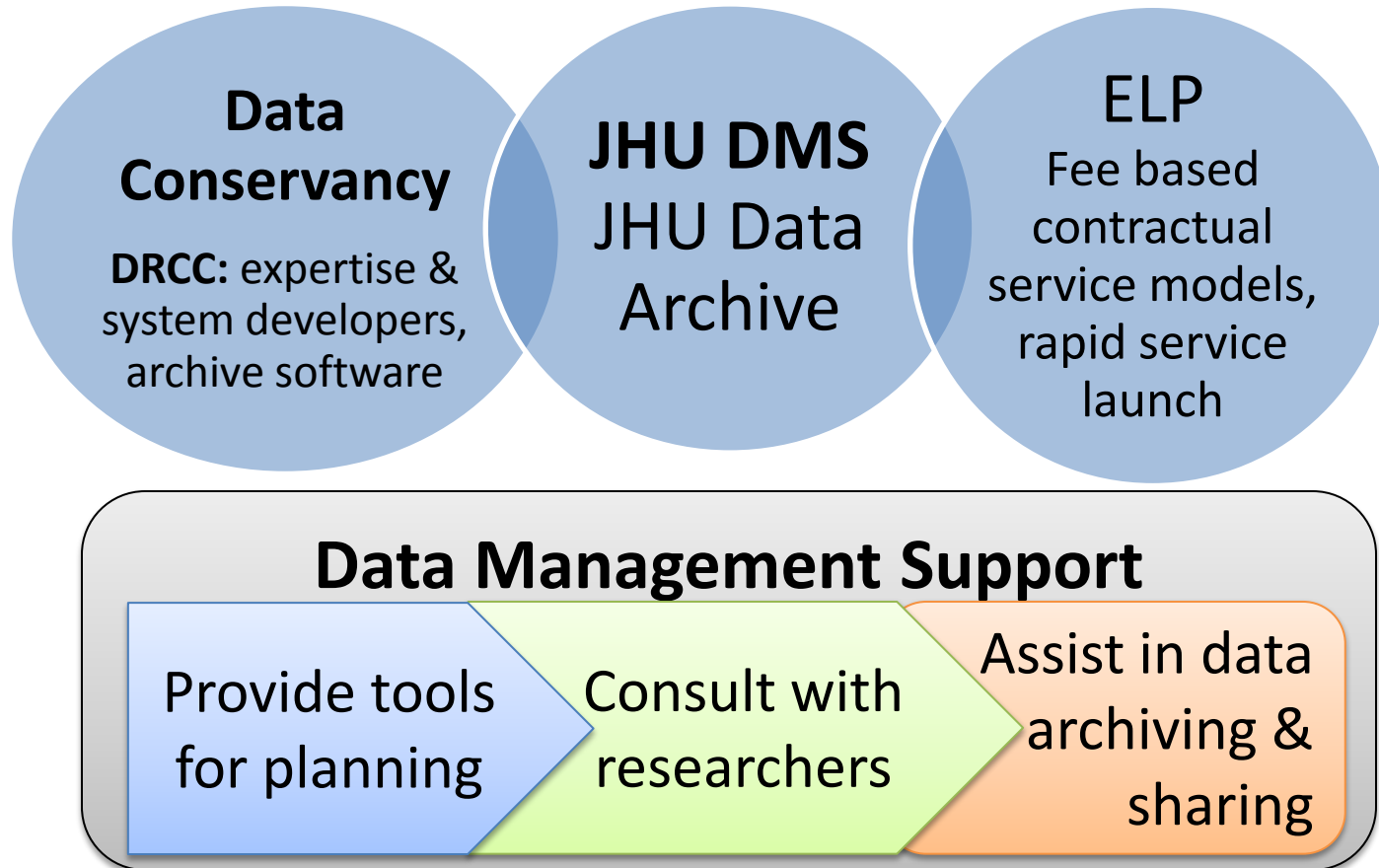
This policy does not attempt to determine relations surrounding collaborative efforts such as authorship.

