

Research Data Management

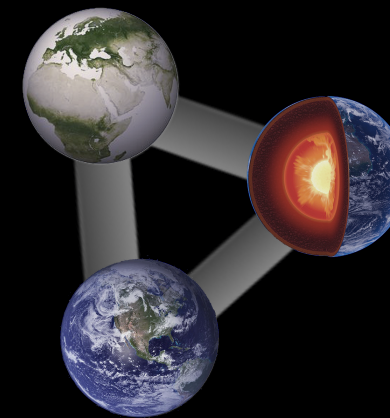
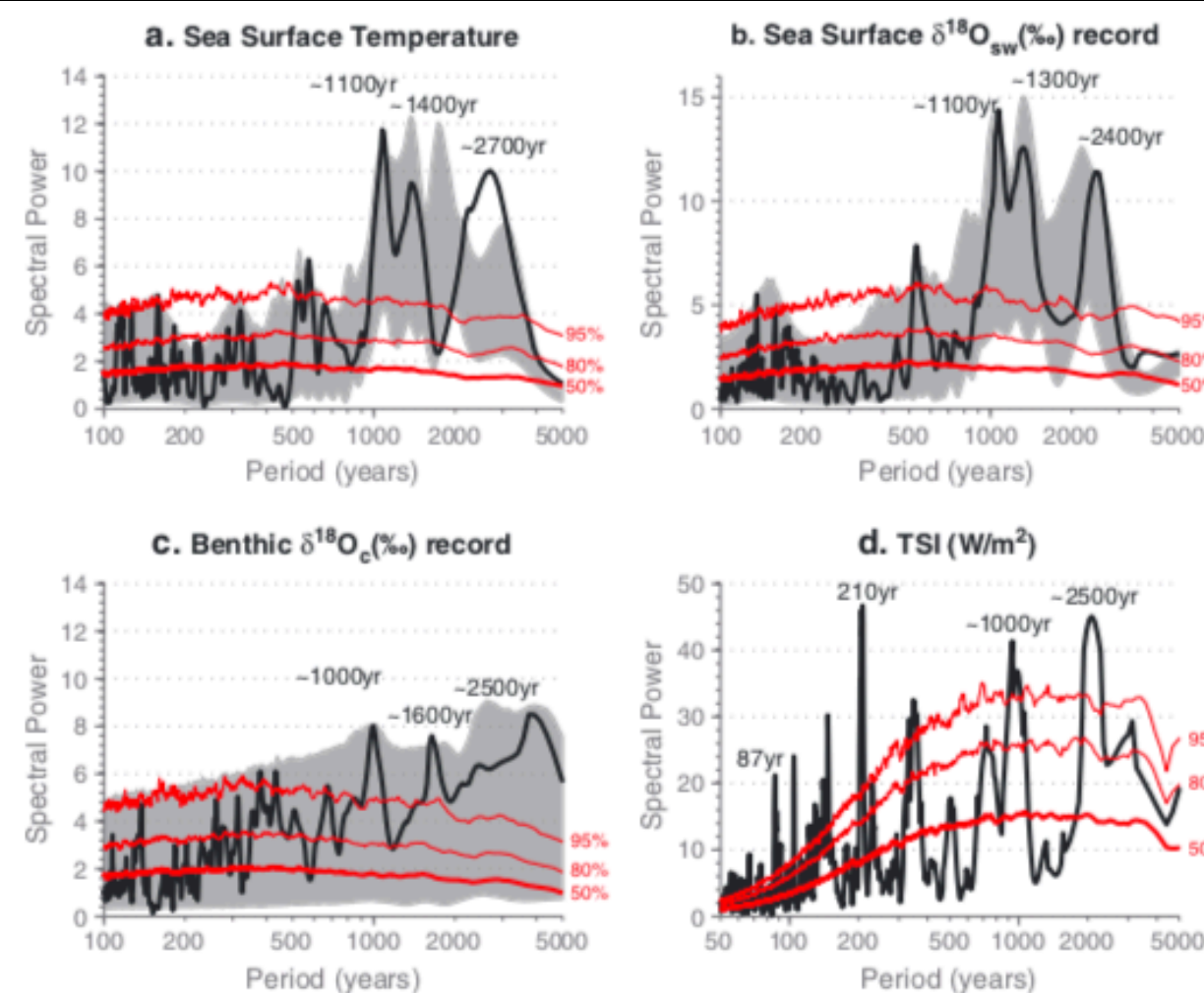
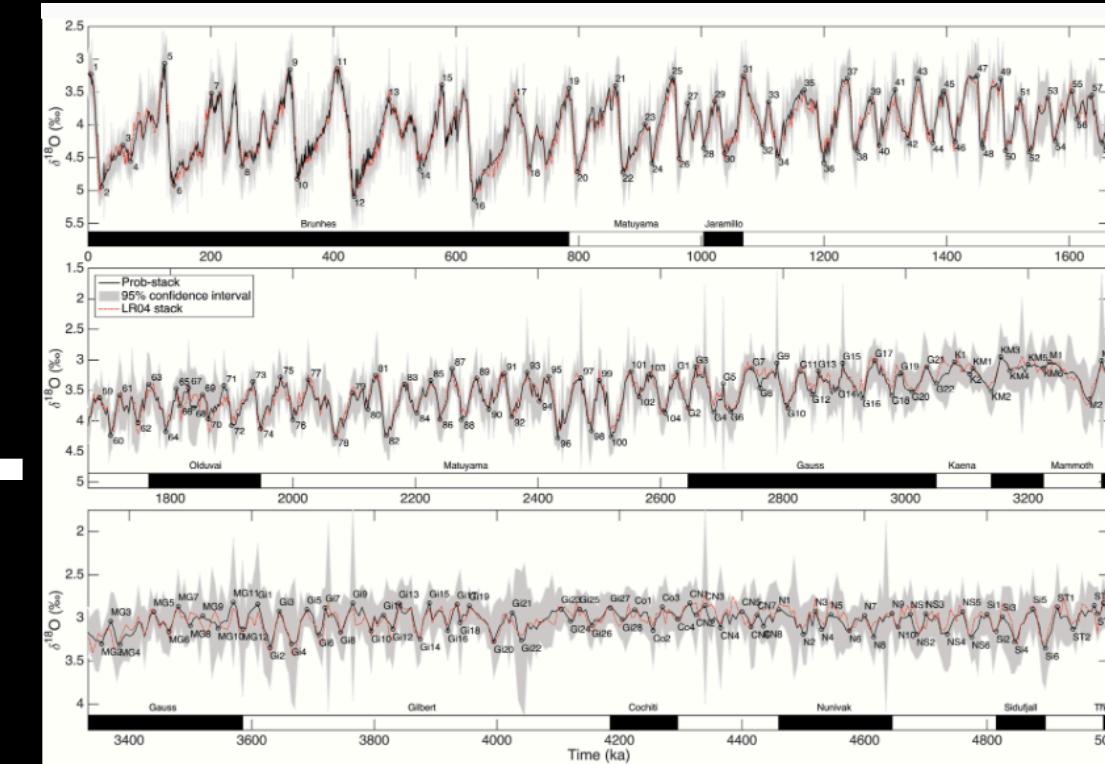
Deborah Khider



