Data Management Plans & LaTeX

EEOB/BCB 546

March 23, 2022

Data Management Plans

Reproducible Research

Biological Data

Biology has become a data-intensive field. Without a comprehensive plan for collecting, storing, maintaining, and disseminating your data and research products your work will not be reproducible and your contribution will be limited.

Reproducible Research

Biological Data

Biology has become a data-intensive field. Without a comprehensive plan for collecting, storing, maintaining, and disseminating your data and research products your work will not be reproducible and your contribution will be limited.

Research Products

In addition to data, biology research also yields other products that are necessary for reproducibility and are tools that can advance the field (*e.g.*, software, scripts, databases, tutorials).

Planning for Data

Write it down

A Data Management Plan is a written plan for dealing with scientific data and all of the products of a research project. This plan accounts for how data (and software, tools, etc.) will be handled during a research project and *after* the project is completed.

Proposing Reproducible Research

White House Mandate

In 2013 a mandate from President Obama required that results of all federally funded research be made publicly available and that these funding agencies develop plans for data management.

Proposing Reproducible Research

White House Mandate

In 2013 a mandate from President Obama required that results of all federally funded research be made publicly available and that these funding agencies develop plans for data management.

DMPs are required for funding

NSF, NIH, USDA, NASA, DOE, HHS, CDC, FDA, NIST, NOAA, USAID, AHRQ etc.

These and many other funding agencies require data management plans for all proposed research projects and/or awards.

Requirements of Data Management Plans

Suggested Standards

In a DMP, you have to consider:

- **data type:** genomes, 3D models, CT scans, images, spreadsheets, alignments, field observations, audio/video recordings, etc.

Requirements of Data Management Plans

Suggested Standards

In a DMP, you have to consider:

- **data type:** genomes, 3D models, CT scans, images, spreadsheets, alignments, field observations, audio/video recordings, etc.
- **data format:** file types, file sizes, how metadata will be stored, etc.

Requirements of Data Management Plans

Suggested Standards

In a DMP, you have to consider:

- **data type:** genomes, 3D models, CT scans, images, spreadsheets, alignments, field observations, audio/video recordings, etc.
- **data format:** file types, file sizes, how metadata will be stored, etc.
- **data storage & preservation:** keeping data safe (backup), long-term storage and curation

Requirements of Data Management Plans

Suggested Standards

In a DMP, you have to consider:

- **data type:** genomes, 3D models, CT scans, images, spreadsheets, alignments, field observations, audio/video recordings, etc.
- **data format:** file types, file sizes, how metadata will be stored, etc.
- **data storage & preservation:** keeping data safe (backup), long-term storage and curation
- **data sharing:** public access of data, sharing agreements, access to physical collections, privacy issues (like for clinical data), timing of data availability

Requirements of Data Management Plans

Suggested Standards

In a DMP, you have to consider:

- **data type:** genomes, 3D models, CT scans, images, spreadsheets, alignments, field observations, audio/video recordings, etc.
- **data format:** file types, file sizes, how metadata will be stored, etc.
- **data storage & preservation:** keeping data safe (backup), long-term storage and curation
- **data sharing:** public access of data, sharing agreements, access to physical collections, privacy issues (like for clinical data), timing of data availability
- **publishing & dissemination:** venues for reporting results, access to published papers (archiving or open access)

Requirements of Data Management Plans

Suggested Standards

In a DMP, you have to consider:

- **data type:** genomes, 3D models, CT scans, images, spreadsheets, alignments, field observations, audio/video recordings, etc.
- **data format:** file types, file sizes, how metadata will be stored, etc.
- **data storage & preservation:** keeping data safe (backup), long-term storage and curation
- **data sharing:** public access of data, sharing agreements, access to physical collections, privacy issues (like for clinical data), timing of data availability
- **publishing & dissemination:** venues for reporting results, access to published papers (archiving or open access)
- **roles & responsibilities:** which members of the project will carry out components of the DMP

Exercise

Take a few minutes to write down the various aspects of data management needed for your research. Consider the following questions:

- What kinds of data are you generating?
- What are the current standards in your field for storing, preserving, and sharing data?
- Are there any potential limitations or restrictions you have to consider when sharing or storing your data?

NSF BIO Data Management Plan

From the BIO Directorate Guide

Description of Data Used and/or Generated

- Describe the types of data, physical samples or collections, software, curriculum materials, and other materials to be produced in the course of the project.
- Describe the standards to be used for all the data types anticipated, including data or file format and metadata.