4. RETENTION AND ARCHIVING: The Primary research project is responsible for selecting and archiving Research Data, and for determining detail and for an adequate period of time about accuracy, authenticity, privacy, and confidentiality governing the conduct of research. The Primary responsible for educating all participants in the study regarding Research Data, and for protection of the University's rights and ability to meet obligations related to the Research Data. The Primary Responsible Investigator should also consult with University officials regarding the development of any contingency plans.

> Research Data should be stored using a method that permits a complete retrospective audit if necessary. Unless ethical/professional/local or funding body guidance requires otherwise, Research Data should be archived in a durable form and in a secure location that is immune to subsequent tampering and falsification for **a minimum period of 5 years** after the date of any publication upon which it is based.

DataConservancy

JOHNS HOPKINS UNIVERSITY   THE SHERIDAN LIBRARIES

Building upon data management, archiving, & service expertise

Data Conservancy

DRCC: expertise & system developers, archive software

JHU DMS JHU Data Archive

ELP
Fee based contractual service models, rapid service launch

Data Management Support

Provide tools for planning

Consult with researchers

Assist in data archiving & sharing

DataConservancy

JOHNS HOPKINS
U N I V E R S I T Y

THE SHERIDAN
LIBRARIES

What we do

Assistance with Preparing Data Management Plans and Planning

Data Management Training Workshops

Free: funded by schools

Archiving Research Data into JHU Data Archive

2% of total direct cost of grant

DataConservancy

JOHNS HOPKINS UNIVERSITY

THE SHERIDAN LIBRARIES

Consulting on data management plans

Final Plans

Contact from Researcher

JHU DMS Comments on Plan

Background Research

Info on PI

Standards

Data Repositories

Researcher Drafts Data Management Plan

In-Person Consultation

Questionnaire

DataConservancy

JOHNS HOPKINS UNIVERSITY   THE SHERIDAN LIBRARIES
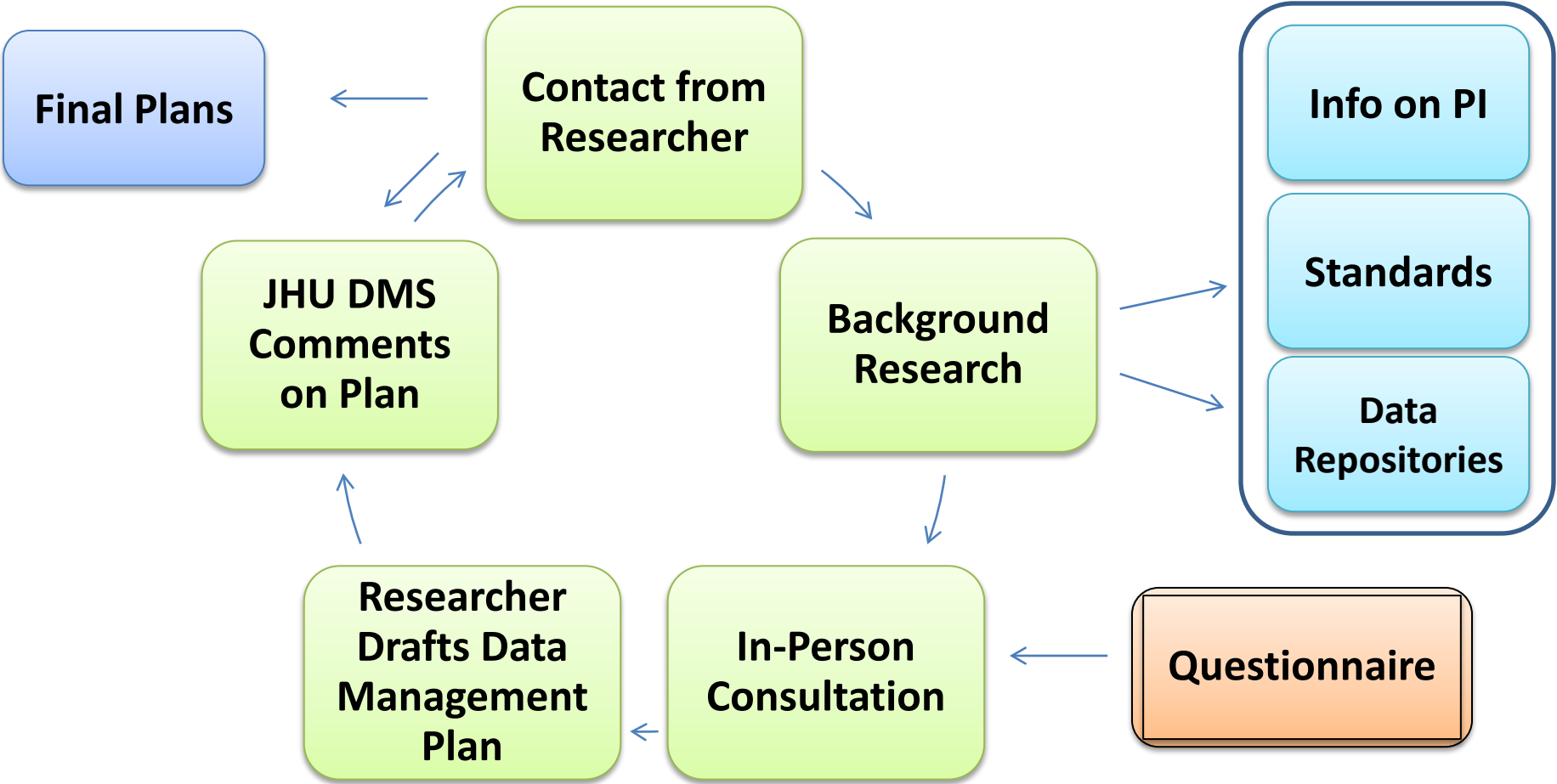
Questionnaire for DMP Consultation:
(sample questions)