USC Viterbi
School of Engineering
Information Sciences Institute

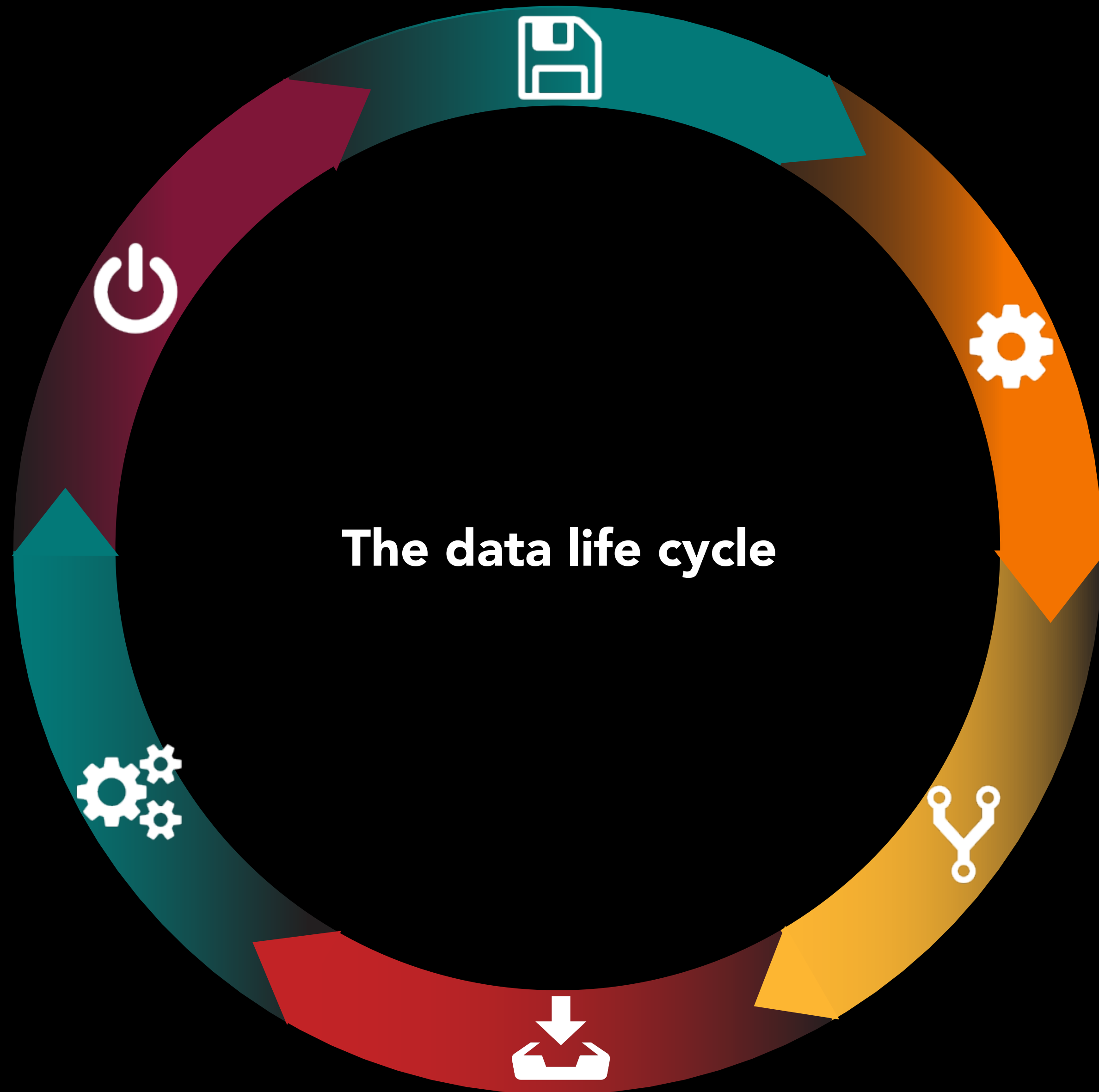


My Data Journey

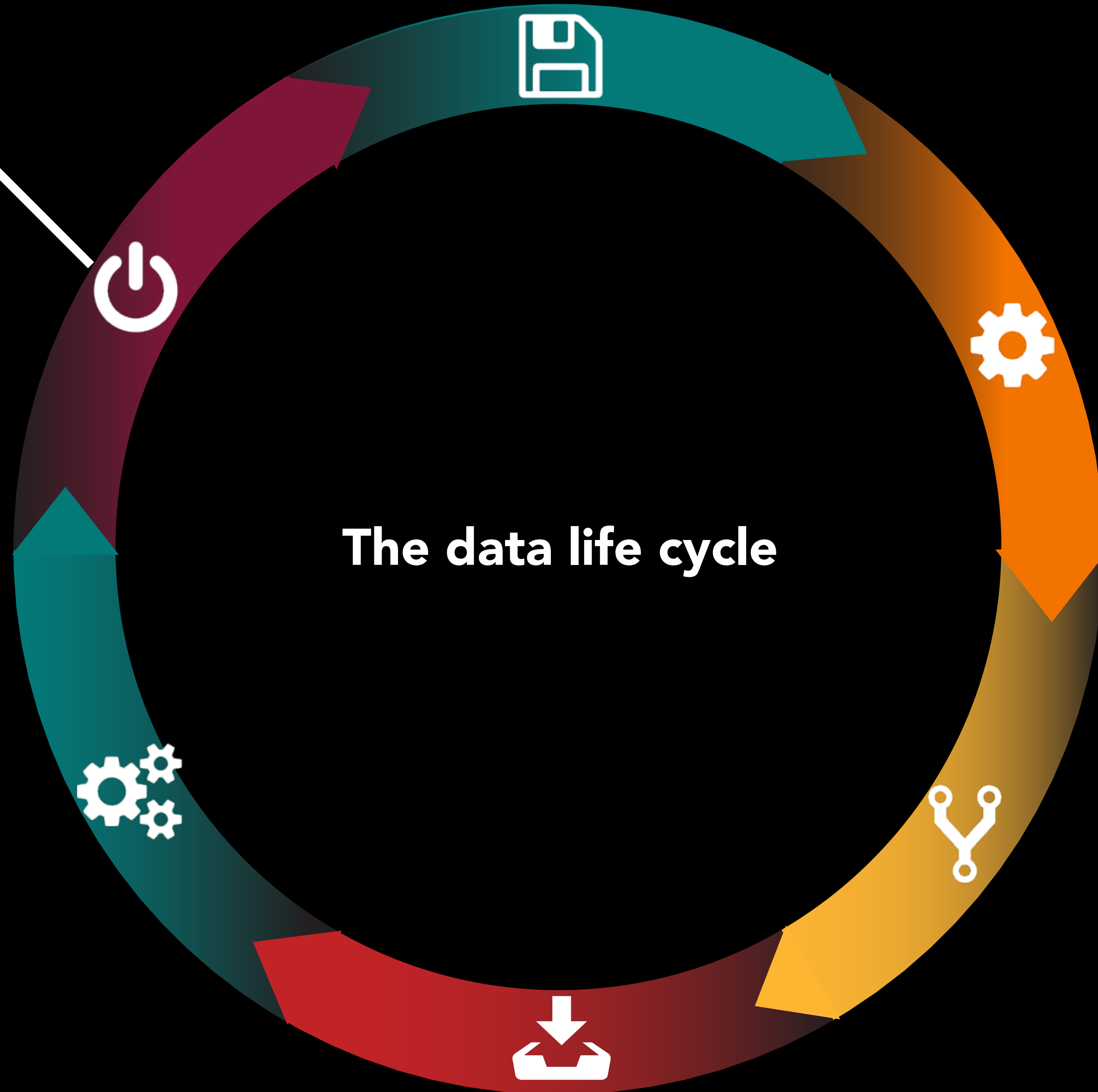


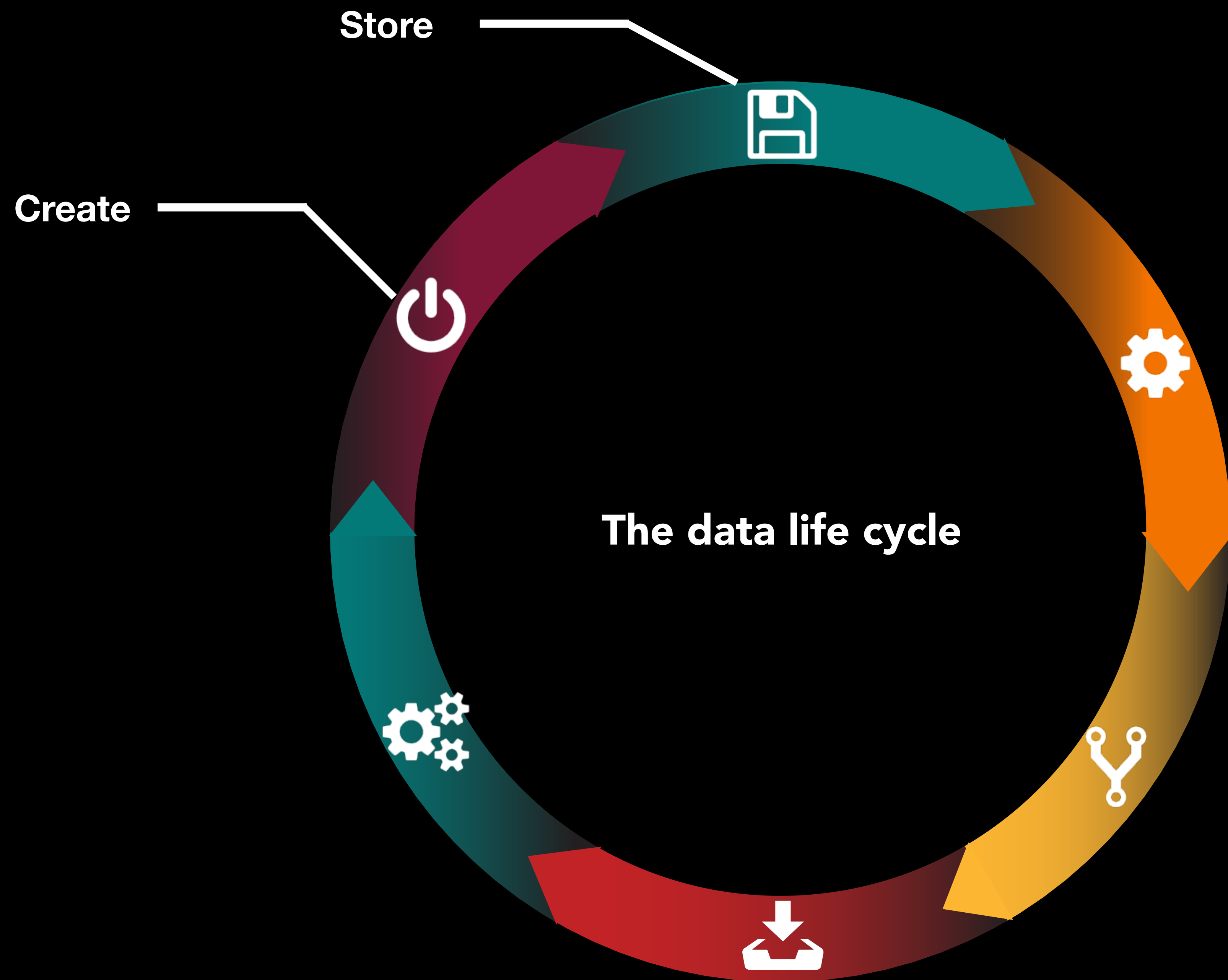
LinkedEarth

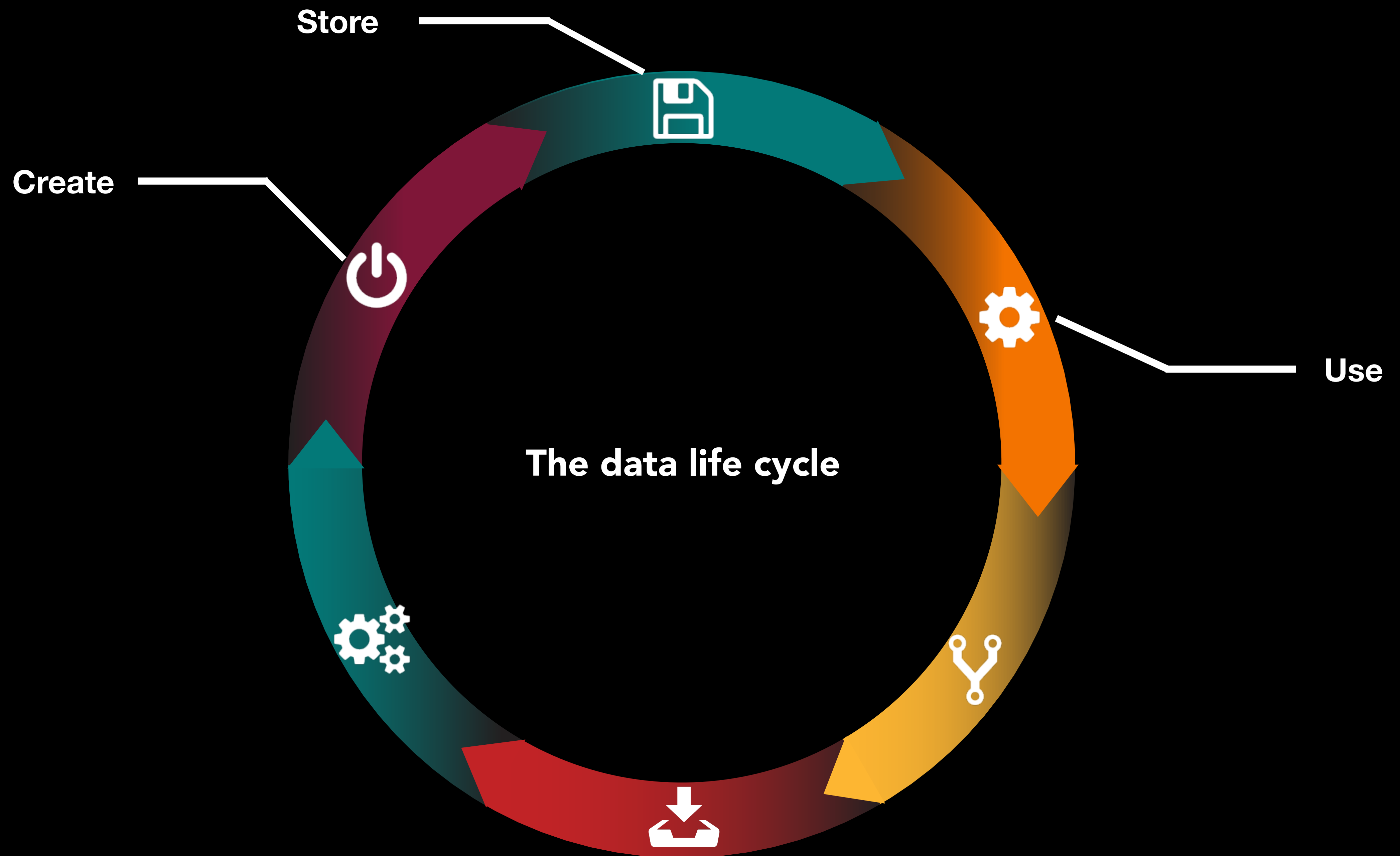


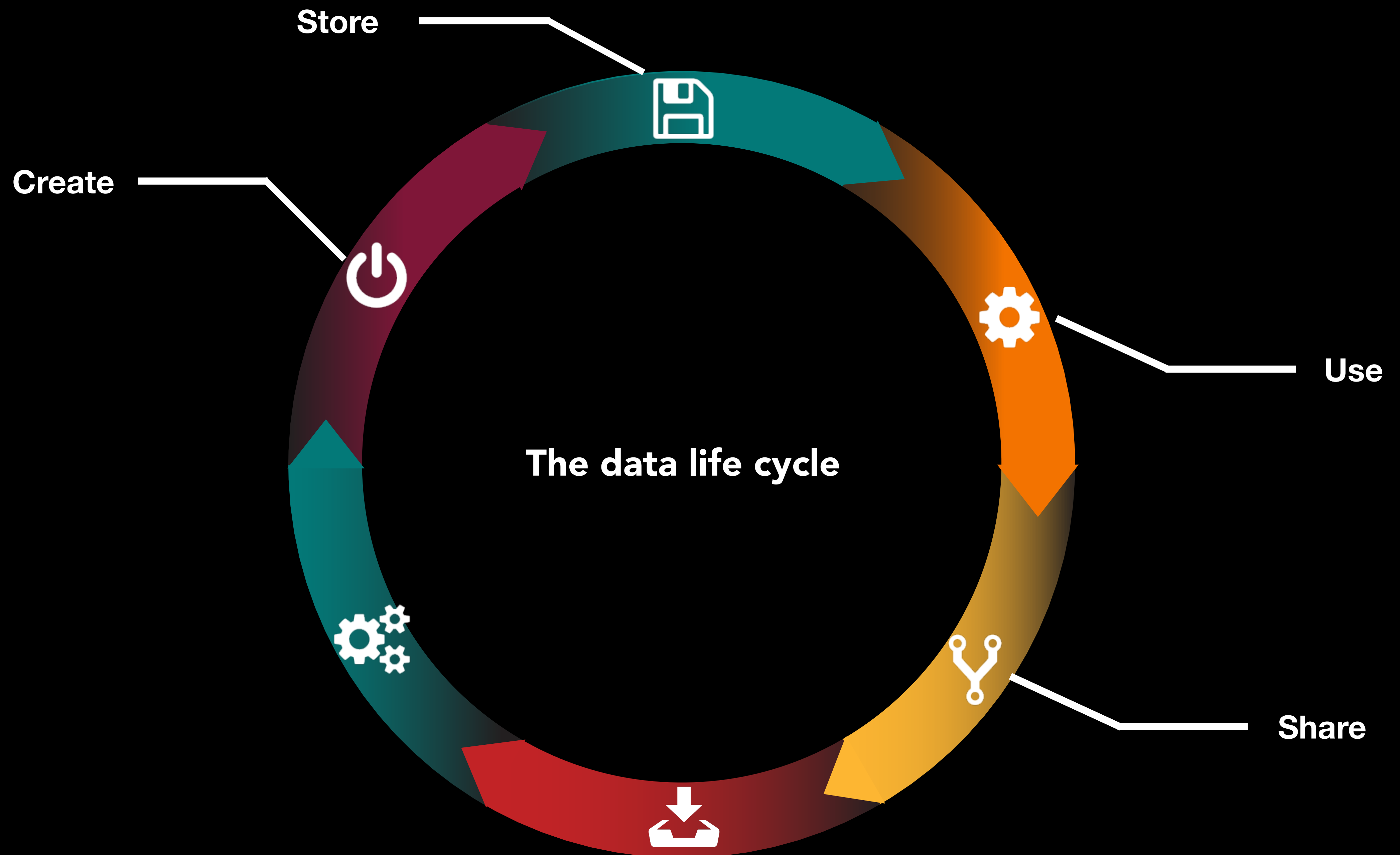


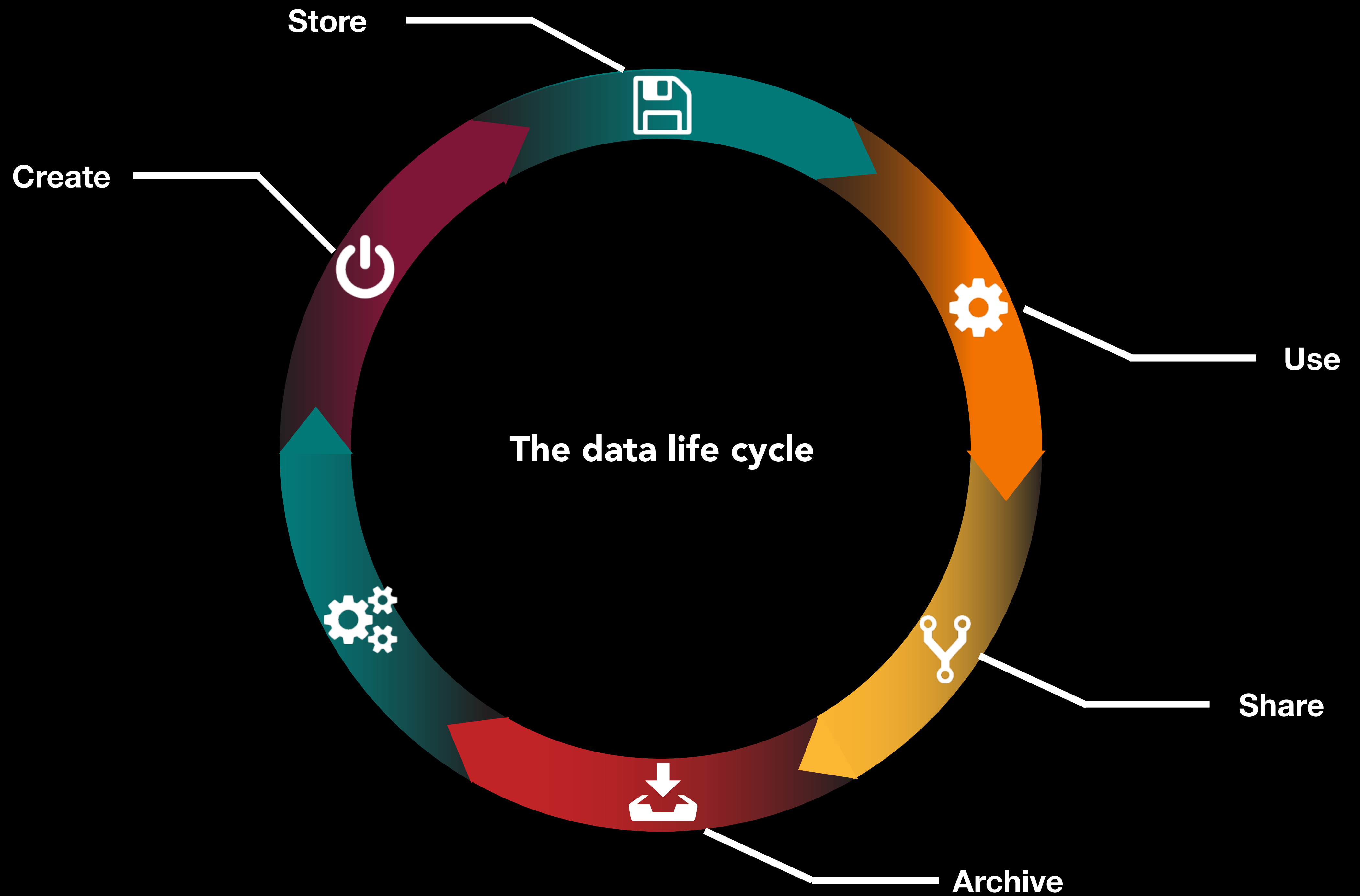