Accountability

- Describe the roles and responsibilities of all parties with respect to the management of the data

Data Sharing

- Describe the dissemination methods will be used to make the data and metadata available to others during the period of the award and any modifications or additional technical information regarding data access after the grant ends.
- Describe the PI's policies for data sharing, public access and re-use, including re-distribution by others and the production of derivatives.

Protection of Data: Security and Integrity

- Where appropriate, include provisions for protection of privacy, confidentiality, security, intellectual property rights and other rights.

Data Preservation

- Where relevant, describe plans for archiving data, samples, and other research products, and for on-going access to these products through their lifecycle of usefulness to research and education.

Examples of Data Management Plans

NSF-BIO Projects

NSF DEB-1556615/DEB-1556701

The DMP for a collaborative award entitled: *Advancing Bayesian Phylogenetic Methods for Synthesizing Paleontological and Neontological Data*

- Overleaf URL: `http://bit.ly/2KxWJVq`

NSF IOS-1546719

The DMP for a collaborative award entitled: *The Genetics of Highland Adaptation in Maize*

- Overleaf URL (entire proposal): `http://bit.ly/2o6NR0K`

Writing Data Management Plans

Many of you will write grant proposals during your research careers.

- the ISU library provides resources to assist with writing DMPs, including a detailed guide and self assessment

Writing Data Management Plans

Many of you will write grant proposals during your research careers.

- the ISU library provides resources to assist with writing DMPs, including a detailed guide and self assessment
- there is an online tool called DMPTool that provides a templates for various agency DMPs

Writing Data Management Plans

Many of you will write grant proposals during your research careers.

- the ISU library provides resources to assist with writing DMPs, including a detailed guide and self assessment
- there is an online tool called DMPTool that provides a templates for various agency DMPs
  - it is important to note that online proposal tools like this may not be completely current with the latest requirements of any given agency, so it is critical that you always verify that your document meets these requirements

Writing Data Management Plans

Many of you will write grant proposals during your research careers.

- the ISU library provides resources to assist with writing DMPs, including a detailed guide and self assessment
- there is an online tool called DMPTool that provides a templates for various agency DMPs
  - it is important to note that online proposal tools like this may not be completely current with the latest requirements of any given agency, so it is critical that you always verify that your document meets these requirements

What are important data management issues for your research?

Think about your own work and the data you collect, how would you craft a DMP for your project?

- data type, data format, data storage & preservation, data sharing, publishing & dissemination, roles & responsibilities

10 Simple Rules for Creating a Good DMP


PLOS | COMPUTATIONAL BIOLOGY

https://doi.org/10.1371/journal.pcbi.1004525

PERSPECTIVE

Ten Simple Rules for Creating a Good Data Management Plan

William K. Michener*

1. Determine the research sponsor requirements

10 Simple Rules for Creating a Good DMP


PLOS | COMPUTATIONAL BIOLOGY

https://doi.org/10.1371/journal.pcbi.1004525

PERSPECTIVE

Ten Simple Rules for Creating a Good Data Management Plan

William K. Michener*

1. Determine the research sponsor requirements
2. Identify the data to be collected

10 Simple Rules for Creating a Good DMP


PLOS | COMPUTATIONAL BIOLOGY

https://doi.org/10.1371/journal.pcbi.1004525

PERSPECTIVE

Ten Simple Rules for Creating a Good Data Management Plan

William K. Michener*

1. Determine the research sponsor requirements
2. Identify the data to be collected
3. Definte how the data will be organized

10 Simple Rules for Creating a Good DMP


PLOS | COMPUTATIONAL BIOLOGY

https://doi.org/10.1371/journal.pcbi.1004525

PERSPECTIVE

Ten Simple Rules for Creating a Good Data Management Plan

William K. Michener*

1. Determine the research sponsor requirements
2. Identify the data to be collected
3. Definte how the data will be organized
4. Explain how the data will be documented

10 Simple Rules for Creating a Good DMP


PLOS | COMPUTATIONAL BIOLOGY

https://doi.org/10.1371/journal.pcbi.1004525

PERSPECTIVE

Ten Simple Rules for Creating a Good Data Management Plan

William K. Michener*

1. Determine the research sponsor requirements
2. Identify the data to be collected
3. Definte how the data will be organized
4. Explain how the data will be documented
5. Describe how data quality will be assured