Section I: Data Products and Standards
What naming conventions/schema will be used for your data, if any?
Section II: Data Storing and Long-Term Preservation
During the project, how (i.e. media) and where (i.e. location(s)) will data be stored and who is responsible for it?
Section III: Data Sharing
Are there any data with privacy concerns to sharing (e.g., human subjects)? If so, what policies need to be adhered to and how will policies be enforced?

Guide & Worksheet for Grant Proposal Reviewers

	Research product	Source	Format	Size	Preserved (how?)	Shared (how?)
	Tables, images, computer code, curriculum items, physical samples	Data repository, Instrument, interviews	JPG, MATLAB, Excel table	>1TB, 20K files	Discarded, PI retains Archiving service	By request, Website, Archive/Repository
1						
2						
3						
4						
5						

Get it at bit.ly/DMPworksheet

Data management during project:
- ☐ Storage: has a backup plan
- ☐ Location & media used:
- ☐ ☆ 2+ copies with 1 off-site
- ☐ ☆ Specifies who is responsible
- ☐ ☆ Data security / access controls
- ☐ ☆ Has conventions for naming & organizing files
- ☐ ☆ Version control
- ☐ ☆ Collaboration coordination

Data retention after the project:
- ☐ Where is data preserved?
- ☐ How long?
- ☐ Who administers?
- ☐ ☆ Gives reasons for preserving data (especially raw data)
- ☐ ☆ Using an archive service or repository?

Data Sharing
- ☐ Is data publically accessible?
- ☐ When will data be shared?
- ☐ Who administers?
- ☐ ☆ Describes audience to benefit.

Preparation of shared data
- ☐ Uses their research field's metadata standards
- ☐ AND/OR creates description sufficient for re-use
- ☐ Metadata or supplementary files explaining: content/ file structure/ procedures/ codebook or variable-level detail
- ☐ ☆ Metadata associated with digital files

Data sharing policy
- ☐ Gives conditions for re-use
- ☐ Accounts for:
- ☐ ☆ privacy (personal identifiers)/security issues
- ☐ ☆ intellectual property (copyrights, patents)
- ☐ ☆ delays for sharing (e.g., embargos)

Services of archive (if specified for preservation and/or sharing data)

Archive type	Preservation activities	Data sharing services
☐ PI's Institutional repository (documents)	☐ Data integrity checking	☐ Public access to data files
☐ ☆ Data repository	☐ ☆ Migration to new formats, media	☐ ☆ Persistent data citation

- ☐ If a plan states there is no data to manage or share, have they justified it?