Create

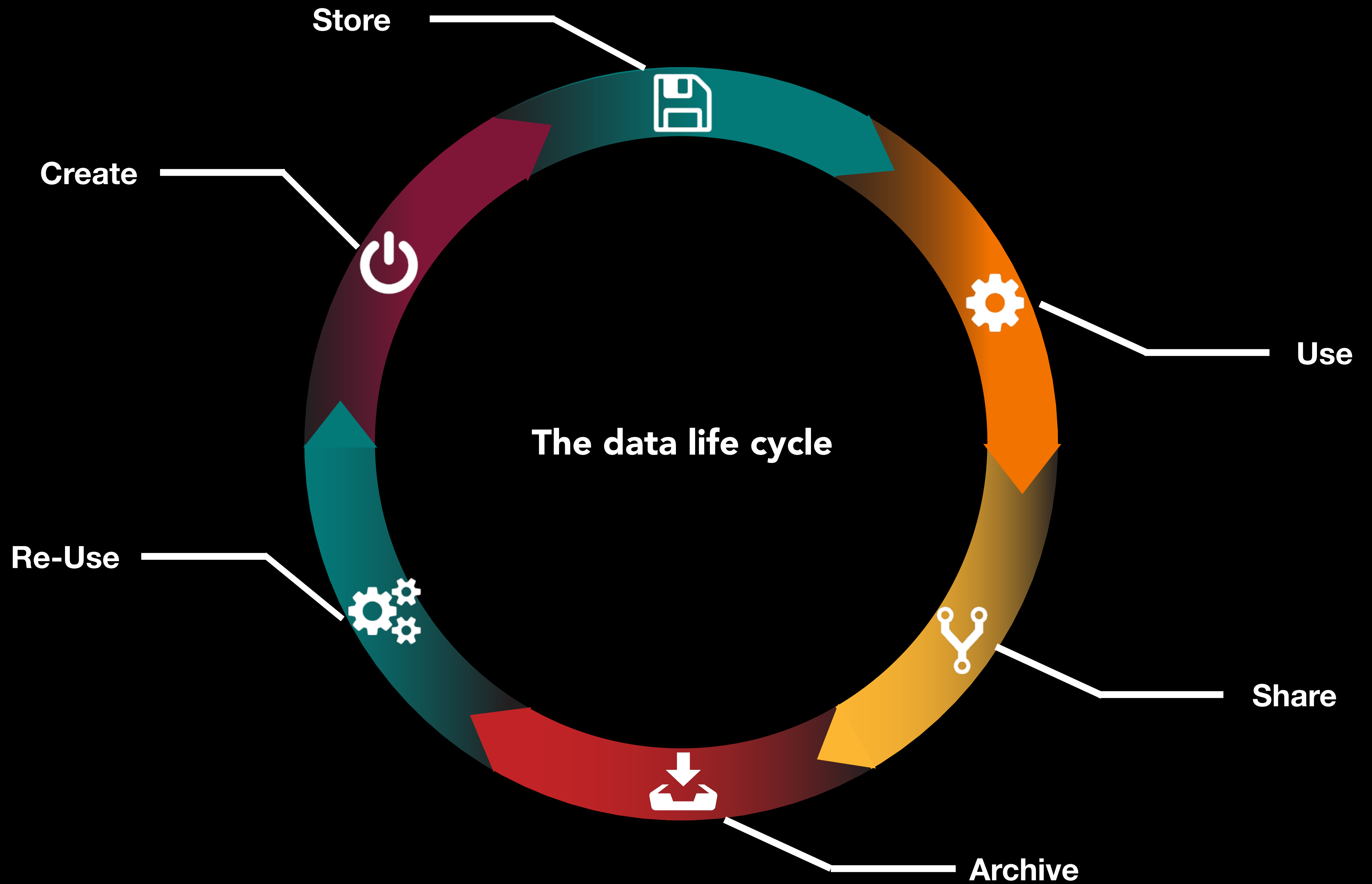










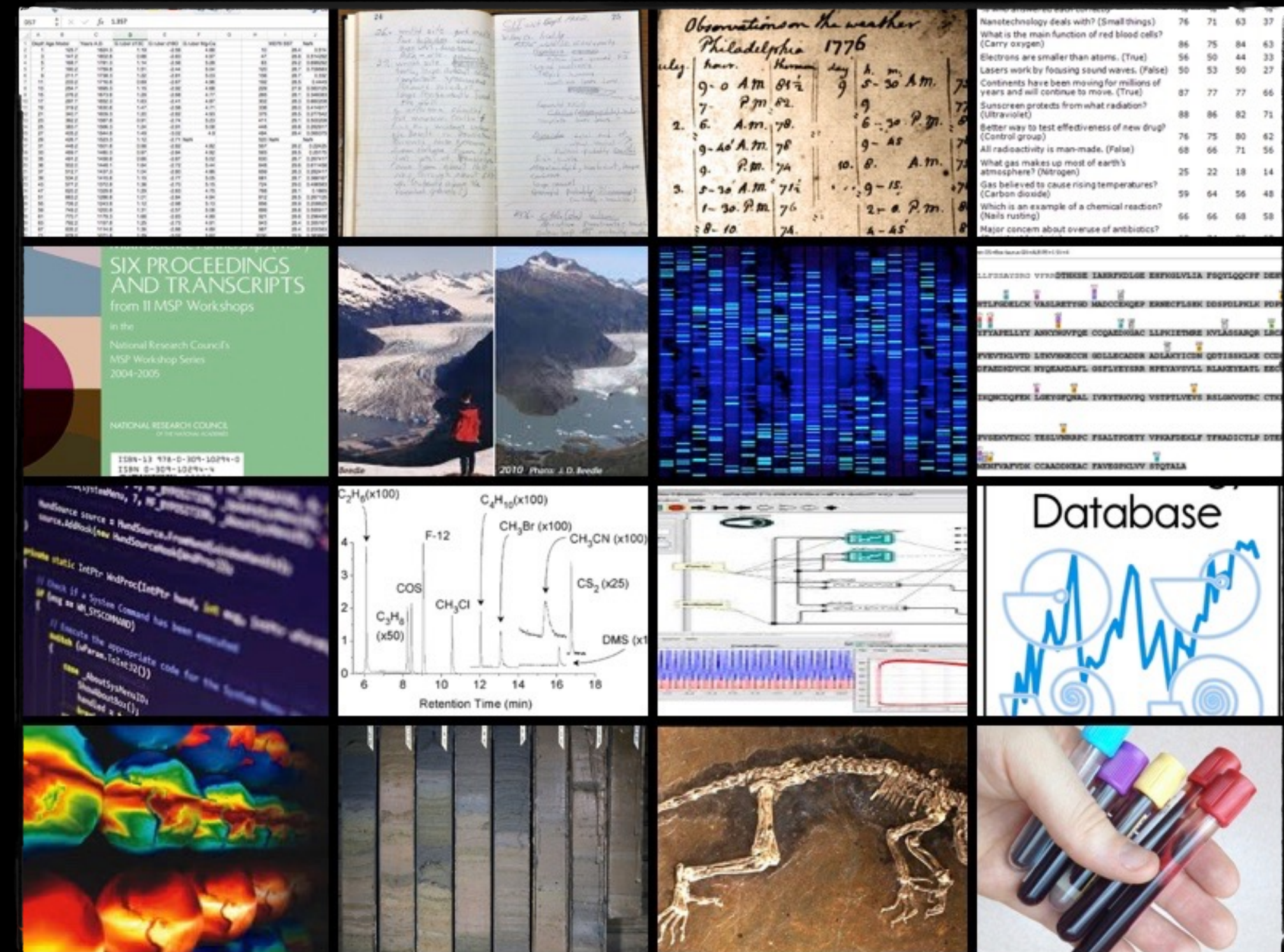


DATA:

"...THE RECORDED FACTUAL
MATERIAL COMMONLY ACCEPTED
IN THE SCIENTIFIC COMMUNITY AS
NECESSARY TO **VALIDATE**
RESEARCH FINDINGS."

DATA:

"...THE RECORDED FACTUAL MATERIAL COMMONLY ACCEPTED IN THE SCIENTIFIC COMMUNITY AS NECESSARY TO **VALIDATE** RESEARCH FINDINGS."



