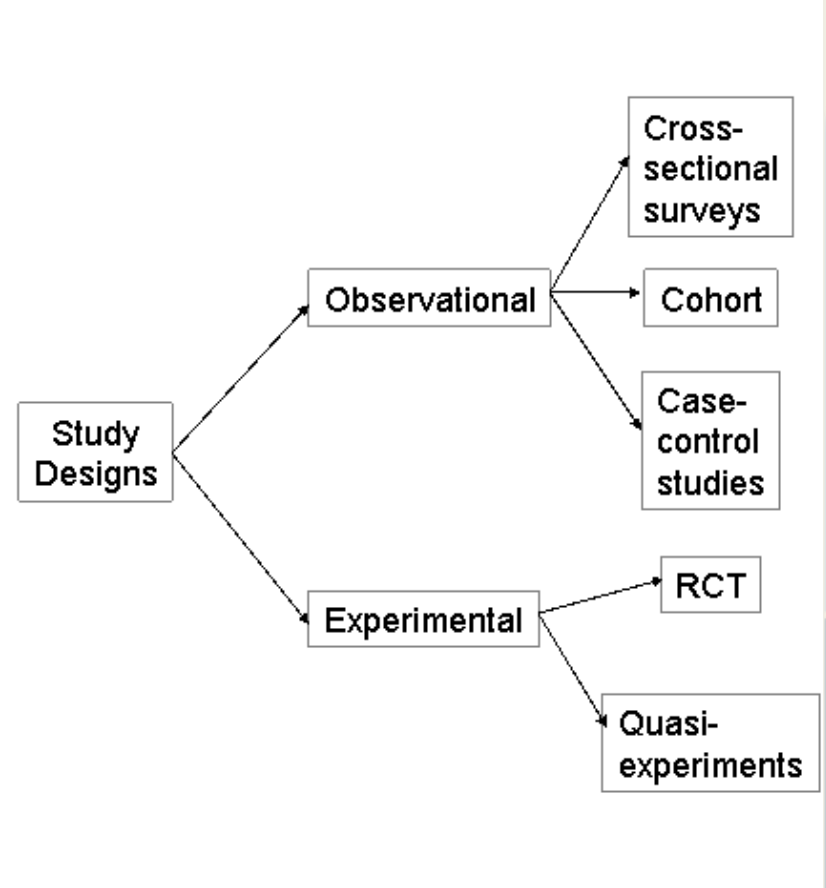
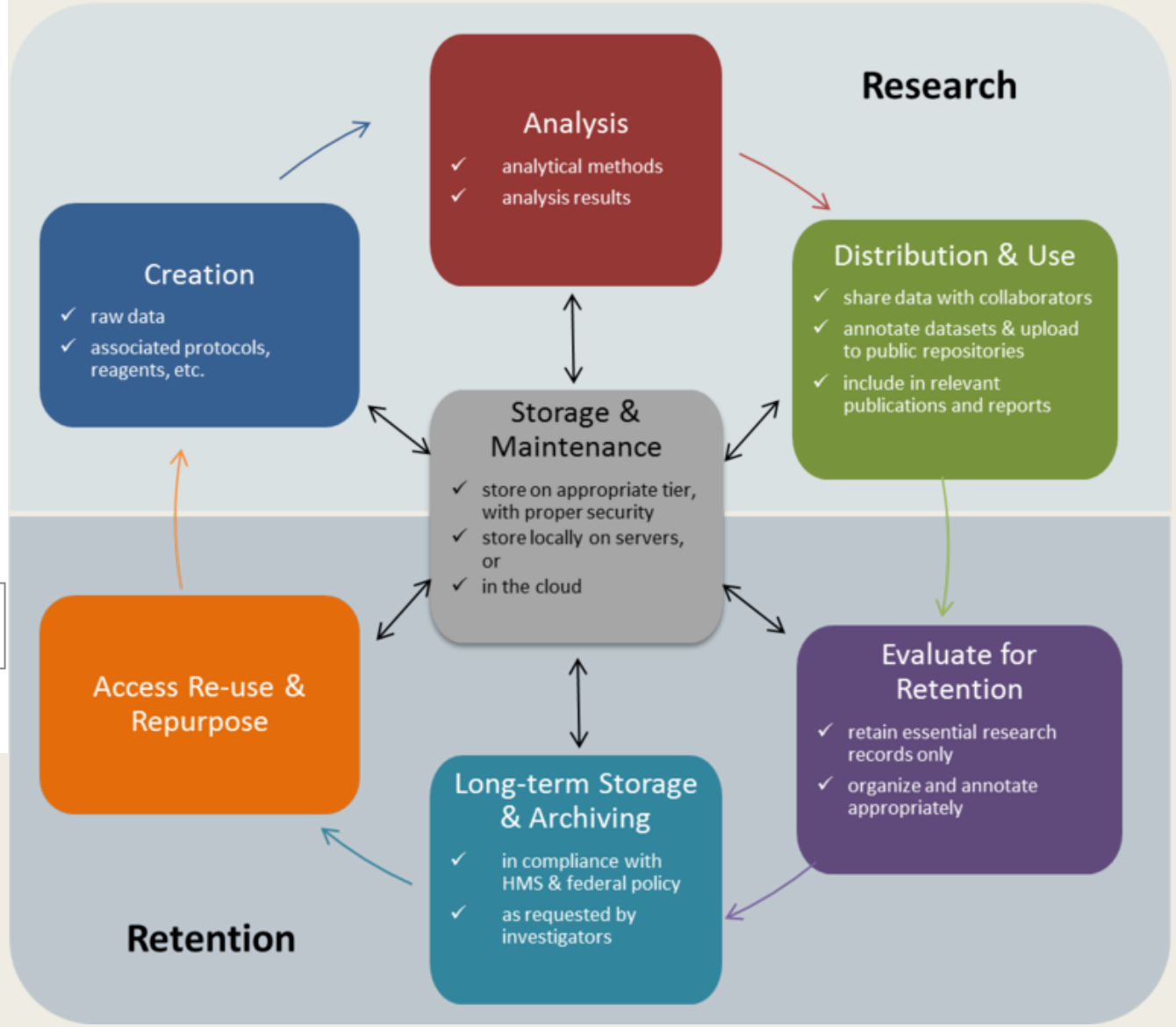
RESEARCH DATA MANAGEMENT

