

# Data Intensive Science

To learn more about the new data intensive scientific paradigm, read *The Fourth Paradigm*, a collection of essays that explore the present and future of 21st Century science.

The book is available in the library or for free online at <http://research.microsoft.com/en-us/collaboration/fourthparadigm/>



OR SCAN THE QR CODE TO GO TO THE SITE!

## POLICIES @ BC

For more information on the research policies of Boston College that apply to digital research data, visit the website of the Offices for Research Compliance and Responsible Conduct of Research.

<http://www.bc.edu/research/rcradmin/default.html>

## What is a Data Management Plan?

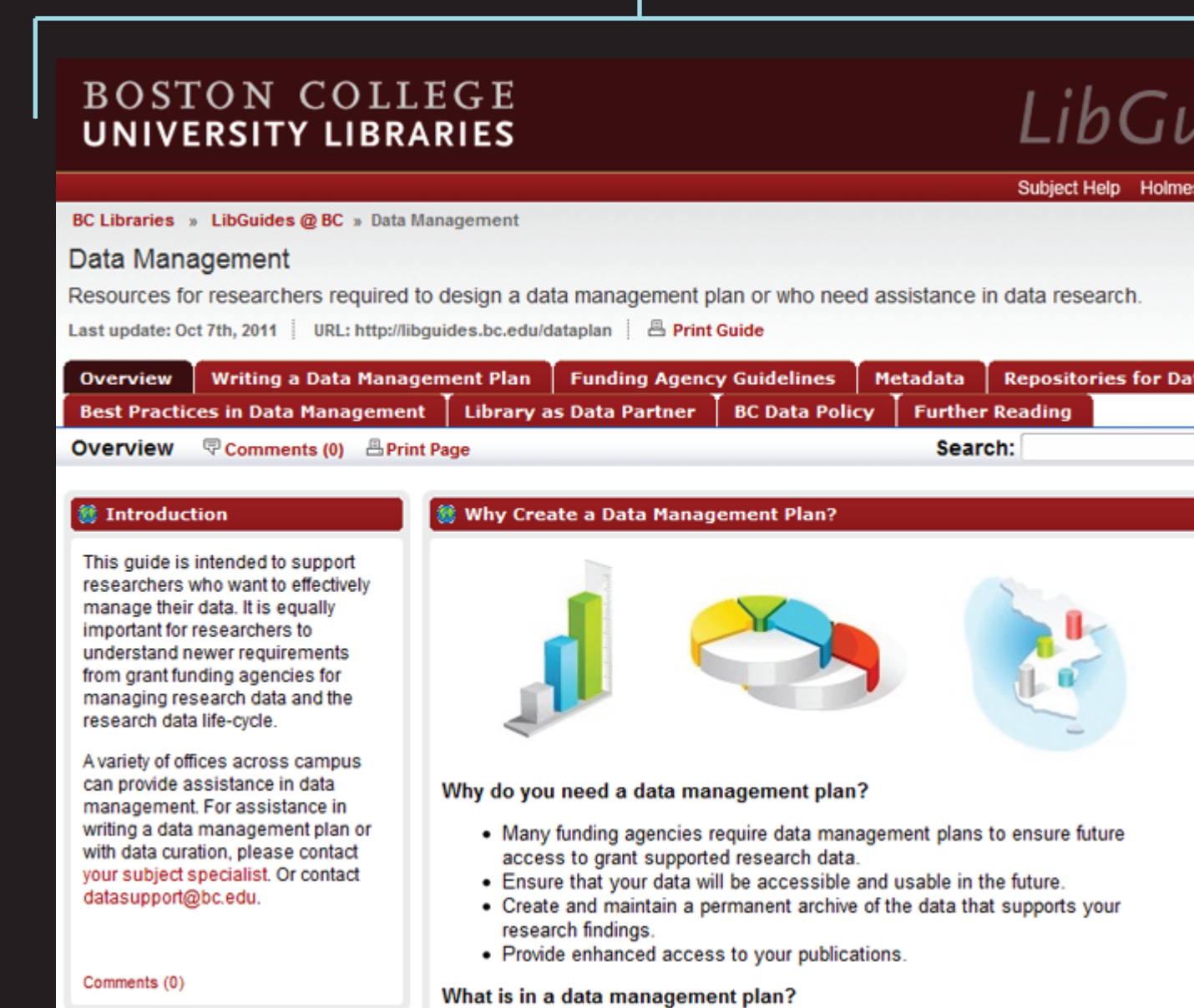
A Data Management Plan is a document that has (among other things) descriptions of the types of data, samples, and physical collections to be generated, the metadata that will describe it, the policies for sharing, accessing, and re-using data, and the methods for archiving and preserving the research data. If you need assistance with developing a data management plan, visit the Data Management LibGuide or contact your subject specialist.