10 Simple Rules for Creating a Good DMP

1. Determine the research sponsor requirements
2. Identify the data to be collected
3. Definte how the data will be organized
4. Explain how the data will be documented
5. Describe how data quality will be assured
6. Present a sound data storage and preservation strategy

10 Simple Rules for Creating a Good DMP

1. Determine the research sponsor requirements
2. Identify the data to be collected
3. Definte how the data will be organized
4. Explain how the data will be documented
5. Describe how data quality will be assured
6. Present a sound data storage and preservation strategy
7. Define the project's data policies

10 Simple Rules for Creating a Good DMP


PLOS | COMPUTATIONAL BIOLOGY

https://doi.org/10.1371/journal.pcbi.1004525

PERSPECTIVE

Ten Simple Rules for Creating a Good Data Management Plan

William K. Michener*

1. Determine the research sponsor requirements
2. Identify the data to be collected
3. Definte how the data will be organized
4. Explain how the data will be documented
5. Describe how data quality will be assured
6. Present a sound data storage and preservation strategy
7. Define the project's data policies
8. Describe how the data will be disseminated

10 Simple Rules for Creating a Good DMP


PLOS | COMPUTATIONAL BIOLOGY

https://doi.org/10.1371/journal.pcbi.1004525

PERSPECTIVE

Ten Simple Rules for Creating a Good Data Management Plan

William K. Michener*

1. Determine the research sponsor requirements
2. Identify the data to be collected
3. Definte how the data will be organized
4. Explain how the data will be documented
5. Describe how data quality will be assured
6. Present a sound data storage and preservation strategy
7. Define the project's data policies
8. Describe how the data will be disseminated
9. Assign roles and responsibilities

10 Simple Rules for Creating a Good DMP


PLOS | COMPUTATIONAL BIOLOGY

https://doi.org/10.1371/journal.pcbi.1004525

PERSPECTIVE

Ten Simple Rules for Creating a Good Data Management Plan

William K. Michener*

1. Determine the research sponsor requirements
2. Identify the data to be collected
3. Definte how the data will be organized
4. Explain how the data will be documented
5. Describe how data quality will be assured
6. Present a sound data storage and preservation strategy
7. Define the project's data policies
8. Describe how the data will be disseminated
9. Assign roles and responsibilities
10. Prepare a realistic budget

Assignment: Write a DMP

Due: April 6, 2022

Write a DMP for your research based on the NSF requirements.

LaTeX for Typesetting

The intention of this lesson is mainly to give you an introduction to LaTeX so that you understand the benefits of using a typesetting system for creating documents.

What is LaTeX?

Pronunciation

LaTeX is pronounced "LAH-tekh" or "LAH-tek" or "LAY-tek" (*i.e.*, don't say the 'x')

What is LaTeX?

Pronunciation

LaTeX is pronounced "LAH-tekh" or "LAH-tek" or "LAY-tek" (*i.e.*, don't say the 'x')

TeX

In 1978 Donald Knuth introduced a typesetting system called TeX that provides "anyone" the ability to produce high-quality typeset documents (like books).

What is LaTeX?

Pronunciation

LaTeX is pronounced "LAH-tekh" or "LAH-tek" or "LAY-tek" (*i.e.*, don't say the 'x')

TeX

In 1978 Donald Knuth introduced a typesetting system called TeX that provides "anyone" the ability to produce high-quality typeset documents (like books).

LaTeX

In 1985 Leslie Lamport released LaTeX, which is a set of macros for document preparation that uses the TeX typesetting program and language.

What are the Advantages of LaTeX?

- it's FREE!

What are the Advantages of LaTeX?

- it's FREE!
- very flexible and creates gorgeous documents

What are the Advantages of LaTeX?

- it's FREE!
- very flexible and creates gorgeous documents
- automation of many tasks like bibliographies and cross-referencing

What are the Advantages of LaTeX?

- it's FREE!
- very flexible and creates gorgeous documents
- automation of many tasks like bibliographies and cross-referencing
- major document format changes can be done by changing a single line (or with generally minimal effort)

What are the Advantages of LaTeX?

- it's FREE!
- very flexible and creates gorgeous documents
- automation of many tasks like bibliographies and cross-referencing
- major document format changes can be done by changing a single line (or with generally minimal effort)
- based on portable files (plain text) making version control, sharing, and collaboration very easy

What are the Advantages of LaTeX?