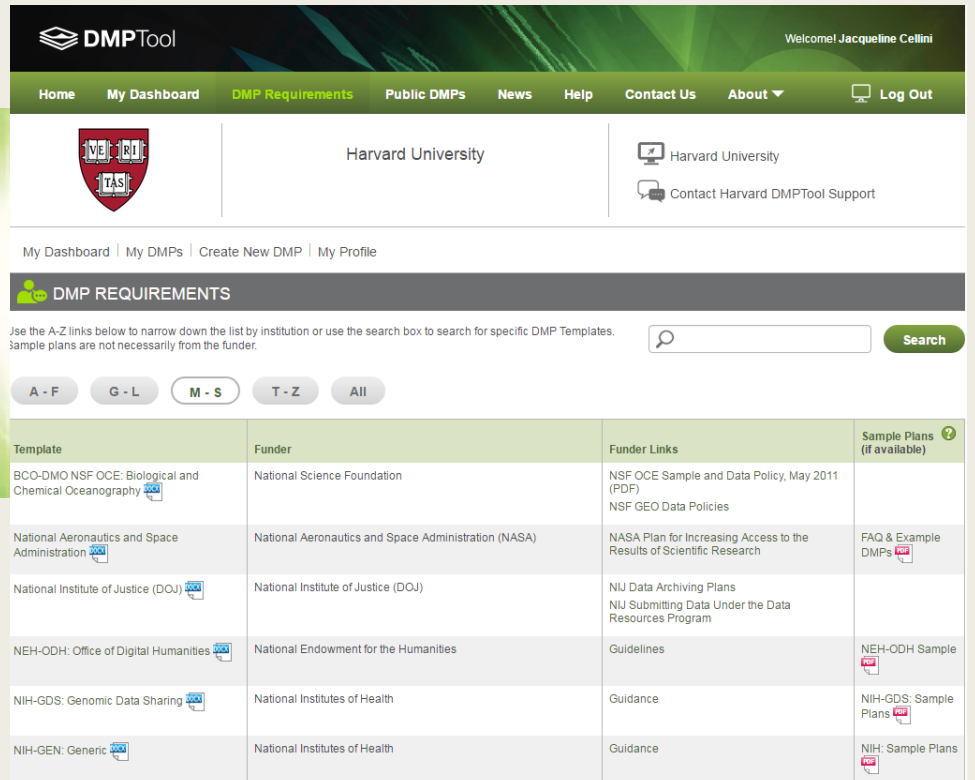
Data lifecycle for biomedical research





Data managed well can be more easily stored, discovered, shared, accessed, interpreted, and reviewed.

A data management plan (DMP) is a written document that describes the data you expect to acquire or generate during the course of a research project, how you will manage, describe, analyze, and store those data, and what mechanisms you will use at the end of your project to share and preserve your data.