This Data Management LibGuide is available online at: <http://libguides.bc.edu/dataplan>

Learn about data management plans, metadata, data set citations, how Boston College librarians can help you and more!



OR SCAN THE QR CODE TO GO TO THE SITE!



## Data Collection: Applying the Data Management Plan

In this stage of the Lifecycle, research data is generated. It is stored, secured and managed according to the details in the Data Management Plan. Boston College [RESEARCH SERVICES](#) provides secure storage for digital research data. Also in this stage, researchers use [METADATA](#) to document the research data that they have collected.

### RESEARCH SERVICES

Do you have research needs that involve computers, data (collecting, transferring, storage, backup, security, management, mining, qualitative or quantitative analysis, publishing, sharing), software, or online surveys? If so, contact Research Services, an office of Information Technology Services:

<http://www.bc.edu/offices/researchservices.html>

## Funding Data

Earlier in 2010, many funding agencies such as the National Science Foundation and the National Institutes of Health, began requiring that researchers include a data management plan in their grant proposals.

Practicing good data management will prevent data loss and ensure the long term preservation and accessibility of your data. Developing a [DATA MANAGEMENT PLAN](#) in order to secure funding for research is the first stage of the digital data research lifecycle.

## Sharing Data

Sharing data to be reused is the crux of data intensive science. Reusing data allows more discoveries and conclusions to be made from one data set, thereby saving money and increasing the impact of the research data. In addition to requiring Data Management Plans, the funding agencies have mandated that publicly funded research data be shared. Various journals like Science and Nature also require that data be publicly available. Research data generated at Boston College can be deposited into [eSCHOLARSHIP](#) and/or a subject specific repository.

### DIGITAL REPOSITORY

eScholarship@BC is a service of the Boston College University Libraries and provides the university community with a publishing platform for the dissemination of scholarly materials. The repository will be accepting data sets as well.

[escholarship.bc.edu](http://escholarship.bc.edu)

## Digital Research Data Lifecycle

## Research Output & Reuse

The final stage of the Lifecycle is the culmination of all of a researcher's hard work. The data is analyzed to produce a scholarly research paper, conference presentation or poster. The data set is now available for others to reuse in their projects. A shared data set could be used as the basis for a new research project, continuing its path on the Digital Research Data Lifecycle. Researchers receive recognition for their data sets in the form of citations.

## Metadata

Metadata renders data sets searchable in an archive or repository, easily located from a citation, and easily understood by people who might want to reuse it. A metadata *standard* is a set group of elements that have been standardized for a particular field. Some disciplines already have established metadata standards for data sets. Below are some examples.

METADATA STANDARDS A FEW EXAMPLES	SOCIAL SCIENCES
	Data Documentation Initiative (DDI)
	HUMANITIES
	Text Encoding Initiative (TEI) Visual Resources Association Core (VRA)
	GENERAL
	Dublin Core (DC) Metadata Object Description Schema (MODS)
	NATURAL SCIENCES
	Ecological Metadata Language (EML) Darwin Core Integrated Taxonomic Information System (ITIS)
	PHYSICAL SCIENCES
	Standard for the Exchange of Earthquake Data (SEED) Federal Geographic Data Committee (FGDC) Geospatial Interoperability Framework (GIF)

## Obsolete Data

What happens to data that is not properly managed? If saved formats are not updated, the data becomes inaccessible. For example, digital data used to be saved on tapes like the roll to the right. By today's storage standards, this format is obsolete. Even 3.5" floppy disks, which are just over twenty years old, are rarely used. Planning for format transfers and using open source storage formats ensures long term preservation of data sets.