DATA:

"...THE RECORDED FACTUAL MATERIAL COMMONLY ACCEPTED IN THE SCIENTIFIC COMMUNITY AS NECESSARY TO **VALIDATE** RESEARCH FINDINGS."

METADATA:

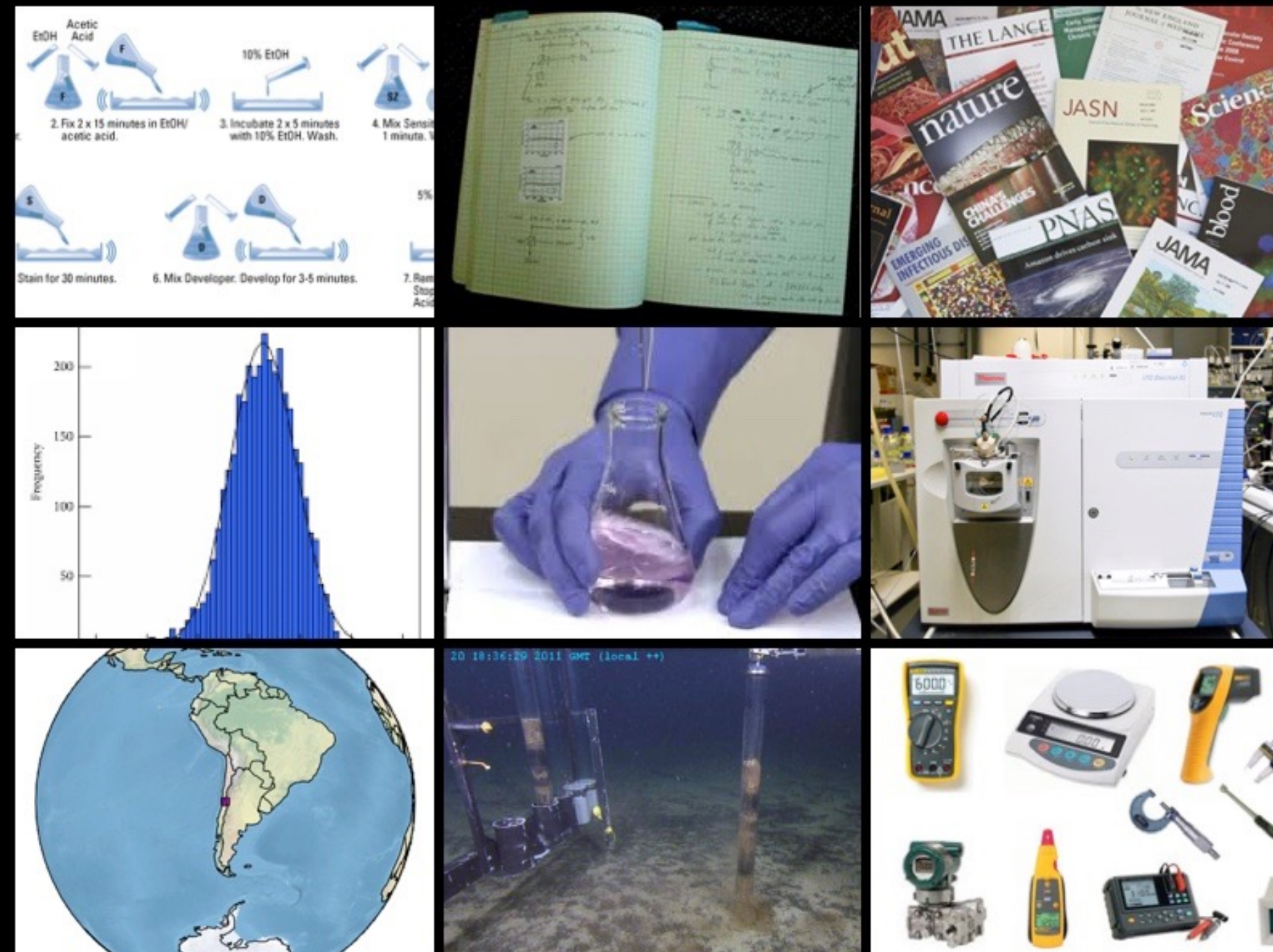
METADATA IS INFORMATION ABOUT THE DATA THAT PROVIDES CONTEXT KEY TO **UNDERSTAND** WHAT THE DATA REPRESENTS.

