

Naming Things

HARVARD LONGWOOD MEDICAL AREA
RESEARCH DATA MANAGEMENT WORKING GROUP



We can help!

- Digital Repositories & Manuscript Submission
- NIH Public Access Policy Compliance
- Copyright & Fair Use
- Author Profiles & CV
- Research Data Management & Data Sharing
- Online Collaboration Tools
- Bioinformatics
- Research Assessment & Impact Metrics
- Scholarly Publishing & Manuscript Preparation

Publishing & Data Services Team



Scott Lapinski, MS

Associate Director, Scholarly
Communication and Open Science
Initiatives

Email

[617-432-6784](mailto:slapinski@countway.harvard.edu)



Reddy Gali, PhD

Bioinformatics Support Specialist

Email

[617-432-7471](mailto:rgali@countway.harvard.edu)



Julie Goldman, MLIS

Research Data Services Librarian

Email



Iris Jahng, MLIS, MA

Digital Scholarship Librarian

Email



Ashley Thomas, MLIS, MA

RDMLA Librarian

Email

Bookmark this website!

<https://datamanagement.hms.harvard.edu>

The screenshot shows the homepage of the Harvard Biomedical Data Management website. At the top, the title "Harvard Biomedical Data Management" is displayed in a serif font, with the subtitle "Best practices & support services for research data lifecycles" below it. A dark red navigation bar contains the following menu items: "About", "Best Practices", "Plan", "Store", "Share", "Resources", and "Support".

The main content area is divided into several sections:

- DATA MANAGEMENT:** A section with a red header. The text below reads: "Data Management is the process of providing the appropriate labeling, storage, and access for data at all stages of a research project. Here you can find best practices, resources, and support services for biomedical research data. Discover the work of the [Data Management Working Group](#)."
- FEATURED RESOURCES:** A section with a red header. It features a large image of a person looking at a laptop screen. Overlaid on the image is the text "Data Management Onboarding Checklist" and a description: "This resource serves as a general, research data management-focused guide to employee/trainee onboarding." Below the image is a navigation bar with left and right arrows and a series of dots.
- SEARCH:** A search bar with a magnifying glass icon.
- Interactions:** Three buttons are visible: "Submit Questions and Feedback", "News & Upcoming Events", and "Subscribe to our Mailing List".
- UPCOMING EVENTS:** A section with a red header. It lists three events:
 - 2019 NOV 15:** Everything you need to know to make your data analysis reproducible
 - 2019 NOV 19:** Introduction to R workshop
 - 2019 NOV 20:** Responsible Conduct of Research (RCR): Research Data ManagementA "More" link with a right-pointing arrow is located at the bottom right of this section.

Learning Objectives

- Understand why naming conventions is essential for data management
- Learn best practices for file naming
- Design naming conventions for your project or group
- Find out who to contact for assistance

Based on the Data Carpentry Lesson File Organization: Naming

<https://datacarpentry.org/rr-organization1/01-file-naming/index.html>

"**Research data management** concerns the organization of data, from its entry to the research cycle through the dissemination and archiving of valuable results. It aims to **ensure reliable verification** of results, and permits **new and innovative research** built on existing information."

Why should you care?

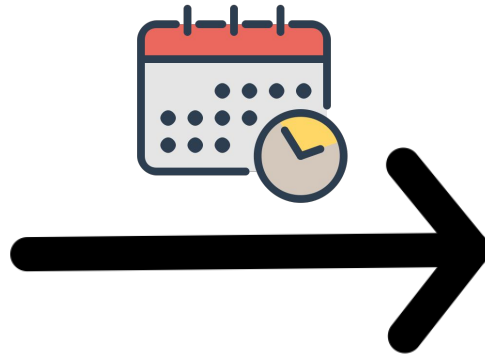
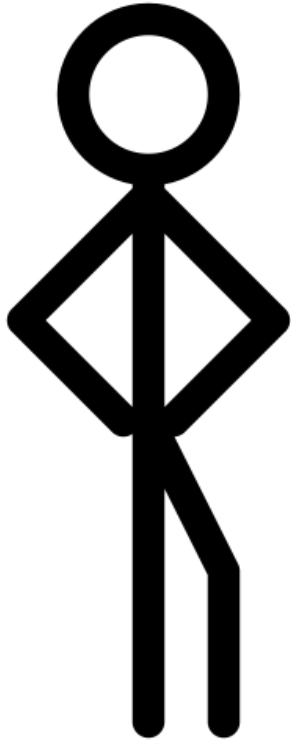
Benefits for yourself, researchers, and science:

