Published on *Data Management Plan Tutorial* (https://www.e-education.psu.edu/dmpt)

Part 3: Policies for Data Sharing and Access

Kim Addresses Issues of Data Access and Restrictions

Having explored data repositories for their metadata requirements and studied the disciplinary standards offered by organizations such as the Digital Curation Center in the U.K., Kim has begun work on applying metadata standards for documenting Dr. Smart's data. Dr. Smart has also informed Kim that he intends to make the data publicly available. He instructs Kim to make certain there are no issues that would prevent broad sharing. Kim will be seeking guidance on the issues and questions that pertain to sharing and access to ensure that management of this research data does not create security or confidentiality breaches, or impinge on any intellectual property rights.

"Strictly Confidential" - Image of a Currier & Winter *carte de visite*

In this part, you will learn about the importance of providing safeguards for protection of confidential information and requirements for restricting data. You will also consider how and when access to your data will be provided, and to whom. We also touch on the importance of data citation.

3.1 Access and Availability of Data

A major reason why federal funding agencies require DMPs is to encourage researchers to think as early in the project as possible about how and when data will be shared and made available. Funding agencies expect you to be clear in the DMP about your approach or policy for sharing and giving access to your data. Sharing also makes possible reuses and repurposing of data, as explained later in this tutorial.

Ways of sharing data can vary. They depend on the research domain, on the availability of services, on the size of the data set, and other factors. Many of these are covered later in this tutorial in the section which addresses long-term preservation of data. Preserving data ensures long-term access to it. What you want to make sure your DMP states is that you will deposit your data in a repository - whether it's a repository dedicated to data from your research domain or an institutional repository that accepts data sets. Penn State has such a repository - ScholarSphere [2], about which you may learn more here [3], in the tutorial.

You should note in the DMP when data will be made available - i.e., during the project, or afterward? Will there be any embargoes? If so, why and for how long? Will data be shared indefinitely, or will there be temporal constraints?

Will all the data be shared, or only a portion? Have you adequately addressed levels of access, e.g., who is allowed to use the data, etc.? (You may already have addressed this in the section of the plan related to data access.)

3. Policies for Access and Sharing
After publication of manuscripts based on the data we collect, we will share our data and metadata with the NutNet community via data updates sent annually as .csv files from the existing central relational database. Other NutNet users will need to contact Lind for access to the data.

We will also submit both of our datasets (abundance and stoichiometry) to the U of M Digital Conservancy, an archive for digital preservation. Borer has access to this resource as a faculty member. This will occur within a year of publication. The data will be publicly available via the Digital Conservancy, which provides a permanent URL for digital documents.

Part of an example data management plan provided by DataONE.
Click to expand to provide more information

3. Policies for Access and Sharing

After publication of manuscripts based on the data we collect, we will share our data and metadata with the NutNet community via data updates sent annually as .csv files from the existing central relational database. Other NutNet users will need to contact Lind for access to the data.

We will also submit both of our datasets (abundance and stoichiometry) to the U of M Digital Conservancy, an archive for digital preservaton. Borer has access to this resource as a faculty member. This will occur within a year of publication. The data will be publicly available via the Digital Conservancy, which provides a permanent URL for digital documents.

Credit: Data Management Plan: NSF General [4]

In DMPs, researchers often state that their findings will be published as articles and other genres relevant to their particular disciplines. Increasingly, publications are linking to data sets that have been deposited into repositories; this is another reason why depositing your data into repositories is important.

The NSF and other funding agencies that require DMPs now frown on the previously common practice of sharing data sets upon request, such as via email. Researchers' email addresses are not permanent, for one thing, and on-demand access unnecessarily burdens the researcher with the redundant act of locating and attaching, or pointing to, the requested data set. By submitting the data set to a repository, data become publicly discoverable, findable, and accessible.

While typically not mentioned in the DMP, the citation of data marks another method of sharing data and attributing the researchers who created them. The organization DataCite [5] has examples of citing data [6]. The most crucial component of such a citation is a Digital Object Identifier or DOI [7]. DOIs are assigned to ensure persistent access. Even though the digital object may undergo changes over time, the DOI stays the same.