DATA:

"...THE RECORDED FACTUAL MATERIAL COMMONLY ACCEPTED IN THE SCIENTIFIC COMMUNITY AS NECESSARY TO **VALIDATE** RESEARCH FINDINGS."

METADATA:

METADATA IS INFORMATION ABOUT THE DATA THAT PROVIDES CONTEXT KEY TO **UNDERSTAND** WHAT THE DATA REPRESENTS.



DATA:

"...THE RECORDED FACTUAL MATERIAL COMMONLY ACCEPTED IN THE SCIENTIFIC COMMUNITY AS NECESSARY TO **VALIDATE** RESEARCH FINDINGS."

METADATA:

METADATA IS INFORMATION ABOUT THE DATA THAT PROVIDES CONTEXT KEY TO **UNDERSTAND** WHAT THE DATA REPRESENTS.

DATA MANAGEMENT:

ACTIONS THAT CONTRIBUTE TO EFFECTIVE **STORAGE, PRESERVATION,** AND **REUSE** OF DATA AND METADATA THROUGHOUT THE RESEARCH LIFECYCLE.

Why do you need to know about data management?

- Scientists are changing

Why do you need to know about data management?

- Scientists are changing

Open Data



Open Access



Open Publications

Open Source



Why do you need to know about data management?

- Scientists are changing
- Publishers are changing

Data & Software for Authors

WHAT IS NEEDED?

AGU requires that the underlying data needed to understand, evaluate, and build upon the reported research be available at the time of peer review and publication. Additionally, authors should make available software that has a significant impact on the research. This entails:

1. Depositing the data and software in a trusted repository, as appropriate, and preferably with a DOI
2. Including an [Availability Statement](#) as a separate paragraph in the Open Research section explaining to the reader where and how to access the data and software
3. And including [citation\(s\)](#) to the deposited data and software, in the Reference Section.

Click on the headings below for detailed information on:

- [Models & Simulations](#)
- [Journal-Specific Data Guidance](#)
- [International Geo Sample Numbers](#)

WHAT DATA NEEDS TO BE AVAILABLE?

Primary and processed data used for your research should be preserved and made available.

Generally, the underlying data are considered to be the types of data usually preserved in domain repositories for each discipline. These may include raw data, but are usually the processed or refined data that support and lead to the described results and allow other readers to assess your conclusions and build off your work.

In your paper, cite these data, as well as any data you used from other sources, and include information about access to the data in the availability statement. For model or simulation data, follow [journal specific guidance](#) on prioritizing preserved output; in general, availability of software is most important.

Very large data (greater than 1 terabyte or TB) can be a challenge to preserve as there often fees and additional resources required. One option to consider, institutions often offer solutions for data preservation and compliance. Again, refer to the [journal specific guidance](#) for more information or email DataHelp@agu.org.

Why do you need to know about data management?

- Scientists are changing
- Publishers are changing

Data & Software for Authors



WHAT IS NEEDED?

AGU requires that the underlying data needed to understand, evaluate, and build upon the reported research be available at the time of peer review and publication. Additionally, authors should make available software that has a significant impact on the research. This entails:

1. Depositing the data and software in a trusted repository, as appropriate, and preferably with a DOI
2. Including an [Availability Statement](#) as a separate paragraph in the Open Research section explaining to the reader where and how to access the data and software
3. And including [citation\(s\)](#) to the deposited data and software, in the Reference Section.

Click on the headings below for detailed information on:

- [Models & Simulations](#)
- [Journal-Specific Data Guidance](#)
- [International Geo Sample Numbers](#)

WHAT DATA NEEDS TO BE AVAILABLE?

Primary and processed data used for your research should be preserved and made available.

Generally, the underlying data are considered to be the types of data usually preserved in domain repositories for each discipline. These may include raw data, but are usually the processed or refined data that support and lead to the described results and allow other readers to assess your conclusions and build off your work.

In your paper, cite these data, as well as any data you used from other sources, and include information about access to the data in the availability statement. For **model or simulation data**, follow [journal specific guidance](#) on prioritizing preserved output; in general, availability of software is most important.

Very large data (greater than 1 terabyte or TB) can be a challenge to preserve as there often fees and additional resources required. One option to consider, institutions often offer solutions for data preservation and compliance. Again, refer to the [journal specific guidance](#) for more information or email DataHelp@agu.org.

Why do you need to know about data management?

- Scientists are changing
- Publishers are changing
- Funders are changing




The screenshot shows the top of the White House website. The header includes the text "the WHITE HOUSE PRESIDENT BARACK OBAMA" and links for "Contact Us" and "Get Email Updates". Below this is a dark blue navigation bar with the White House seal and links for "BRIEFING ROOM", "ISSUES", "THE ADMINISTRATION", "PARTICIPATE", and "1600 PENN". A search bar is located on the right side of the navigation bar.

Below the navigation bar, the breadcrumb "HOME · BLOG" is visible. The main heading of the blog post is "Expanding Public Access to the Results of Federally Funded Research". The date and author information are "FEBRUARY 22, 2013 AT 12:04 PM ET BY MICHAEL STEBBINS". There are three social media icons: Twitter, Facebook, and Email.

The summary text reads: "Summary: The Obama Administration is committed to the proposition that citizens deserve easy access to the results of research their tax dollars have paid for. That's why, in a policy memorandum released today, OSTP Director John Holdren has directed Federal agencies with more than \$100M in R&D expenditures to develop plans to make the results of federally funded research freely available to the public—generally within one year of publication."

Why do you need to know about data management?

- Scientists are changing
- Publishers are changing
- Funders are changing
- Universities are changing

 UNIVERSITY OF CALIFORNIA

UC Publication Management

Manage your publications.
Participate in the UC Open Access Policy.
Increase the impact of your work.

Select your campus to get started:

[UC Berkeley >>](#)[UC Riverside >>](#)





[UC Davis >>](#)[UC Santa Barbara >>](#)

[UC Irvine >>](#)[UC Santa Cruz >>](#)

[UCLA >>](#)[UC San Diego >>](#)

[UC Merced >>](#)[UCSF >>](#)

Managing your publications

-  We'll scan the web for publications you've authored.
-  Log in (at left) to review what we've found.
-  Claim publications that are yours; reject those that aren't.
-  Upload your manuscript for public display on [eScholarship](#).

Resources and support

Learn more about the [UC Open Access Policy](#).

Get answers to [Frequently Asked Questions](#).

Find out who to contact for [additional support](#).


Logout notice

To protect your accounts from unauthorized access, please lock your workstation or exit your browser after logging out of this site.

Why do you need to know about data management?

- Sharing your data is not only required but it also helps you!

Sharing Detailed Research Data Is Associated with Increased Citation Rate

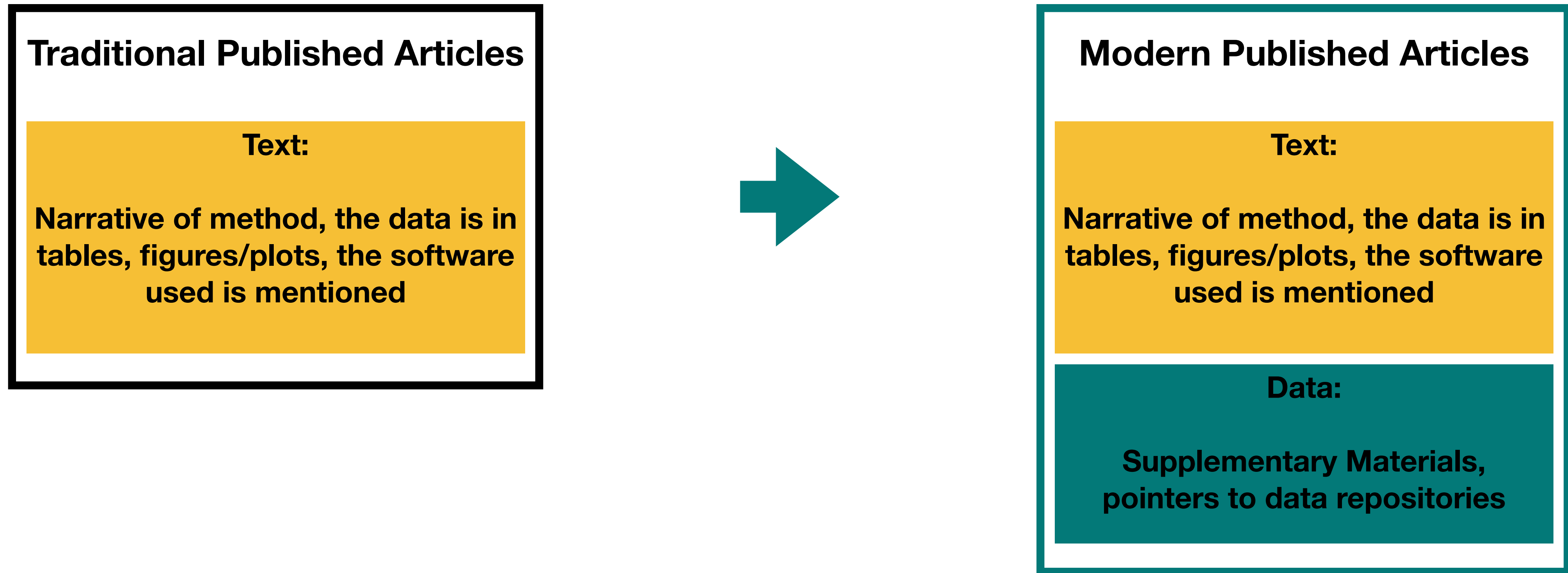
Heather A. Piwowar , Roger S. Day, Douglas B. Fridsma

Published: March 21, 2007 • DOI: 10.1371/journal.pone.0000308 • Featured in PLOS Collections

