Towards FAIR data: Findable, Accessible, Interoperable, and Reusable

“Good data management is not a goal in itself, but rather is the key conduit leading to knowledge discovery and innovation, and to subsequent data and knowledge integration and reuse by the community after the data publication process.”


Mercè Crosas, Ph.D.  @mercecrosas
Chief Data Science and Technology Officer
Institute for Quantitative Social Science
Harvard University
“Good data management is not a goal in itself”

“Good data management is not a goal in itself”

- Enables continuity of research projects
- Facilitates data sharing and re-use
- Reduces research and data storage costs
- Helps with data reproducibility
Connecting Computing Resources and Data Management is critical

HMS Data Management Working Group
(established in 2014, grassroots)
Countway Library
HMS IT
HMS Basic Sciences
HMS Sponsored Programs Admin
Harvard Chan Bioinformatics
HMS Research Computing
HMS Academic and Research Integrity

Research Computing Council
(established in 2016 by CIO; Margulies, Cuff)
HUIT
FAS Research Computing
HMS Research Computing
HBS Research Computing
IQSS
HU Library

Harvard Data Group
(established in 2016 by Office of Vice-Provost of Research; Tahmassian and Crosas)
Office of Vice-Provost of Research
IQSS
HUIT
HMS IT
HU Library
Countway Library
HU Office of Sponsored Programs
HU Office of Sponsored Programs
HMS Basic Sciences
HMS Sponsored Programs Admin
Harvard Data Group has concrete Tasks

• Build a website for research data management @Harvard, coordinating with all existing resources (Spring-Summer 2017)
• Create a research data management training module, with custom modules for various research domains (2017-2018)
• Data User Agreements sub-group to coordinate DUA tracking and workflows, as part of data management support (2017)
• More in the future
A Single Entry to Data Management Avoids Confusion

- For researchers, not for librarians, archivists, or trainers
- Cite scholarly work, evidence-based studies
- Concise; point to other resources as needed
- “good enough data management”:
  - what you need to know
  - what Harvard can offer
  - other resources you can use
RDM @Harvard will link to HMS Data Management and Library sites

Harvard Biomedical Data Management
Best practices & support services for research data lifecycles

Data Management
Data Management is the process of providing the appropriate labeling, storage, and access for data at all stages of a research project. We recognize that best practices for each of these aspects of data management can and often do change over time, and are different for different stages in the data lifecycle.

Early and attentive management at each step of the data lifecycle will ensure the discoverability and longevity of your research.
Data Management Training offered for Medical School & School of Public Health

- Organized by the HMS Data Management Working Group
- Based on the data lifecycle for biomedical research
- Has been offered a few times in 2016/2017
- Will be combined with Harvard wide training
Extension of Harvard Dataverse Curation Services

• Led by Sonia Barbosa (IQSS)

• 6 month pilot program with Harvard librarians

• Offers extended curation services to Harvard affiliates (and all users, when possible)

• Evaluating cost-based model

• Plus, office hours once a week
Data Management Support is not Sufficient

Layers of Support

Data Management Support

Data Science Support

Research Computing & Security Support
DataFest 2017 Brings Data Science Basic Training to researchers and staff

Inaugural DataFest reflects a growing interest

Conference builds awareness of data science resources at Harvard

February 3, 2017 | ✓

By Kareem Carr, Harvard Correspondent

The proof of Harvard’s growing interest in data science became even clearer the third week of January when the inaugural session of the
More technology integration and ease-of-use, less training and support
Data Repositories can help Integrate the Data Lifecycle

**Data Acquisition and Planning**

What do I need to know before bringing research data into Harvard? How do I prepare for a data management plan?

- Data User Agreement, Data Management Plan, Harvard Policies, licensed data.

*More >*

**Data Storage**

Where and how should I store my research data? What are the options at Harvard? What do I need to know about security?

- Data files, documentation, logbooks, notebooks, security levels, and permits.

*More >*

**Compute and Analysis**

What are the options for research computing at Harvard? Which tools or methods should I use for my research?

- Harvard Research Computing, data science and computational help.

*More >*

**Data Sharing and Archival**

What is Data Sharing and why is it important? What do Funders and Journals require? Can I get help on data curation?

- Harvard Dataverse repository, domain repositories, Open Data policies.

*More >*

**Preservation Services**

What is long-term preservation? What services do Harvard offer for preservation of data collections?

- Harvard Library services, format migration, suitable medium.

*More >*

**Data Disposal**

Are there some cases when I need to destroy my data? How should I do it? What services do Harvard offer?

- Contractual obligations, method of disposal, documentation

*More >*
Dataverse is an open-source platform for building any type of data repository, including institutional repositories. A growing community of developers and users. http://dataverse.org
An Integrated Data Management and Computing Solution

- Data Acquisition and Planning
- Data Storage
- Compute and Analysis
- Data Sharing and Archival

**E Lab Notebooks, Instruments, Surveys, ...**

- Assign Security Level, DUA
- Track Provenance
- Metadata

**Storage and Computing**

- DOI, Metadata, DUA, Restrictions
- Link citations

**Journals**

DOI, metadata, and DUA are assigned after data collection; Data repository enables **data-centric computing**
Machine-readable, FAIR Data Management Plans can help track data management
In summary:

• Coordinate, coordinate, coordinate (across groups)

• Integrate, integrate, integrate (across technologies)

THANKS!

Mercè Crosas @mercecrosas mercecrosas.com

With contributions from Caroline Shamu, Radhika Khetani, and Sonia Barbosa