- it's FREE!
- very flexible and creates gorgeous documents
- automation of many tasks like bibliographies and cross-referencing
- major document format changes can be done by changing a single line (or with generally minimal effort)
- based on portable files (plain text) making version control, sharing, and collaboration very easy
- beautiful rendering of math and graphics

What are the Advantages of LaTeX?

- it's FREE!
- very flexible and creates gorgeous documents
- automation of many tasks like bibliographies and cross-referencing
- major document format changes can be done by changing a single line (or with generally minimal effort)
- based on portable files (plain text) making version control, sharing, and collaboration very easy
- beautiful rendering of math and graphics
- excellent online editors

What are the Advantages of LaTeX?

- it's FREE!
- very flexible and creates gorgeous documents
- automation of many tasks like bibliographies and cross-referencing
- major document format changes can be done by changing a single line (or with generally minimal effort)
- based on portable files (plain text) making version control, sharing, and collaboration very easy
- beautiful rendering of math and graphics
- excellent online editors
- you can make modular files that compile into a single PDF

What are the Advantages of LaTeX?

- it's FREE!
- very flexible and creates gorgeous documents
- automation of many tasks like bibliographies and cross-referencing
- major document format changes can be done by changing a single line (or with generally minimal effort)
- based on portable files (plain text) making version control, sharing, and collaboration very easy
- beautiful rendering of math and graphics
- excellent online editors
- you can make modular files that compile into a single PDF
- you can comment out sections or include based on conditional statements

What are the Advantages of LaTeX?

- it's FREE!
- very flexible and creates gorgeous documents
- automation of many tasks like bibliographies and cross-referencing
- major document format changes can be done by changing a single line (or with generally minimal effort)
- based on portable files (plain text) making version control, sharing, and collaboration very easy
- beautiful rendering of math and graphics
- excellent online editors
- you can make modular files that compile into a single PDF
- you can comment out sections or include based on conditional statements
- etc. (I could write much more here)

What are the Advantages of a WYSIWYG?

What you see is what you get (WYSIWYG)

Word processing tools like Microsoft Word or Mac OS X Pages or LibreOffice Writer or Google Docs are all excellent tools for creating documents.

What are the Advantages of a WYSIWYG?

What you see is what you get (WYSIWYG)

Word processing tools like Microsoft Word or Mac OS X Pages or LibreOffice Writer or Google Docs are all excellent tools for creating documents.

- editing is intuitive and easy to start right away

What are the Advantages of a WYSIWYG?

What you see is what you get (WYSIWYG)

Word processing tools like Microsoft Word or Mac OS X Pages or LibreOffice Writer or Google Docs are all excellent tools for creating documents.

- editing is intuitive and easy to start right away
- almost everyone has Word installed & knows how to use it

What are the Advantages of a WYSIWYG?

What you see is what you get (WYSIWYG)

Word processing tools like Microsoft Word or Mac OS X Pages or LibreOffice Writer or Google Docs are all excellent tools for creating documents.

- editing is intuitive and easy to start right away
- almost everyone has Word installed & knows how to use it
- you are required to use it for many documents as an academic

What are the Disadvantages of a WYSIWYG?

- many are costly

What are the Disadvantages of a WYSIWYG?

- many are costly
- formatting is often mysterious and challenging to get exactly right

What are the Disadvantages of a WYSIWYG?

- many are costly
- formatting is often mysterious and challenging to get exactly right
- they are TERRIBLE for math

What are the Disadvantages of a WYSIWYG?

- many are costly
- formatting is often mysterious and challenging to get exactly right
- they are TERRIBLE for math
- files are difficult to manage using version control, thus collaboration is sometimes clunky (requiring versioned copies of files, tracking changes)

What are the Disadvantages of a WYSIWYG?

- many are costly
- formatting is often mysterious and challenging to get exactly right
- they are TERRIBLE for math
- files are difficult to manage using version control, thus collaboration is sometimes clunky (requiring versioned copies of files, tracking changes)
- bibliography management is nearly impossible without second-party software

What are the Disadvantages of a WYSIWYG?