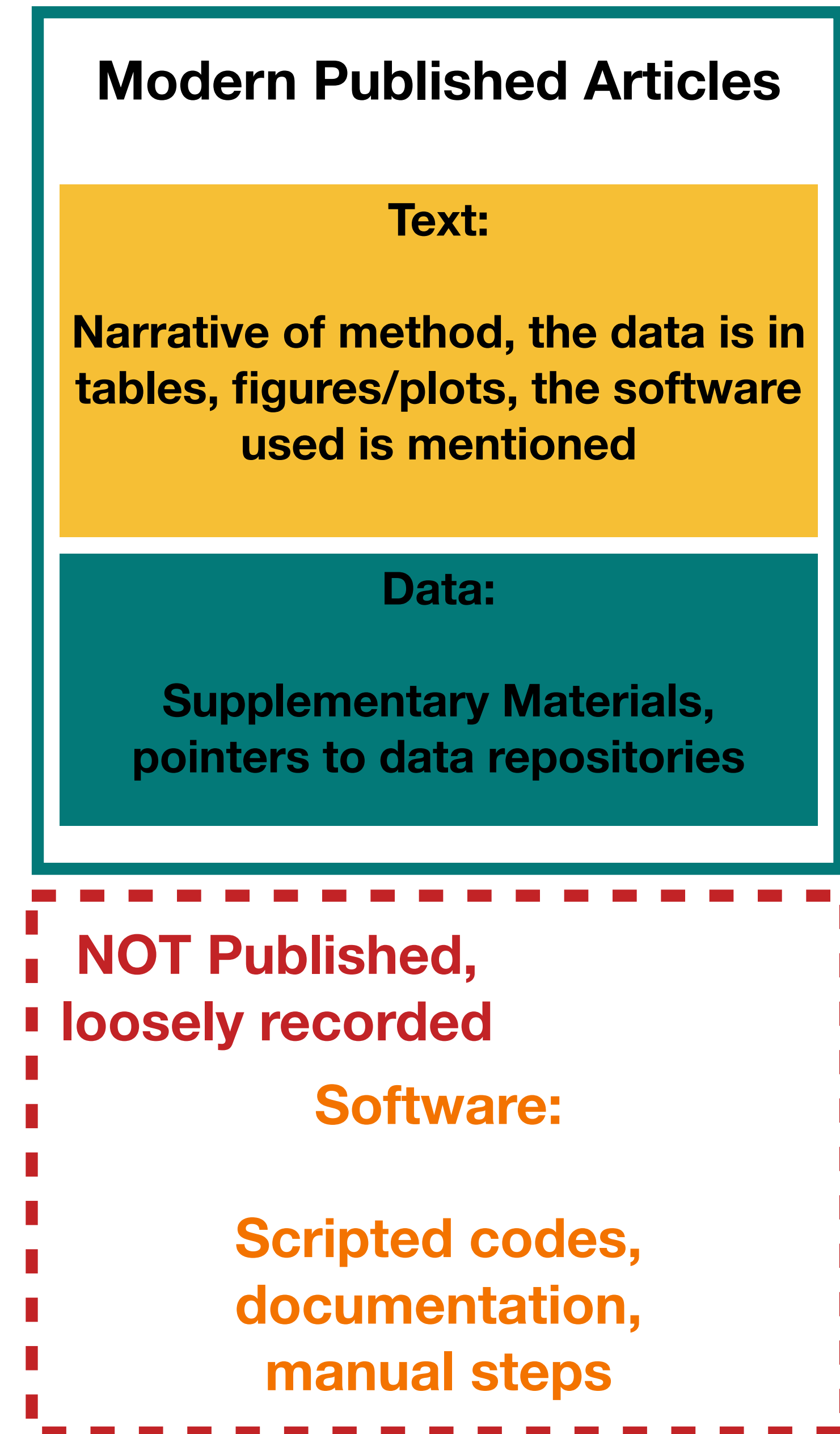
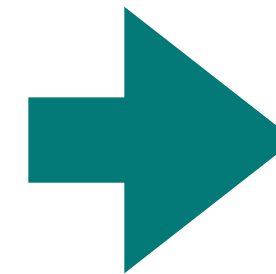
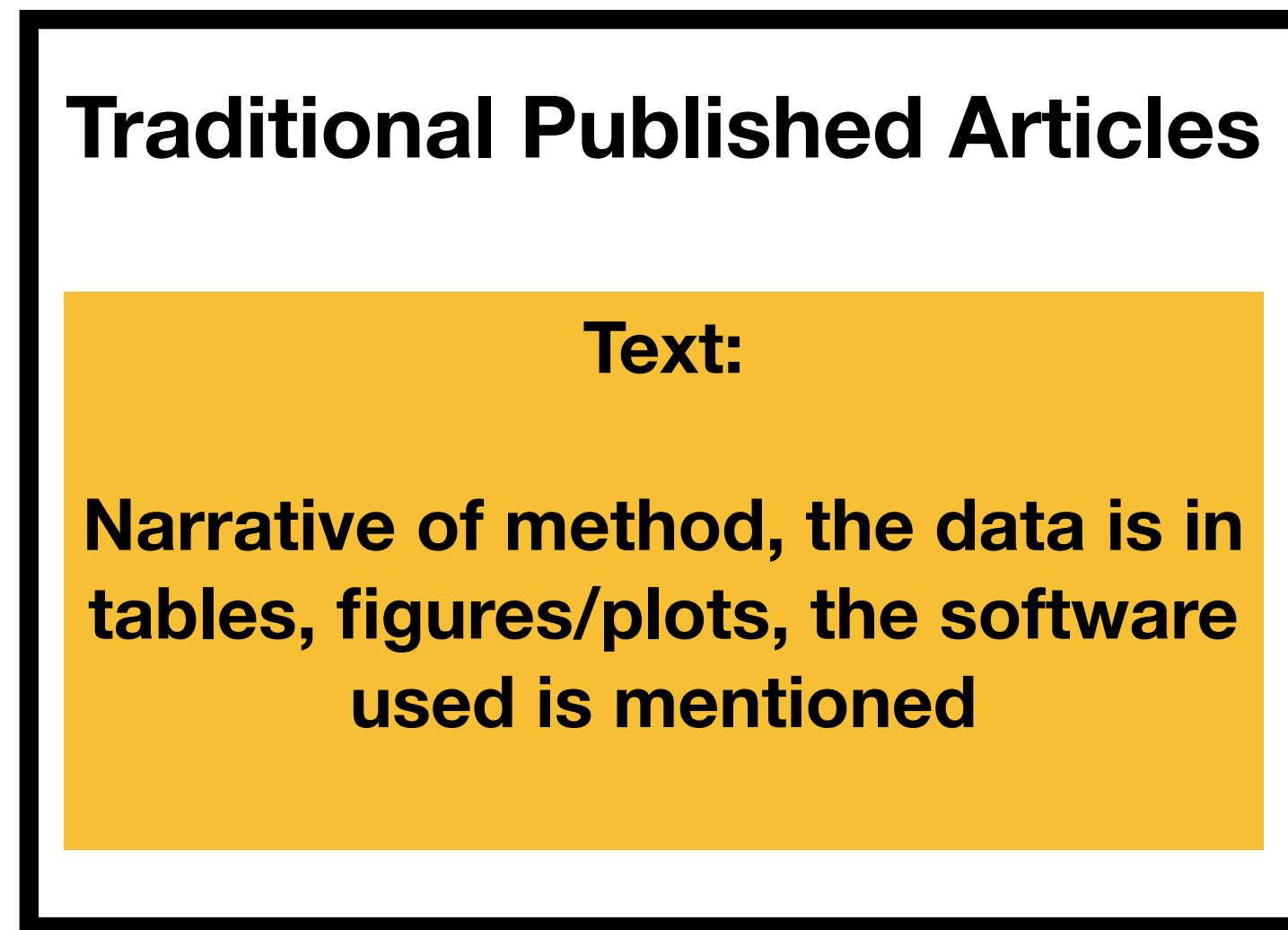
Principal Findings

We examined the citation history of 85 cancer microarray clinical trial publications with respect to the availability of their data. The 48% of trials with publicly available microarray data received 85% of the aggregate citations. Publicly available data was significantly ($p = 0.006$) associated with a 69% increase in citations, independently of journal impact factor, date of publication, and author country of origin using linear regression.

Modern Scientific Articles



Modern Scientific Articles



Scientific Paper of the Future

Modern Paper

Text:

Narrative of the method, some data is in tables, figures/plots, and the software used is mentioned

Data:

Include data as supplementary materials and pointers to data repositories

Reproducible Publication

Software:

For data preparation, data analysis, and visualization

Provenance and methods:

Workflow/scripts specifying dataflow, codes, configuration files, parameter settings, and runtime dependencies

Open Science

Sharing:

Deposit data and software (and provenance/workflow) in publicly shared repositories

Open licenses:

Open source licenses for data and software (and provenance/workflow)

Metadata:

Structured descriptions of the characteristics of data and software (and provenance/workflow)