Another benefit to researchers of having your data made available in a repository is that it becomes more widely accessible and citable. Citation of data is increasingly required by editors and having a permanent URL or repository identifier helps others find your data, as Andrew Stephenson describes in the following video:

Link to YouTube video. [8]

Making publicly funded data available is a common good, however, there are instances where it is important to restrict data. Thus, in a DMP researchers should note what data will be shared and whether there are any restrictions to sharing.
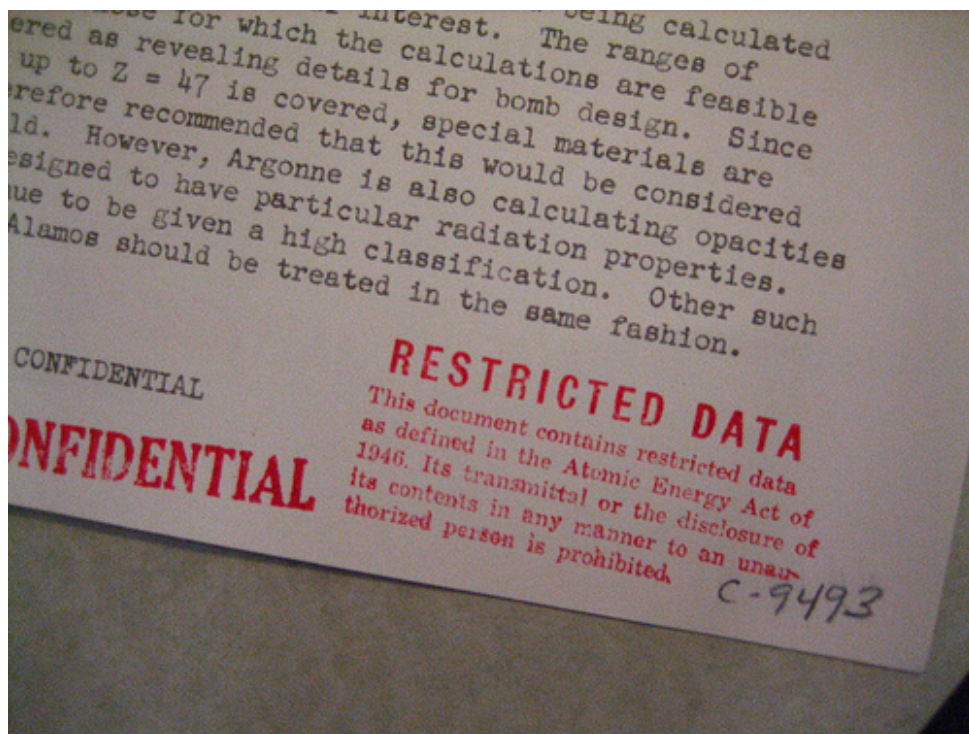
A "restricted data" stamp from the 1950s on documentation about opacity classifications.
Credit: restricted data on Flickr [9]. CC-BY 2.0.

3.2 Confidentiality and Ethics Issues

Confidential data usually includes personally identifiable information (PII), such as names, addresses, and Social Security numbers - anything that might point to the identity of a person. Health information data, for example, is protected under the Health Information Privacy and Accountability Act (HIPAA), because it is considered private data.

From the website of the U.S. Department of Health and Human Services.
Credit: U.S. Department of Health and Human Services [10]

Data collection stemming from human participant research - such as survey data - and thus requiring approval from an Institutional Review Board, or IRB, typically produces confidential data, access to which must be restricted because of PII. Sharing such data would be a violation of both research ethics and human subjects' privacy.

If your research data encompasses any of the above ethics and confidentiality concerns, then you should note these in the DMP. If your project will be generating survey data, then you should also state an intention to comply with Penn State's IRB requirements set by research administration guidelines and policies. Consult Penn State's Human Subjects Research (IRB) site [11]. Review RA 22: HIPAA and Research at Penn State University [12], or, if applicable, RA 23: HIPAA and the Milton S. Hershey Medical Center and Penn State College of Medicine [13].

Anonymity In Genetic Research Can Be Fleeting

by VERONIQUE LACAPRA

January 17, 2013   3:47 PM

From NPR's blog on health news, "Shots."

Listen to the Story

All Things Considered
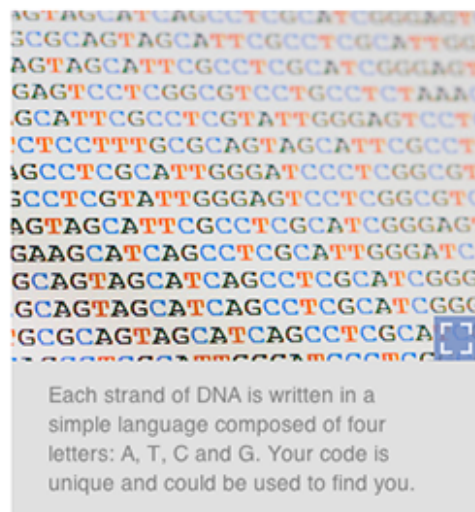
3 min 48 sec

Playlist
Download
Transcript

Each strand of DNA is written in a simple language composed of four letters: A, T, C and G. Your code is unique and could be used to find you.

iStockphoto.com

People who volunteer for medical research usually expect to remain anonymous. That includes people who donate their DNA for use in genetic studies.