- Your future self will thank you!
- Facilitate and ensure seamless team transitions
- Conduct analysis effectively when collaborating with others
- Check and verify research results
- Support FAIR principles: Findable, Accessible, Interoperable, & Reusable

Compliance:

- Be **compliant with University and School policies**
- Be **compliant with funding organizations** that require a data management plan and data sharing
- Be **compliant with journals** that require to submit your data accompanying the article

Naming Things



Three Principles for (File) Names

- Machine readable
- Human readable
- Plays well with default ordering

Examples of really awesome file names



- 2013-06-26_BRAFWTNEGASSAY_Plasmid-Cellline-100-1MutantFraction_H01.csv
- 2013-06-26_BRAFWTNEGASSAY_Plasmid-Cellline-100-1MutantFraction_H02.csv
- 2013-06-26_BRAFWTNEGASSAY_Plasmid-Cellline-100-1MutantFraction_H03.csv
- 2013-06-26_BRAFWTNEGASSAY_Plasmid-Cellline-100-1MutantFraction_platefile.csv
- 2014-02-26_BRAFWTNEGASSAY_FFPEDNA-CRC-1-41_A01.csv
- 2014-02-26_BRAFWTNEGASSAY_FFPEDNA-CRC-1-41_A02.csv
- 2014-02-26_BRAFWTNEGASSAY_FFPEDNA-CRC-1-41_A03.csv
- 2014-02-26_BRAFWTNEGASSAY_FFPEDNA-CRC-1-41_A04.csv

Machine Readable

- Easy to search for files later
- Easy to narrow file lists based on names
- Easy to extract info from file names

Machine Readable

- Regular expression and globbing friendly! Be sure to avoid:
 - spaces in file names
 - punctuation
 - accented characters
 - case sensitivity (“test” vs “Test”)
- Easy to compute on
 - deliberate use of delimiters

Searching "data"

Plasmid

Search: This Mac "data" Shared

Save +

Name	Kind	Last Opened
2013-06-26_BRAFWTNEGASSAY...-100-1MutantFraction_A01.csv	comma-separated values	2014-05-08 8:05 PM
2013-06-26_BRAFWTNEGASSAY...-100-1MutantFraction_A02.csv	comma-separated values	2014-05-08 8:05 PM
2013-06-26_BRAFWTNEGASSAY...-100-1MutantFraction_A03.csv	comma-separated values	2014-05-08 8:05 PM
2013-06-26_BRAFWTNEGASSAY...e-100-1MutantFraction_B01.csv	comma-separated values	2014-05-08 8:05 PM
2013-06-26_BRAFWTNEGASSAY...e-100-1MutantFraction_B02.csv	comma-separated values	2014-05-08 8:05 PM
2013-06-26_BRAFWTNEGASSAY...e-100-1MutantFraction_B03.csv	comma-separated values	2014-05-08 8:05 PM
2013-06-26_BRAFWTNEGASSAY...-100-1MutantFraction_C01.csv	comma-separated values	2014-05-08 8:05 PM
2013-06-26_BRAFWTNEGASSAY...-100-1MutantFraction_C02.csv	comma-separated values	2014-05-08 8:05 PM
2013-06-26_BRAFWTNEGASSAY...-100-1MutantFraction_C03.csv	comma-separated values	2014-05-08 8:05 PM
2013-06-26_BRAFWTNEGASSAY...-100-1MutantFraction_D01.csv	comma-separated values	2014-05-08 8:05 PM
2013-06-26_BRAFWTNEGASSAY...-100-1MutantFraction_D02.csv	comma-separated values	2014-05-08 8:05 PM
2013-06-26_BRAFWTNEGASSAY...-100-1MutantFraction_D03.csv	comma-separated values	2014-05-08 8:05 PM
2013-06-26_BRAFWTNEGASSAY...e-100-1MutantFraction_E01.csv	comma-separated values	2014-05-08 8:05 PM
2013-06-26_BRAFWTNEGASSAY...e-100-1MutantFraction_E02.csv	comma-separated values	2014-05-08 8:05 PM
2013-06-26_BRAFWTNEGASSAY...e-100-1MutantFraction_E03.csv	comma-separated values	2014-05-08 8:05 PM
2013-06-26_BRAFWTNEGASSAY...e-100-1MutantFraction_F01.csv	comma-separated values	2014-05-08 8:05 PM
2013-06-26_BRAFWTNEGASSAY...e-100-1MutantFraction_F02.csv	comma-separated values	2014-05-08 8:05 PM
2013-06-26_BRAFWTNEGASSAY...e-100-1MutantFraction_F03.csv	comma-separated values	2014-05-08 8:05 PM
2013-06-26_BRAFWTNEGASSAY...-100-1MutantFraction_G01.csv	comma-separated values	2014-05-08 8:05 PM
2013-06-26_BRAFWTNEGASSAY...-100-1MutantFraction_G02.csv	comma-separated values	2014-05-08 8:05 PM
2013-06-26_BRAFWTNEGASSAY...-100-1MutantFraction_G03.csv	comma-separated values	2014-05-08 8:05 PM
2013-06-26_BRAFWTNEGASSAY...-100-1MutantFraction_H01.csv	comma-separated values	2014-05-08 8:05 PM
2013-06-26_BRAFWTNEGASSAY...-100-1MutantFraction_H02.csv	comma-separated values	2014-05-08 8:05 PM
2013-06-26_BRAFWTNEGASSAY...-100-1MutantFraction_H03.csv	comma-separated values	2014-05-08 8:05 PM
2013-06-26_BRAFWTNEGASSAY...0-1MutantFraction_platefile.csv	comma-separated values	2014-05-08 3:06 PM