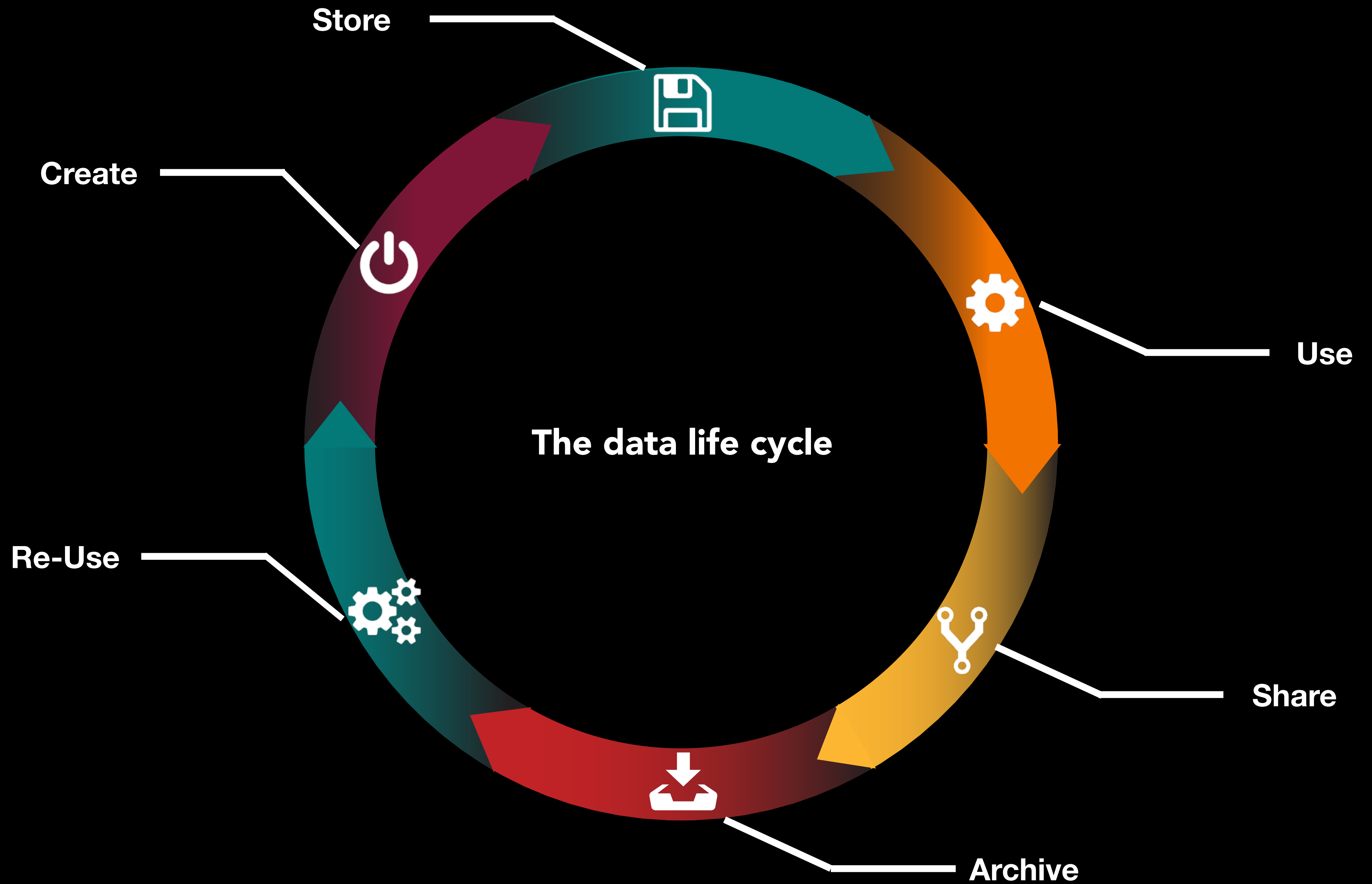
Digital Scholarship

Persistent identifiers:

For data, software, and authors (and provenance/workflow)

Citations:

Citations for data and software (and provenance/workflow)





Store



AND





AND





Store



AND





Store



AND





Store



AND

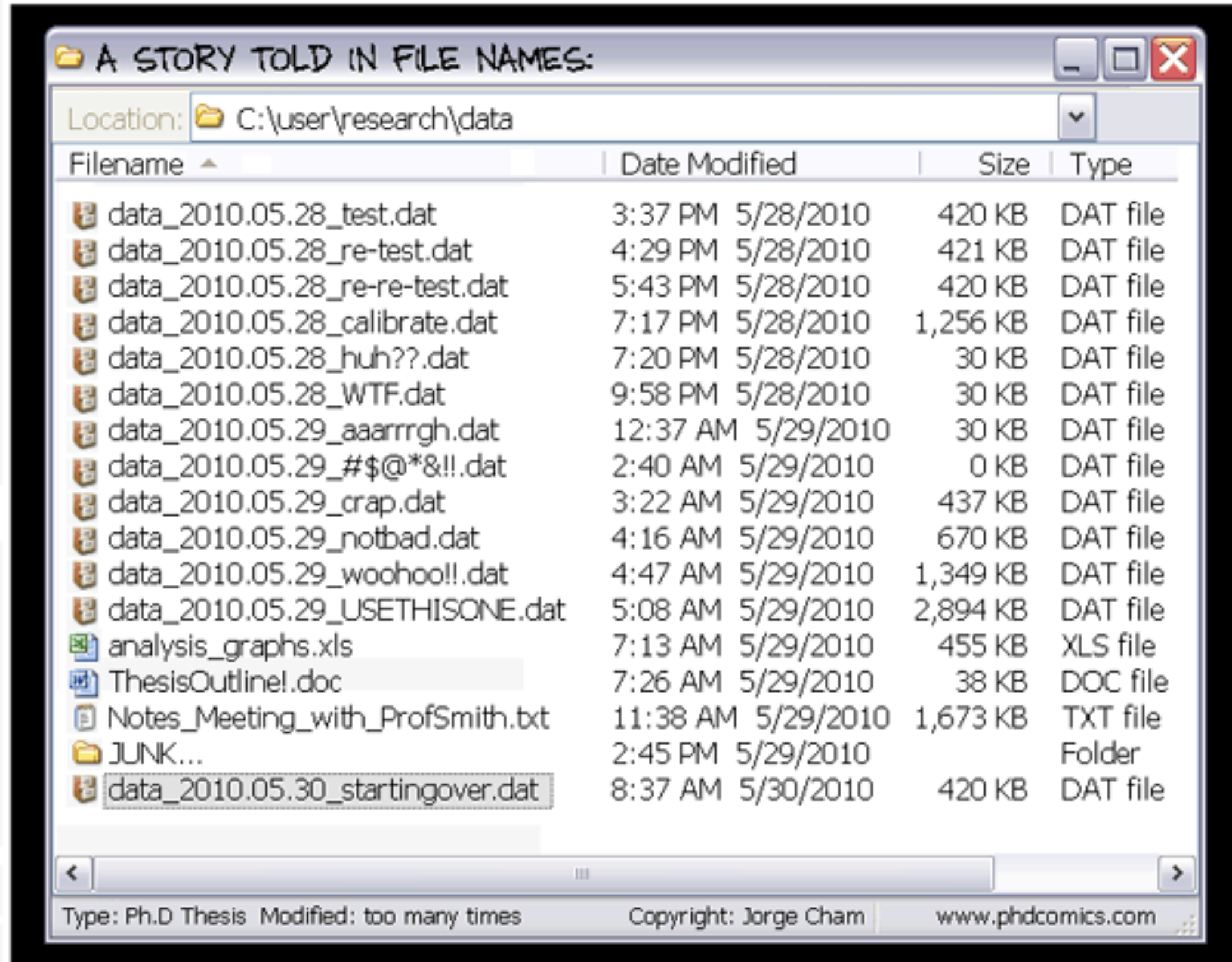


Keep in mind:

1. Some data backup is better than none
2. Automated backups are better than manual
3. Your data is only as safe as the last backup

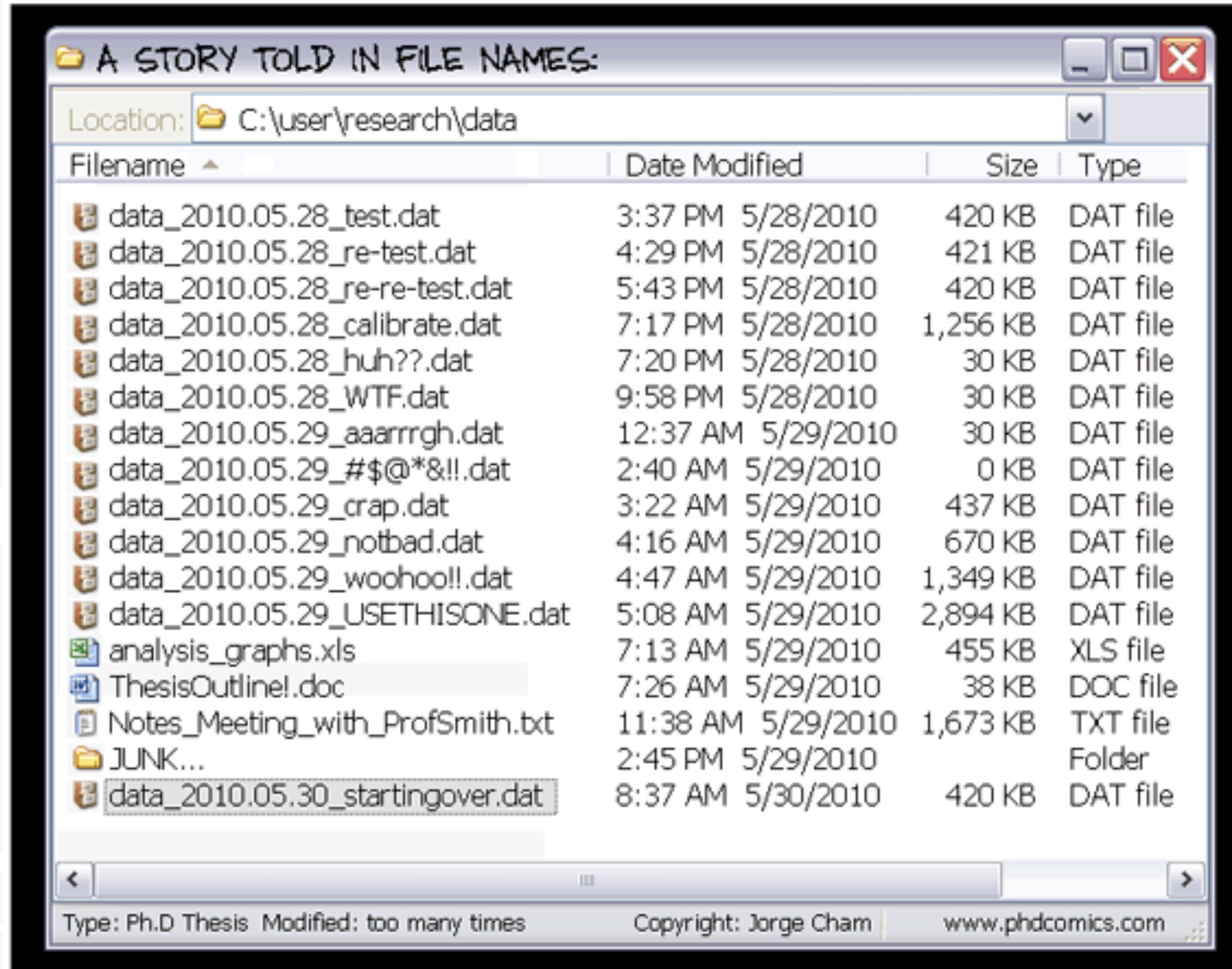


File Name and Organization





File Name and Organization



Morgan Edwards

@mangoedwards

[Follow](#)

I can't send you the original data because I don't remember what my excel file names mean anymore [#overlyhonestmethods](#)