Raw data

What is being measured or observed? This is the data that is being generated during the research project.

Processed data

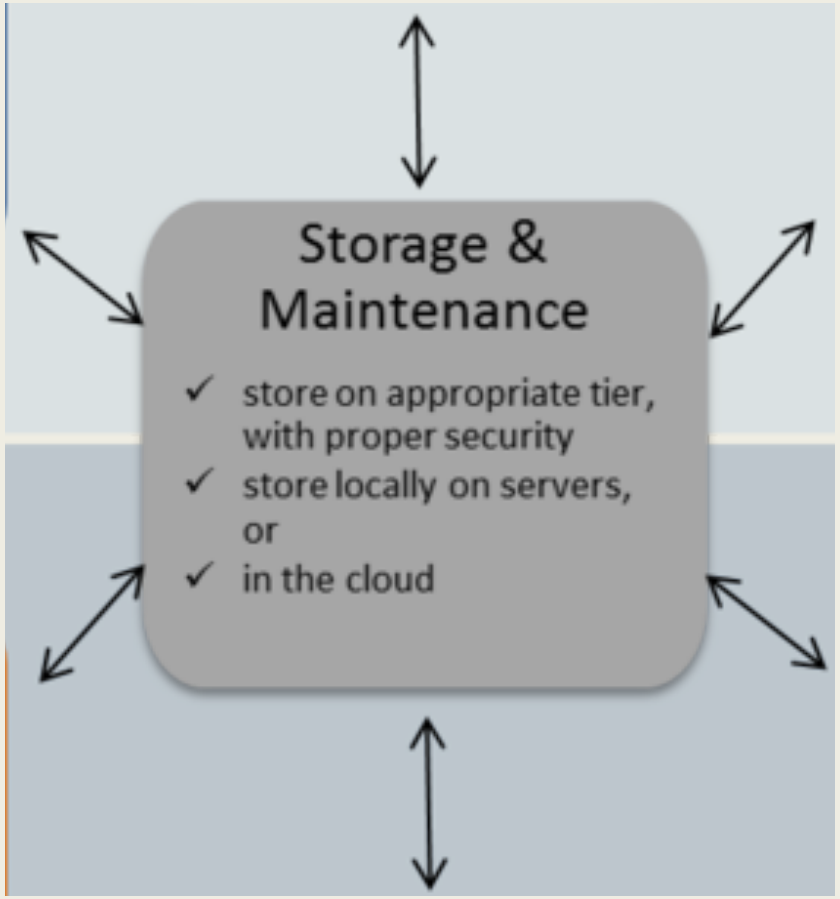
How can the raw data be made useful- able to be manipulated?

Analyzed data

What does the data tell us? Is it significant? How so?