Globbing → in Shell

```
Jennifers-MacBook-Pro-3:2014-03-21 jenny$ ls *Plasmid*
```

```
2013-06-26_BRAFWTNEGASSAY_Plasmid-Cellline-100-1MutantFraction_A01.csv
```

```
2013-06-26_BRAFWTNEGASSAY_Plasmid-Cellline-100-1MutantFraction_A02.csv
```

```
2013-06-26_BRAFWTNEGASSAY_Plasmid-Cellline-100-1MutantFraction_A03.csv
```

```
2013-06-26_BRAFWTNEGASSAY_Plasmid-Cellline-100-1MutantFraction_B01.csv
```

```
...
```

```
2013-06-26_BRAFWTNEGASSAY_Plasmid-Cellline-100-1MutantFraction_H03.csv
```

```
2013-06-26_BRAFWTNEGASSAY_Plasmid-Cellline-100-1MutantFraction_plate.csv
```

Regular Expressions → in R

```
> list.files(pattern = "Plasmid") %>% head
```

```
[1] "2013-06-26_BRAFWTNEGASSAY_Plasmid-Cellline-100-1MutantFraction_A01.csv"  
[2] "2013-06-26_BRAFWTNEGASSAY_Plasmid-Cellline-100-1MutantFraction_A02.csv"  
[3] "2013-06-26_BRAFWTNEGASSAY_Plasmid-Cellline-100-1MutantFraction_A03.csv"  
[4] "2013-06-26_BRAFWTNEGASSAY_Plasmid-Cellline-100-1MutantFraction_B01.csv"  
[5] "2013-06-26_BRAFWTNEGASSAY_Plasmid-Cellline-100-1MutantFraction_B02.csv"  
[6] "2013-06-26_BRAFWTNEGASSAY_Plasmid-Cellline-100-1MutantFraction_B03.csv"
```

Breaking it Down

```
> flist <- list.files(pattern = "Plasmid") %>% head
```

```
> stringr::str_split_fixed(flist, "[_\\.]", 5)
```

	[,1]	[,2]	[,3]	[,4]	[,5]
[1,]	"2013-06-26"	"BRAFWTNEGASSAY"	"Plasmid-Cellline-100-1MutantFraction"	"A01"	"csv"
[2,]	"2013-06-26"	"BRAFWTNEGASSAY"	"Plasmid-Cellline-100-1MutantFraction"	"A02"	"csv"
[3,]	"2013-06-26"	"BRAFWTNEGASSAY"	"Plasmid-Cellline-100-1MutantFraction"	"A03"	"csv"
[4,]	"2013-06-26"	"BRAFWTNEGASSAY"	"Plasmid-Cellline-100-1MutantFraction"	"B01"	"csv"
[5,]	"2013-06-26"	"BRAFWTNEGASSAY"	"Plasmid-Cellline-100-1MutantFraction"	"B02"	"csv"
[6,]	"2013-06-26"	"BRAFWTNEGASSAY"	"Plasmid-Cellline-100-1MutantFraction"	"B03"	"csv"

date

assay

sample set

well

file type

Human Readable

- Name contains information on content
- Connects to concept of a slug from semantic URLs
- Easy to figure out what something is based on its name