DataConservancy

JOHNS HOPKINS
UNIVERSITY

THE SHERIDAN
LIBRARIES

Data Management Training Workshops

Current Topic

- Creation of Data Management Plans
- Dec 5th and Jan 8th, 12-1 sign-up at datamanagement@jhu.edu

Future Topics

- Backup & security
- File organization & documentation
- Sharing & finding data

DataConservancy

JOHNS HOPKINS UNIVERSITY   THE SHERIDAN LIBRARIES

What we do

Assistance with Preparing Data Management Plans and Planning

Data Management Training Workshops

Archiving Research Data into JHU Data Archive

2% of total direct cost of grant

DataConservancy

JOHNS HOPKINS UNIVERSITY

THE SHERIDAN LIBRARIES

What does it mean to preserve digital objects?

Long Term Access – data re-use, active management

JHU Data Archive

ICPSR

Cambridge Structural Database

Archiving – unique durable ID, fixity check

CUAHSI

DATAVERSE

BIRN

Storage – Researcher maintains

Lab/Dept. Server

Cloud storage

CDs/DVDs

JHU Data Archive

- Will be an instance of Data Conservancy Service (DCS) software

- Infrastructure configured for research data

- Will accept data from all disciplines and formats

- Currently, in testing phase (not publicly-accessible)