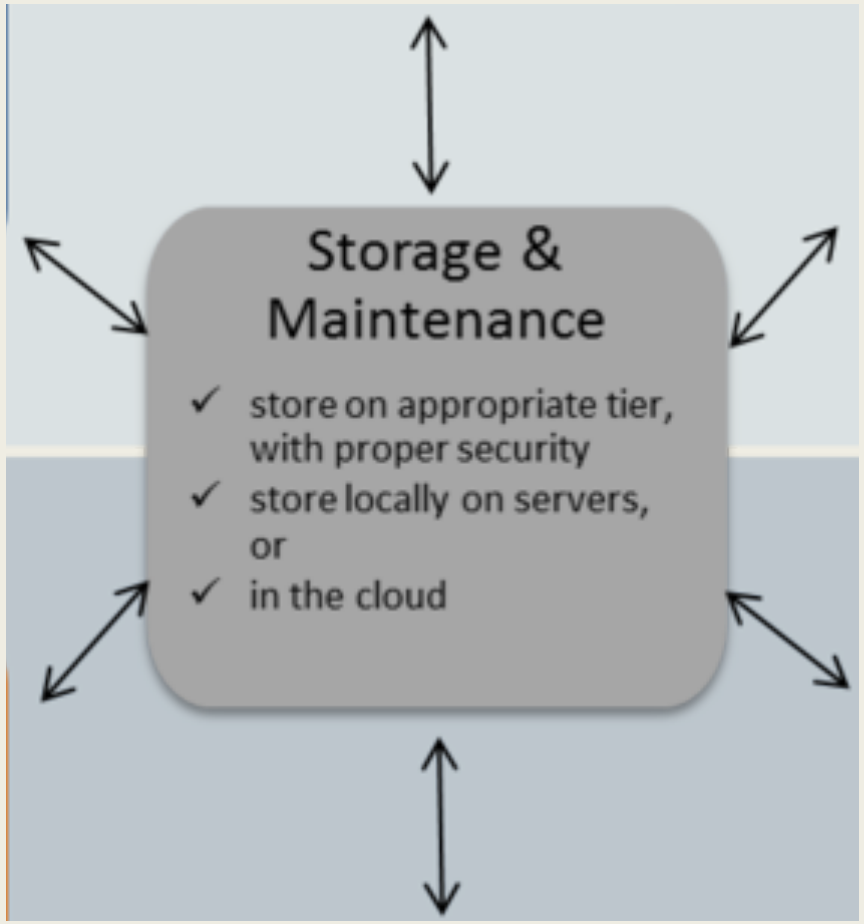
Finalized/published data

How does the data support your research question?



“Good metadata is standardized, consistent and interoperable, and facilitates discovery, preservation and archiving of data.”

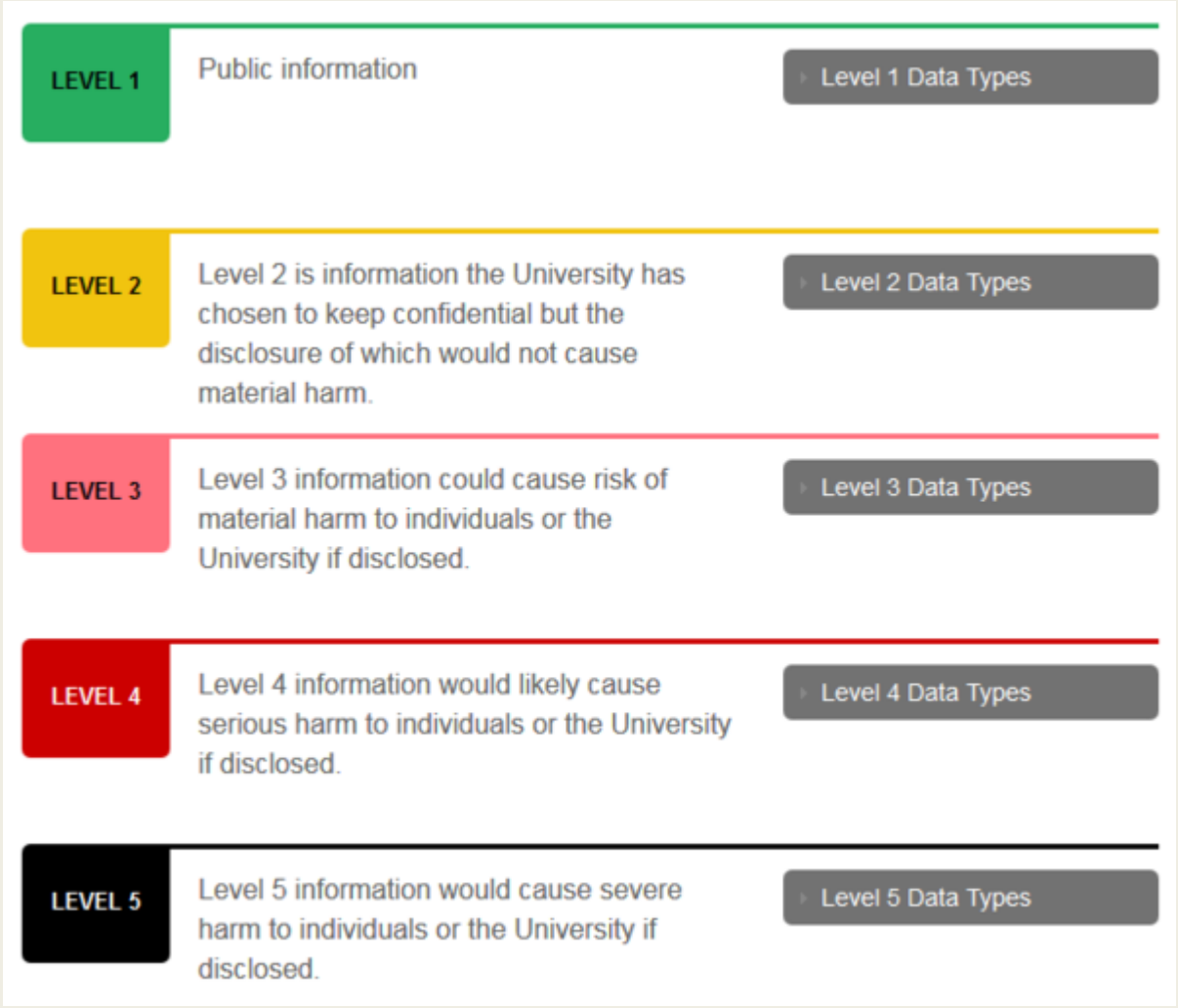
Storage, backup, and security are interrelated.