Human Readable

```
Jennifers-MacBook-Pro-3:analysis jenny$ ls -l
```

```
01_marshall-data.md          01.md
01_marshall-data.r          01.r
02_pre-dea-filtering.md     02.md
02_pre-dea-filtering.r     02.r
03_dea-with-limma-voom.md   03.md
03_dea-with-limma-voom.r   03.r
04_explore-dea-results.md  04.md
04_explore-dea-results.r  04.r
90_limma-model-term-name-fiasco.md 90.md
90_limma-model-term-name-fiasco.r 90.r
helper01_load-counts.r     helper01.r
Helper02_load-exp-des.r   helper02.r
helper03_load-focus-statinf.r helper03.r
Helper04_extract-and-tidy.r helper04.r
Makefile                  tmp.txt
figure
```



**Which set of
file(name)s
do you want
at 3am before
a deadline?**

Embrace the Slug

01_marshall-data.r

02_pre-dea-filtering.r

03_dea-with-limma-voom.r

04_explore-dea-results.r

90_limma-model-term-name-fiasco.r

helper01_load-counts.r

helper02_load-exp-des.r

helper03_load-focus-statinf.r

helper04_extract-and-tidy.r



**Easy to figure out
what the heck
something is based on
its name**

Plays Well with Default Ordering

- Put something numeric first
- Left pad other numbers with zeros
- Use the ISO 8601 standard for dates

Numbering

01_marshall-data.r

02_pre-dea-filtering.r

03_dea-with-limma-voom.r

04_explore-dea-results.r

90_limma-model-term-name-fiasco.r

helper01_load-counts.r

helper02_load-exp-des.r

helper03_load-focus-statinf.r

helper04_extract-and-tidy.r

If you don't left pad, you get this:

10_final-figs-for-publication.R

1_data-cleaning.R

2_fit-model.R



ISO 8601 Standard




PUBLIC SERVICE ANNOUNCEMENT:

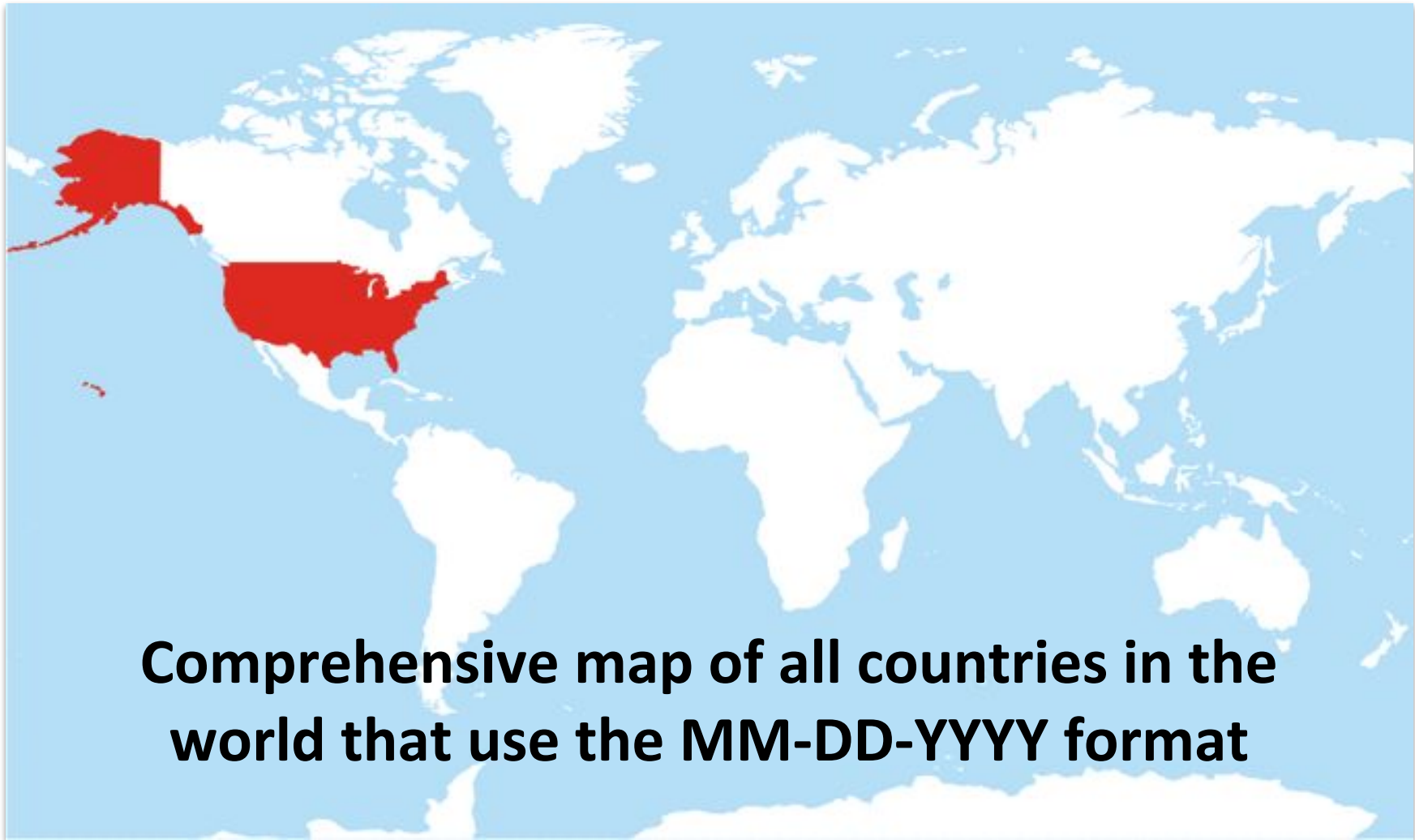
OUR DIFFERENT WAYS OF WRITING DATES AS NUMBERS CAN LEAD TO ONLINE CONFUSION. THAT'S WHY IN 1988 ISO SET A GLOBAL STANDARD NUMERIC DATE FORMAT.