- The DCS software is available at http://dataconservancy.org/

Data Model Overview

Data Model Overview

	Projects	E.g., NSF #123
Collection 1 Metadata File	**Collection 1**	E.g., a publication
Data Item 1	**Data Item 2**	E.g., description of scanning electron microscope images
Data File(s) 1	**Data File(s) 2**	E.g, scanning electron microscope images
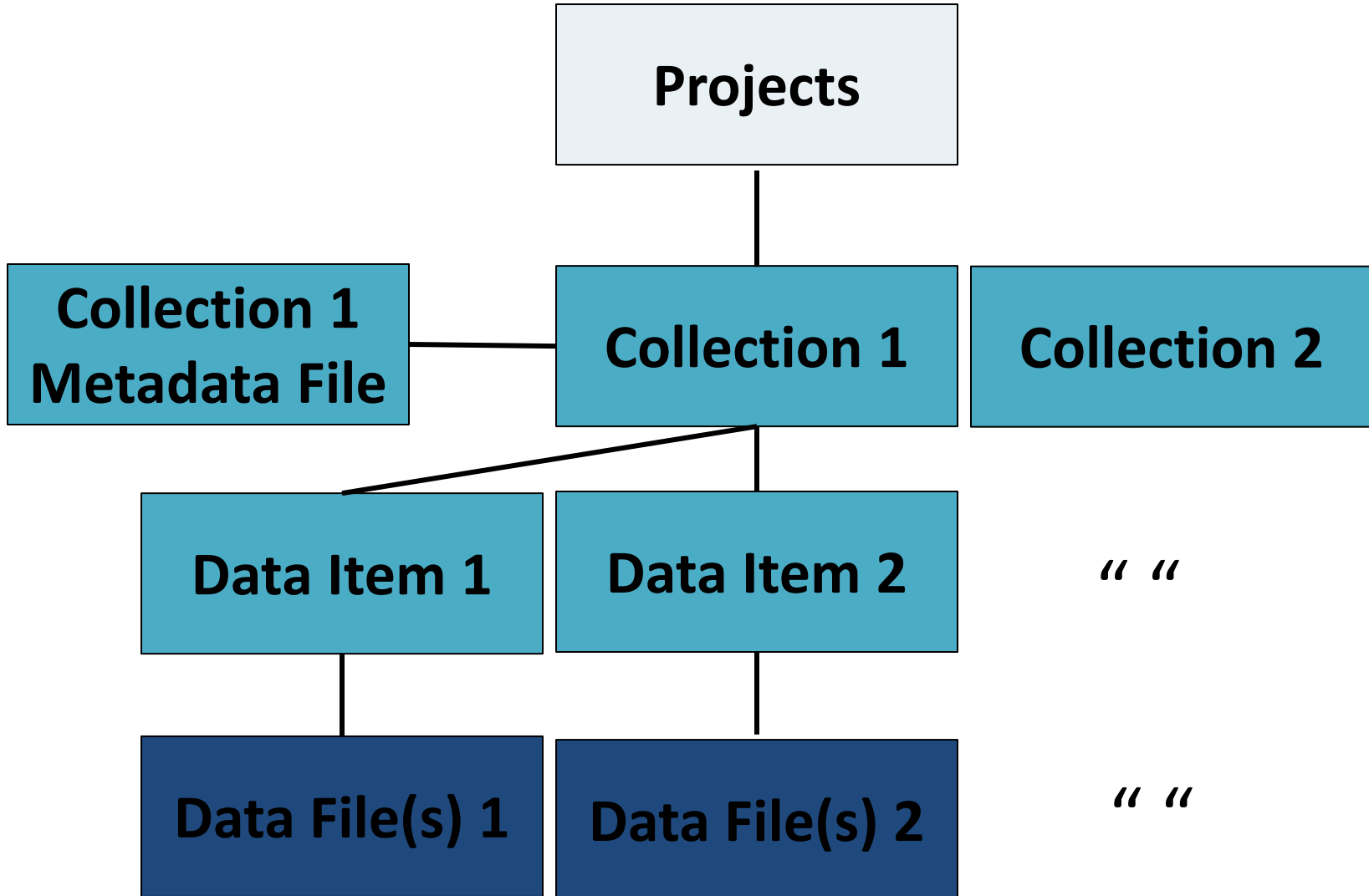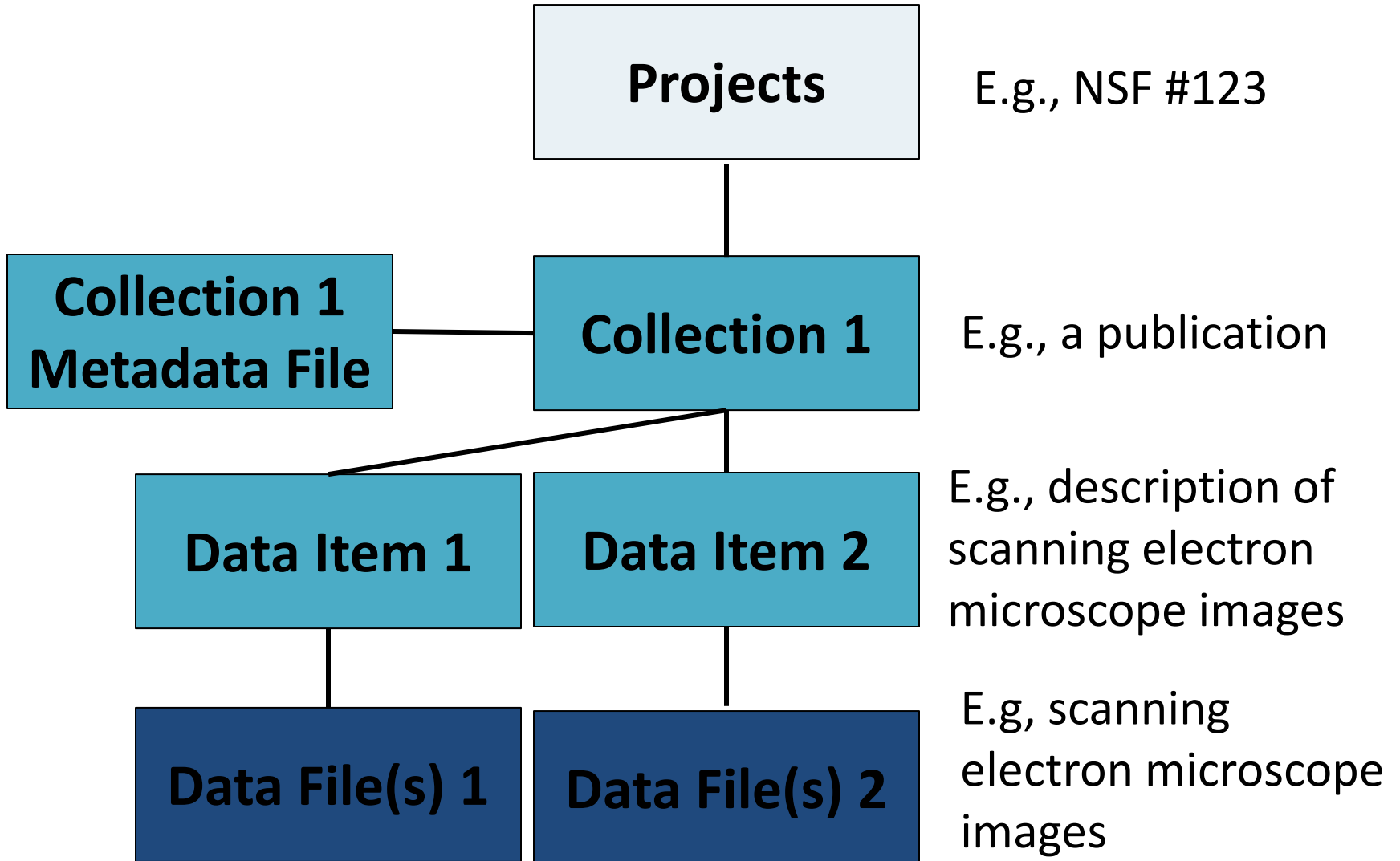
JHU Data Archive Development Timeline