- many are costly
- formatting is often mysterious and challenging to get exactly right
- they are TERRIBLE for math
- files are difficult to manage using version control, thus collaboration is sometimes clunky (requiring versioned copies of files, tracking changes)
- bibliography management is nearly impossible without second-party software
- figure and other references within a single document are difficult

What are the Disadvantages of a WYSIWYG?

- many are costly
- formatting is often mysterious and challenging to get exactly right
- they are TERRIBLE for math
- files are difficult to manage using version control, thus collaboration is sometimes clunky (requiring versioned copies of files, tracking changes)
- bibliography management is nearly impossible without second-party software
- figure and other references within a single document are difficult
- default autoformatting

What are the Disadvantages of a WYSIWYG?

- many are costly
- formatting is often mysterious and challenging to get exactly right
- they are TERRIBLE for math
- files are difficult to manage using version control, thus collaboration is sometimes clunky (requiring versioned copies of files, tracking changes)
- bibliography management is nearly impossible without second-party software
- figure and other references within a single document are difficult
- default autoformatting
- often not backward/forward compatible

What are the Disadvantages of a WYSIWYG?

- many are costly
- formatting is often mysterious and challenging to get exactly right
- they are TERRIBLE for math
- files are difficult to manage using version control, thus collaboration is sometimes clunky (requiring versioned copies of files, tracking changes)
- bibliography management is nearly impossible without second-party software
- figure and other references within a single document are difficult
- default autoformatting
- often not backward/forward compatible
- troubleshooting is difficult

What are the Disadvantages of a WYSIWYG?

- many are costly
- formatting is often mysterious and challenging to get exactly right
- they are TERRIBLE for math
- files are difficult to manage using version control, thus collaboration is sometimes clunky (requiring versioned copies of files, tracking changes)
- bibliography management is nearly impossible without second-party software
- figure and other references within a single document are difficult
- default autoformatting
- often not backward/forward compatible
- troubleshooting is difficult
- etc. (I could write much more here)

What are the Disadvantages of LaTeX?

- there is a substantial learning curve

What are the Disadvantages of LaTeX?

- there is a substantial learning curve
- it is challenging to collaborate with people who don't know how to use it

What are the Disadvantages of LaTeX?

- there is a substantial learning curve
- it is challenging to collaborate with people who don't know how to use it
- sometimes error messages are cryptic

What are the Disadvantages of LaTeX?

- there is a substantial learning curve
- it is challenging to collaborate with people who don't know how to use it
- sometimes error messages are cryptic
- it can take some effort to figure out how to do something you've never done before (*e.g.*, create a numbered list that goes in reverse order, which is impossible to do in Word, by the way)

What are the Disadvantages of LaTeX?

- there is a substantial learning curve
- it is challenging to collaborate with people who don't know how to use it
- sometimes error messages are cryptic
- it can take some effort to figure out how to do something you've never done before (*e.g.*, create a numbered list that goes in reverse order, which is impossible to do in Word, by the way)
- if your file is very large and contains lots of images (like a long presentation) it can take several seconds to compile

What are the Disadvantages of LaTeX?

- there is a substantial learning curve
- it is challenging to collaborate with people who don't know how to use it
- sometimes error messages are cryptic
- it can take some effort to figure out how to do something you've never done before (*e.g.*, create a numbered list that goes in reverse order, which is impossible to do in Word, by the way)
- if your file is very large and contains lots of images (like a long presentation) it can take several seconds to compile
- it is not for everyone ☹

How do Researchers Use LaTeX

Some Examples

LaTeX is useful for many types of documents including journal articles, grant proposals, and presentations (like these slides!). Let's look at a few of these.

- a scientific manuscript
- a CV
- a presentation

Thesis Template

Iowa State has a LaTeX template for writing your graduate thesis.

LaTeX Tools

Download LaTeX

https://www.latex-project.org/get

LaTeX Interactive Development Environments

TeX distributions often come with free LaTeX IDEs that allow you to compose and compile documents. There are also many nice 3rd party products (though many are not free).

- LyX (offers a WYSIWYMean environment)
- VS Code with LaTeX Workshop
- TeXMaker
- TeXPad (Mac OS X)

Online LaTeX Tools

Web-based LaTeX editors

In the last several years, collaborating online with LaTeX has gotten easier and easier with web-tools.

- Overleaf (https://www.overleaf.com)
- LaTeX Base (https://latexbase.com)
- Papeeria (https://papeeria.com)
- Authorea (https://www.authorea.com)

Let's get started!