Access: limiting the availability of your data

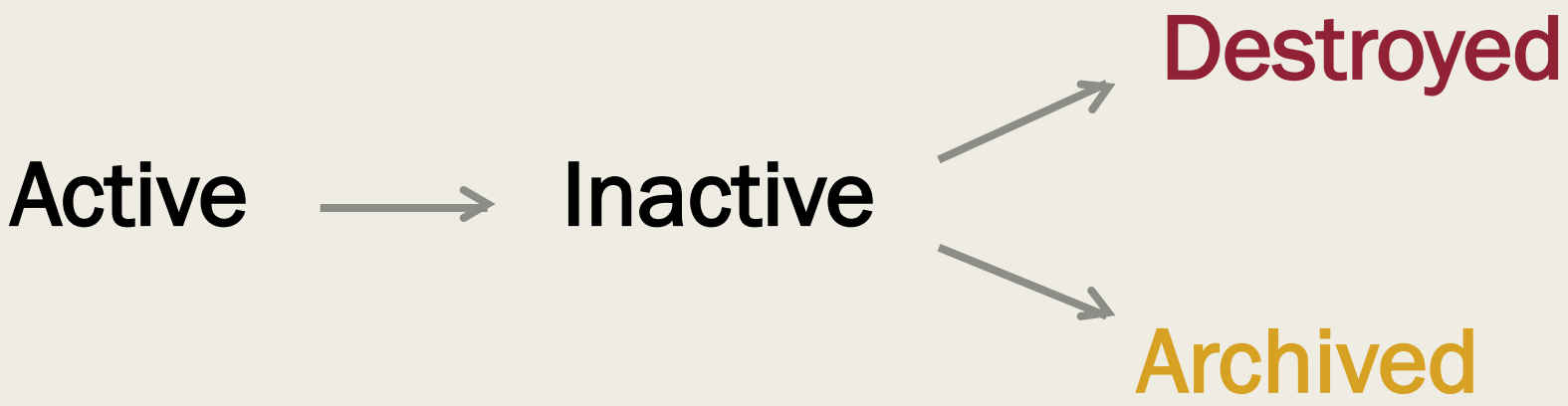
Systems: protecting your hardware and software

Data Integrity: Ensure that your data is not manipulated in an unauthorized way



Who owns the data?

4 Types of Records



Data Sharing

When establishing data sharing and access policies and provisions, consider *whom* you will share your data with, *how* it will be shared, and *when* in the research process you will share it.

Citation / Attribution of Data

Acknowledgement of the use of someone else's information or work is a long-accepted practice in scholarly communication.

The following elements are generally considered the core elements of a data citation:

- Author/Creator(s): the creators of the data; can be one or more people or organizations
- Title: the title of the data set
- Version: the exact version or edition of the data set used
- Publication Date: the date when the data set was published or released
- Publisher/Archive: the data center or repository that is archiving and distributing the data
- Identifier/Locator: URL or other linkable locator for the data; a persistent, permanent URL such as a DOI (Digital Object Identifier) or a handle is preferred

Data Repositories

- Institutional repository (“IR”)
- Disciplinary repository (“DR”)
- Open Repository (“OR)

Retention

Data retention requirements are put in place by funding agencies and sponsoring institutions for a number of reasons, including:

- the need to make research findings available for corroboration,
- to promote the reuse of data within and across disciplines, to support open data initiatives,
- and the need to protect intellectual property rights.

Appraisal & Archiving

Questions?

Meghan Kerr, HMS Archivist/Records Manager
Meghan_Kerr@hms.harvard.edu

Jacqueline Cellini, Reference Librarian HMS Countway Library
Jacqueline_Cellini@hms.Harvard.edu

Heather Mumford, Archivist, Harvard T.H. Chan School of Public Health
heather_mumford@hms.harvard.edu

http://datamanagement.hms.harvard.edu/

References: http://www.edc.pitt.edu/about/data-management/
https://www.med.uottawa.ca/sim/data/Study_Designs_e.htm
Editor: Lamar Soutter Library, University of Massachusetts Medical School
the title of the work: New England Collaborative Data Management Curriculum
the URL where the original work can be found: http://library.umassmed.edu/necdmc
DCC. (2013). *Checklist for a Data Management Plan.* v.4.0. Edinburgh: Digital Curation Centre. Available online: http://www.dcc.ac.uk/resources/data-management-plans