Fall 2012

Pilot release of software

Deployment of test instance of JHU Data Archive

Winter 2013

Deployment of JHU Data Archive

Archiving of data from three different researchers

Spring 2013

XML metadata schema validation

Indexing of project and collection metadata

DataConservancy

JOHNS HOPKINS
UNIVERSITY

THE SHERIDAN
LIBRARIES

JHU Data Archive (winter of 2013)

Current Features

- Create projects, collections within a project, and data items within a collection
- Ability to archive and retrieve data files
- Assign DOIs for easy data citation

DataConservancy

JOHNS HOPKINS UNIVERSITY   THE SHERIDAN LIBRARIES

Deposit interface of test instance

Viewing Differentiated Smooth Muscle Cell project

Name:	Differentiated Smooth Muscle Cell
Description:	This project includes three published studies conducted on the use of differentiated smooth muscle cells in artificial vascularization
Funding entity:	March of Dimes
Award No:	123
Start date:	9/1/07
End date:	5/1/12
Storage allocated:	2.00 TB
Storage used:	5.00 GB
Project Administrators:	admin admin

Edit project

View project's activities log

View collections in this project | Add Collection

DataConservancy

JOHNS HOPKINS
UNIVERSITY

THE SHERIDAN
LIBRARIES

Title:	Collagen I Stained Cell
Citable Locator:	**Get citable locator**
Associated Project:	**Differentiated Smooth Muscle Cell**
Publication Date:	
Contacts:	
Citation:	*ESIP*: Hanjaya-Putra, D. (2012). *Collagen I Stained Cell. Version [Version].* Johns Hopkins Data Management Service. [locator]. Accessed 9 Nov 2012.
Description:	These data were published in Vo et. al. 2010. Smooth-Muscle-Like Cells Derived from Human Embryonic Stem Cells Support and Augment Cord-Like Structures. Stem Cell Reviews and Reports Volume 6, Number 2 (2010), 237-247, DOI: 10.1007/s12015-010-9144-3.
Alternate ID:	
Creators:	Donny Hanjaya-Putra
Deposit status:	Deposited

Collection's Metadata Files

Add Metadata File

Content

Deposit data

View Data For Collection

[Done]

Data deposit page of test instance

Data item page of test instance

Developing Capabilities For JHU Data Archive

Short-term
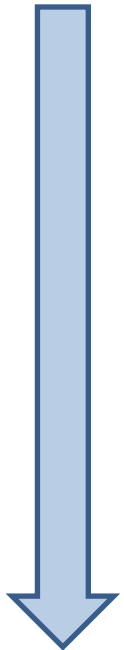
Long-term

- Creation of sub-collections
- Multiple files per data item
- Search interface
- Ability to embargo
- Ability to search for same properties (lat/long) across disparate data sets
- Support long-term access (preservation)

DataConservancy

JOHNS HOPKINS
UNIVERSITY

THE SHERIDAN
LIBRARIES

JHU DMS Current Archiving Projects

Political Science

- Transcription in original language and English translations
- De-identified documents
- Data will be transferred to us this winter
- ~1 megabyte