But now researchers have shown that in some cases, they can trace research subjects' DNA back to them with ease. And they say the risk of being identified from genetic information will only increase.

An international team of scientists wondered how difficult it would be to identify people whose DNA had been posted anonymously on the Internet. They decided to focus on 10 men whose entire genetic blueprints had been made public for research purposes.

Share

4 Comments

Keeping data private and thus not traceable back to human subjects is a persistent challenge. Credit: NPR [14]

3.3 Intellectual Property, Embargoes, and Protections

One way to help ensure that confidential data, or data related to patented research, for example, is not shared prematurely is to implement an embargo. Embargoes are intended to bar data sharing and access for a limited period of time, thus protecting the data and the intellectual property rights of the researchers. Typically renewable, an embargo allows the researcher to decide when data may be released. The DMP should state whether an embargo will be imposed on the collected data or not.

Statement by an artist requiring a confidentiality agreement from patrons.
Credit: mlinksva on Flickr [15]. CC BY 2.0.

Most universities provide guidance and policies on intellectual property rights for research. Penn State does at the General University Reference Utility (GURU) site [16]. It advises on ownership and management of intellectual property (Policy IP01 - Ownership and Management of Intellectual Property [17]), as well as guidance on the intellectual property rights of students (Guideline IPG01 - Faculty Guidance on Student Intellectual Property Rights [18]), and (Guideline IPG02 - Special Student Intellectual Property Agreement Forms [19]),While a DMP does not ordinarily contain this depth of information, it is a recommended practice for researchers to review their institutions' policies and guidelines on intellectual property rights.

RESEARCH ADMINISTRATION POLICIES AND GUIDELINES

Research Administration Policies

- Introduction - Research Administration Policies
- RA01 - Determining the Allowability, Allocability, and Reasonableness of Costs on Sponsored Projects
- RA02 - Acceptability and Evaluating of Services, Equipment or Property Donated to Federally Sponsored Projects
- RA03 - Preparing Budgets (Cost Estimates) For Sponsored Projects
- RA04 - Making Revisions to Budgets and Program Plans on Federally Sponsored Projects
- RA05 - Significant Financial Interest Disclosure for Sponsored Project's Investigators (replaced 06-17-09 by RAG20)
- RA06 - Cost Recovery for Sponsored Projects
- RA07 - Use of Equipment Purchased on Federally Sponsored Projects
- RA08 - Reimbursement Salary Caps and Plan Effort Confirmation (formerly Salary Rate Limitations and Effort Confirmation)
- RA09 - Determining and Disposing of Program Income on Federally Sponsored Projects
- RA10 - Addressing Allegations of Research Misconduct (Formerly Handling Inquiries/Investigations Into Questions of Ethics in Research and in Other Scholarly Activities)
- RA11 - Patents and Copyrights (Intellectual Property) (Replaced 01-07-13 by IP01)
- RA12 - Technology Transfer and Entrepreneurial Activities (Faculty Research)
- RA13 - Co-authorship of Scholarly Reports, Papers and Publications (Replaced 01-07-13 by IP02)
- RA14 - The Use of Human Participants in Research
- RA15 - Care and Use of Vertebrate Animals
- RA16 - Administration of Sponsored Project Subcontracts of the University
- RA17 - Courseware (Replaced 01-07-13 by IP03)
- RA18 - Compliance with Federal Export Regulations
- RA19 - Administration of Sponsored Project Subawards of the University
- RA20 - Disclosure and Management of Significant Financial Interests (formerly *Individual Conflict of Interest*)
- RA21 - Institutional Financial Conflict of Interest Involving Sponsored Projects, Dedicated Gifts, Research, Scholarship, and Technology Transfer (Replaced 09-10-13 by AD83)
- RA22 - HIPAA and Research at Penn State University
- RA23 - HIPAA and Research at The Milton S. Hershey Medical Center And Penn State College Of Medicine
- RA24 - Position Statement on Research

Research Administration Guidelines

- RAG01 - Implementing Revised OMB Circular A-21 (Section F.6.b - Departmental Administration Expense) REVISED 4/26/96, and Further Amended 8/29/97)

Screenshot of Penn State's General University Reference Utility (GURU) page.
Credit: Penn State University [20]

In IRB-approved research, an informed consent agreement between the researcher and the study participants is de rigeur. Depending on the nature and sensitivity of the data, however, researchers may wish to consider including an option in the consent form that allows data sharing. Refer to the guidance provided by the UK Data Archive [21] on informed consent forms, as suggested by the MIT Libraries [22].