THIS IS *THE* CORRECT WAY TO WRITE NUMERIC DATES:

2013-02-27

THE FOLLOWING FORMATS ARE THEREFORE DISCOURAGED:

02/27/2013 02/27/13 27/02/2013 27/02/13
20130227 2013.02.27 27.02.13 27-02-13
27.2.13 2013. II. 27. 27/2-13 2013.158904109
MMXIII-II-XXVII MMXIII ^{LVII}/_{CCCLXV} 1330300800
 $((3+3) \times (111+1) - 1) \times 3/3 - 1/3^3$ 2013
10/1101/1101 02/27/20/13 $\begin{matrix} 2 & 3 & 1 & 4 \\ 0 & 1 & 2 & 3 & 7 \\ 5 & 6 & 7 & 8 \end{matrix}$ 



Comprehensive map of all countries in the world that use the MM-DD-YYYY format

Renaming Files

Mac:

1. In the Finder, select and highlight the files you want to change using Shift or Command.
2. Right click/Control + click on the selected files and choose “Rename X Items”.
3. Select one of the rename options: swap out text, add a set text, or apply a custom sequential format.

File Renaming Tool: [NameChanger](#)

Windows:

1. In the File Manager, select and highlight the files you want to change using Shift or Control.
2. Right click on the selected files and choose “Rename”. One of the file names will become active for editing.
3. Enter a systematic file name and press enter.
4. All files will be renamed using the chosen file name and numbered sequentially (1) (2), etc.

File Renaming Tool: [Rename It!](#)

More suggestions: <https://datamanagement.hms.harvard.edu/file-naming-conventions>

Takeaways

- Remember the three principles for names: machine readable, human readable, and play well with default ordering.
- Check for established file naming conventions in your discipline or in your group. Naming conventions should be documented so that others in your lab or department can follow this standard.
- Payoffs accumulate as your skills evolve and projects get more complex. File names should be descriptive and provide just enough contextual information.
- Check your instruments, software, or other equipment to see if the data outputs can be set with a file naming system.

Submit Questions

<https://datamanagement.hms.harvard.edu>

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

About ▾ Best Practices ▾ Plan ▾ Store ▾ Share ▾ Resources Support

DATA MANAGEMENT

Data Management is the process of providing the appropriate labeling, storage, and access for data at all stages of a research project. Here you can find best practices, resources, and support services for biomedical research data. Discover the work of the [Data Management Working Group](#).

FEATURED RESOURCES

Data Management Onboarding Checklist
This resource serves as a general, research data management-focused guide to employee/trainee onboarding.

2019 NOV 15 Everything you need to know to make your data analysis reproducible

2019 NOV 19 Introduction to R workshop

2019 NOV 20 Responsible Conduct of Research (RCR): Research Data Management

[More ▶](#)

Submit Questions and Feedback

News & Upcoming Events

Subscribe to our Mailing List

Upcoming Data Management Online Seminars

Version Control Basics

Wednesday, May 20, 2020

1:00pm - 1:30pm

Online - Zoom

bit.ly/RDM-Seminars

Tidy Data is Good Data

Tuesday, May 26, 2020

3:00pm - 3:30pm

Online - Zoom

bit.ly/RDM-Seminars

Thank you!

Please fill out this survey

bit.ly/rdm-online