Chemical and Biomolecular Engineering

- Data from six publications
- Images taken by different types of instruments
- Raw PCR data
- Data transferred to us
- ~35 Gigabytes

DataConservancy

JOHNS HOPKINS
UNIVERSITY

THE SHERIDAN
LIBRARIES

Process for Work with Students
(Data already published)

1. Create "data flow diagram", a diagram of what data products have been generated during experiment

2. Meet with researchers to transfer and understand files

3. Create table detailing the content of each folder and sub-folder

4. Create new folders and rename folders and files if necessary.

5. This winter: deposit data into test instance first before deposit into JHU Data Archive
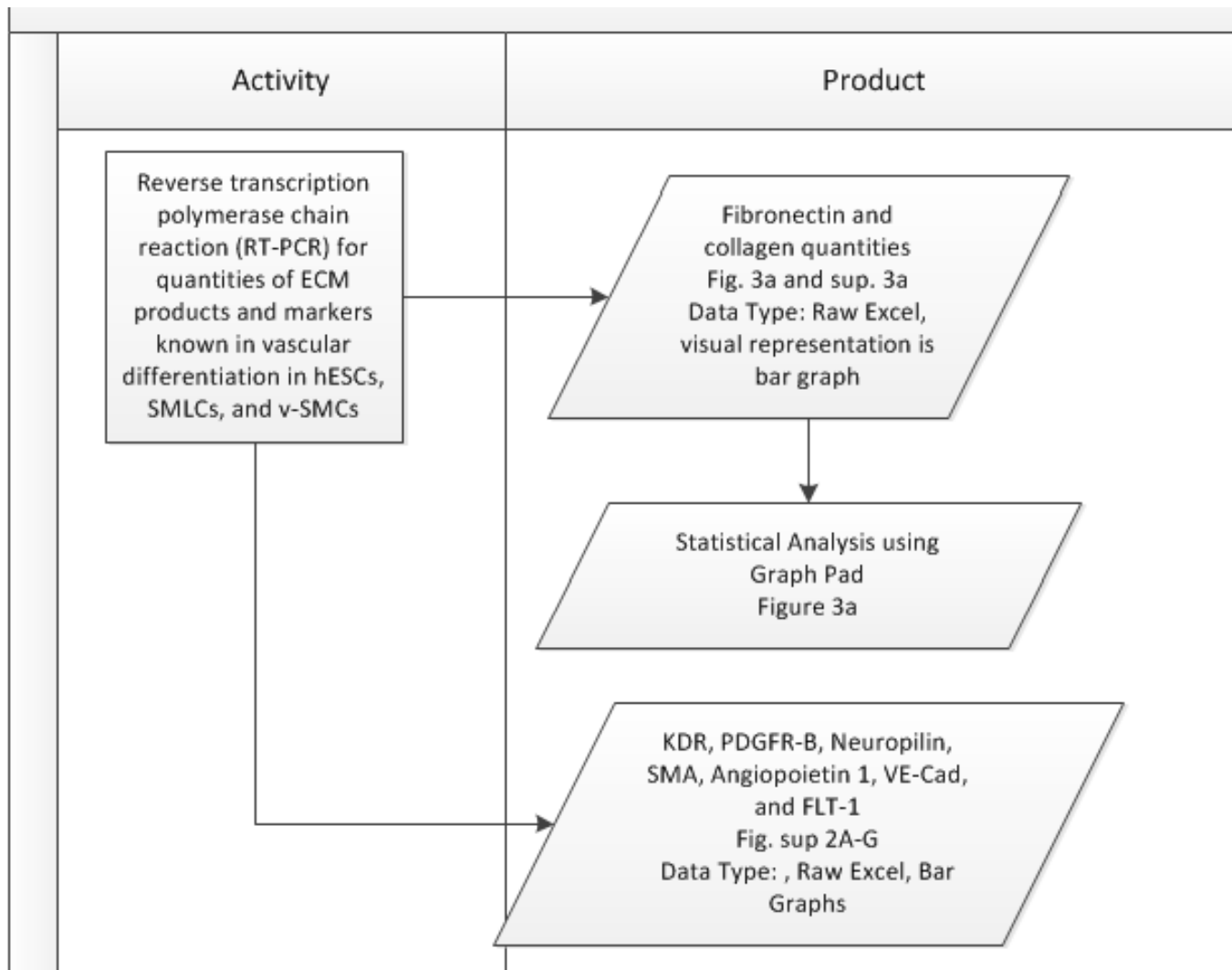
Data Flow Diagram