Finally, in this section of the proposal, where data protections, sharing, and access are discussed, it's important to state outright who on the project will be responsible for monitoring embargoes, consent forms, and non-disclosure/confidentiality agreements. A key part of managing research data is being explicit about who will adopt what role in such management.

Archivists receive formal training related to these issues. Ben Goldman shares some of his expertise regarding security and access to digital data in the video below.
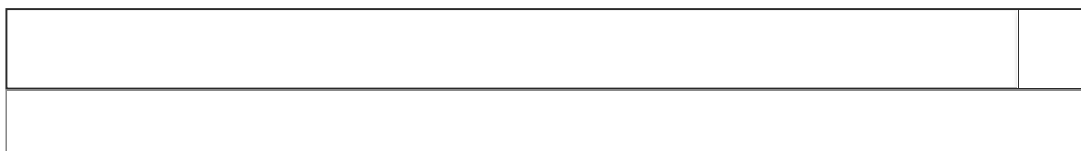
<u>Link to YouTube video.</u> [23]

3.4 Summary

While funding agencies created the DMP requirement to encourage sharing of data, there are data produced on funded research projects that are sensitive and need to be restricted for access and use. These kinds of data tend to include personally identifying information (PII), such as might be collected as the result of demographic surveys or those found in health records. Often such data are protected under federal law, such as the <u>Health Information Privacy and Accountability Act (HIPAA)</u> [10].

If the data resulting from your project are likely to have confidentiality or ethics issues (such as human subject research engenders), then these should be noted in your DMP. Be sure to consult Penn State's Human Subject Research (IRB) site, as well as <u>RA22: HIPAA and Research at Penn State University</u> [12]. If applicable, review <u>RA 23: HIPAA and the Milton S. Hershey Medical Center and Penn State College of Medicine</u> [13].

There can be intellectual property rights issues as well, particularly with patent-pending research, that can prevent the sharing of data for a period of time. In such situations, an embargo may need to be implemented for your data, and you should state such in your DMP, specifying when the embargo will be lifted.

A DMP does not need to go into detail regarding confidential data and intellectual property rights, beyond stating what kind of restrictions on access are likely. For guidance during a project, Penn State provides faculty, students, and staff with <u>research guidelines and policies</u> [20] that can be helpful to draw on.

Finally, if possible, the DMP should state whose role it will be to monitor embargoes, consent forms, non-disclosure/confidentiality agreements, and so on.

Check Your Understanding

Which of the following types of data would be considered confidential?

(a) Social Security numbers
(b) Student grades
(c) Personal medical information
(d) All of the above

Click for answer.

ANSWER: (d) All of the above. Confidential information is often information that might lead to the identification of the human subject or to more information about the human subject who participated in a study. Special measures, such as redaction or anonymization, are typically taken to prevent breaches of confidence.

Source URL: https://www.e-education.psu.edu/dmpt/node/665

Links
[1] http://www.flickr.com/photos/beinecke_library/5166407009/in/set-72157625240109163
[2] https://scholarsphere.psu.edu/
[3] https://www.e-education.psu.edu/dmpt/node/680
[4] http://www.dataone.org/sites/all/documents/DMP_NutNet_Formatted.pdf
[5] http://www.datacite.org/
[6] http://www.datacite.org/whycitedata
[7] http://www.datacite.org/whatisdoi
[8] http://www.youtube.com/watch?v=x2Ftl0sDdAo
[9] http://www.flickr.com/photos/restricteddata/6322624283/in/photostream/
[10] http://www.hhs.gov/ocr/privacy/
[11] http://www.research.psu.edu/orp/humans
[12] http://guru.psu.edu/policies/RA22.html
[13] http://guru.psu.edu/policies/RA23.html
[14] http://www.npr.org/blogs/health/2013/01/17/169609144/anonymity-in-genetic-research-can-be-fleeting
[15] http://www.flickr.com/photos/mlinksva/10753178366/in/photostream/
[16] http://guru.psu.edu/
[17] https://guru.psu.edu/policies/IP01.html
[18] https://guru.psu.edu/policies/ipg01.html
[19] https://guru.psu.edu/policies/IPG02.html
[20] http://guru.psu.edu/policies/#RESEARCH
[21] http://www.data-archive.ac.uk/create-manage/consent-ethics/consent?index=0
[22] http://libraries.mit.edu/guides/subjects/data-management/ethical.html
[23] http://www.youtube.com/watch?v=LEmr5NDtgUA