Process for Work with Students
(Data already published)

1. Create "data flow diagram", a diagram of what data products have been generated during experiment

2. Meet with researchers to transfer and understand files

3. Create table detailing the content of each folder and sub-folder

4. Create new folders and rename folders and files if necessary.

5. This winter: deposit data into test instance first before deposit into JHU Data Archive

DataConservancy

JOHNS HOPKINS
UNIVERSITY

THE SHERIDAN
LIBRARIES

📁 3 FACS	File folder			
📁 4 PCR data	File folder			
📁 5 SMC stretch	File folder			
📁 6 121108 MG livestaining	File folder			
📁 6 Matrigel 12 hr 220109	File folder			
📁 6 matrigel 24 hr 020609	File folder			
📁 6 Matrigel live stain 12 hr 4_29_09	File folder			
📁 6 Tube formation 090827 Vivi's tim...	File folder			
📁 7 quantitative data	File folder			
📁 8 06_25_09	File folder			
📁 8 090513 LSM510	File folder			
📁 8 Confocor3_PKH	File folder			
📁 8 EPC-SMC	File folder			
📁 8 LSM	File folder			
📁 10 EPC SMC Vivi	File folder			
📁 ATVB Figures in EPS	File folder			
📁 Vivi	File folder			
📊 lab_meeting_06_30_09	Microsoft Office PowerPo...	3,459 KB	No	
📊 matrigel 06_25_09	Microsoft Office Access D...	51 KB	No	
📊 Sheet1	Microsoft Office Excel Wo...	6 KB	No	
📄 Vo et al ATVB 11-18	Microsoft Office Word Do...	107 KB	No	

DataConservancy

JOHNS HOPKINS UNIVERSITY

THE SHERIDAN LIBRARIES

Process for Work with Students
(Data already published)

1. Create "data flow diagram", a diagram of what data products have been generated during experiment

2. Meet with researchers to transfer and understand files

3. Create table detailing the content of each folder and sub-folder

4. Create new folders and rename folders and files if necessary.

5. This winter: deposit data into test instance first before deposit into JHU Data Archive

Folder Name	Sub-Folder Name	Sub-Sub Folder	File	Notes	Questions for Donny
2 SMC Markers 2	0801 Nanog and MHC				What is na? ES = embryonic stem cells = hESCS?
2 SMC Markers 2	0806 MHC overnight				
2 SMC Markers 2	1201 SMA PE				
2 SMC Markers 2				Date_Cell Type_Field_what was stained 03/26/08	What is difference between Efield, Afield, etc. ?

Folder Name	Sub-Folder Name	Sub-Sub Folder	File	Notes	Questions for Donny
2 test stains 02_13-09					Does this warrant archiving? **DELETED**

Folder Name	Sub-Folder Name	Sub-Sub Folder	File	Notes	Questions for Donny
3 FACS				Folder naming convention is: yymmdd some yyddmm Cyflogic software v.1.2.1	Are the files in the same folder just different runs? These should be for different markers but it is not possible to see which markers with current

Process for Work with Students
(Data already published)

1. Create "data flow diagram", a diagram of what data products have been generated during experiment

2. Meet with researchers to transfer and understand files

3. Create table detailing the content of each folder and sub-folder

4. Create new folders and rename folders and files if necessary.

5. This winter: deposit data into test instance first before deposit into JHU Data Archive

Questions Asked While Gathering Students' Data

What is the best way to organize the data? By publication? By method or instrument?

Is there sufficient descriptive metadata? Is the data understandable to someone outside his/her lab? His/her discipline?

How much renaming and reorganization of files and folders and documentation of naming conventions should we undertake?

Should we deposit ALL the data that the researcher has given to us?

Contact JHU Data Management Services:

- If you or your team are preparing a data management plan

- If you are considering data sharing and archiving for your new grant project

- For advice on data management planning and preservation

datamanagement@jhu.edu

http://dmp.data.jhu.edu/