9:11 AM - 8 Jan 2013



130



77



File Name and Organization



Project_Date_Description

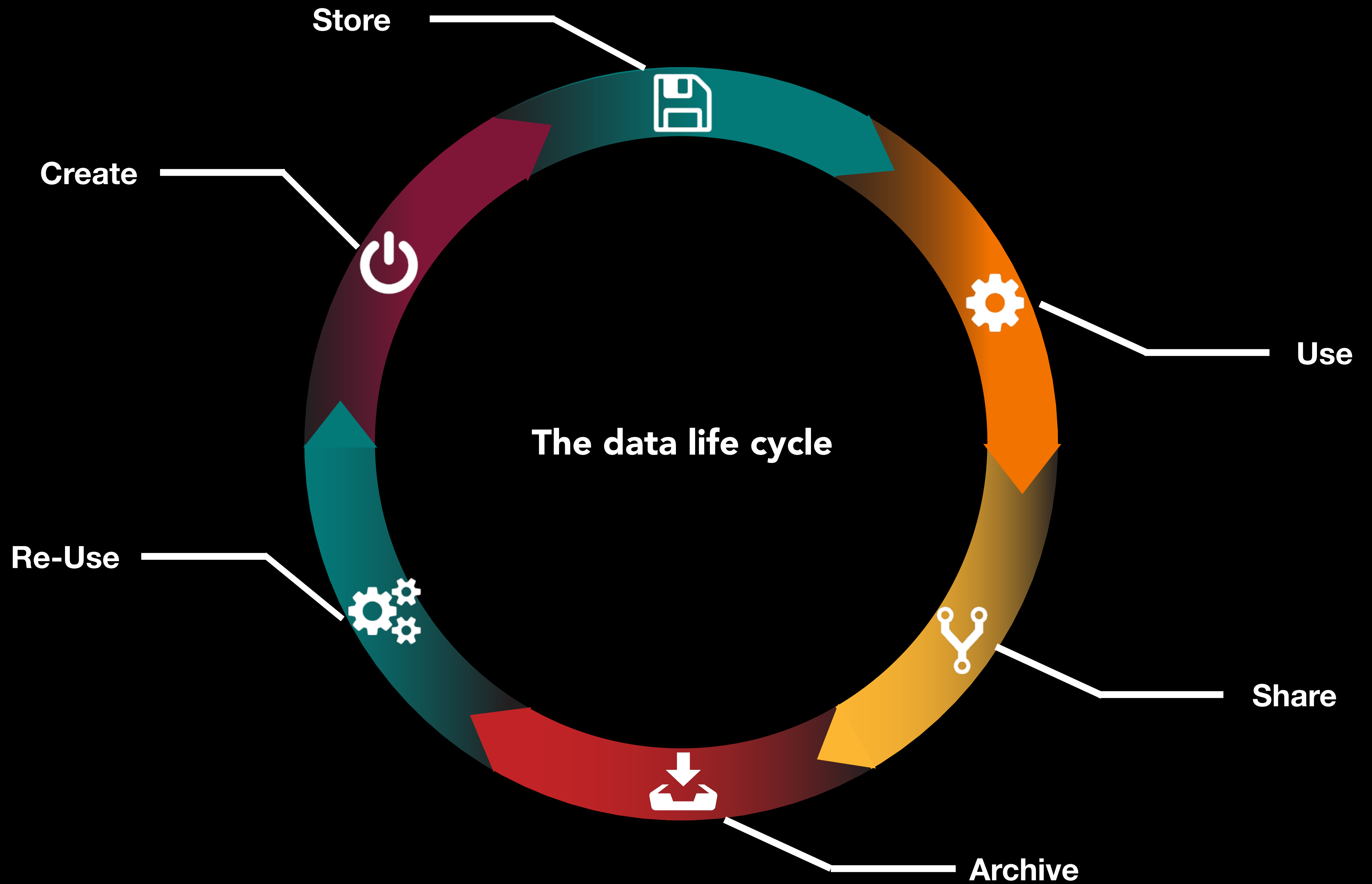


File Name and Organization



KEEP A FILE ABOUT YOUR FILES

► **ReadMe:** Description of what the files/folders contain





KEEP YOUR RAW DATA RAW!

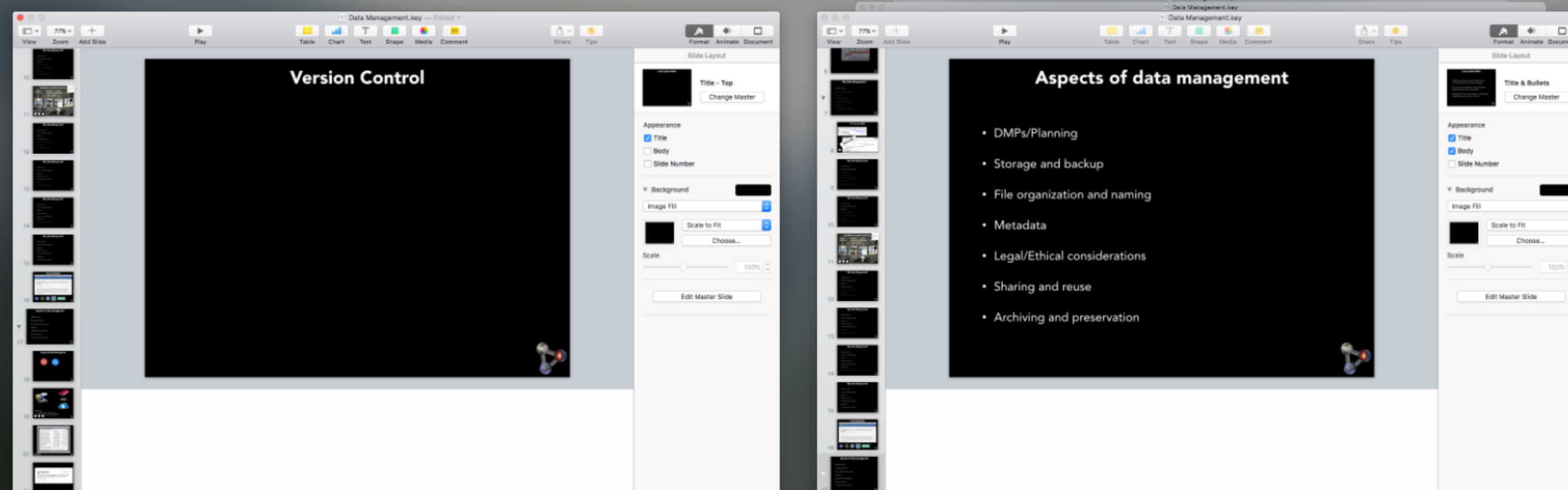
- ▶ Always keep the original data as raw as possible. Create new versions of dataset if you perform any data cleaning.
- ▶ Even more important when calibrating datasets or applying normalization, filters,...



Version Control



Version Control



Current Document

Done

Restore

Today at 2:16 PM

May 2016

Today

Built into Pages, Numbers, Keynote on the Mac...



Version Control

The image shows a Google Docs interface with the 'File' menu open and the 'Revision history' panel visible on the right.

File Menu Options:

- Share...
- New
- Open... ⌘O
- Rename...
- Make a copy...
- Organize...
- Move to trash
- See revision history ⌘+Option+Shift+H
- See new changes
- Language
- Download as
- Publish to the web...
- Email collaborators...
- Email as attachment...
- Page setup...
- Print preview
- Print ⌘P

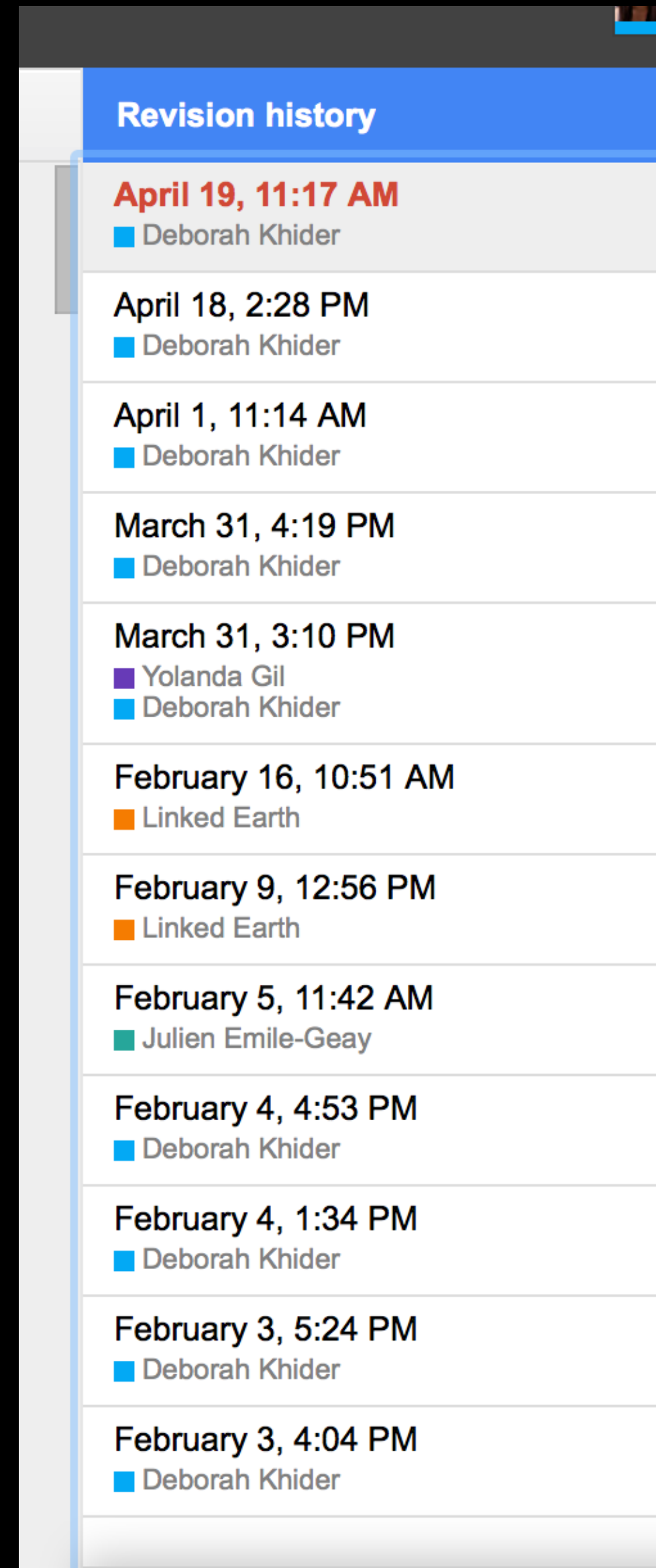
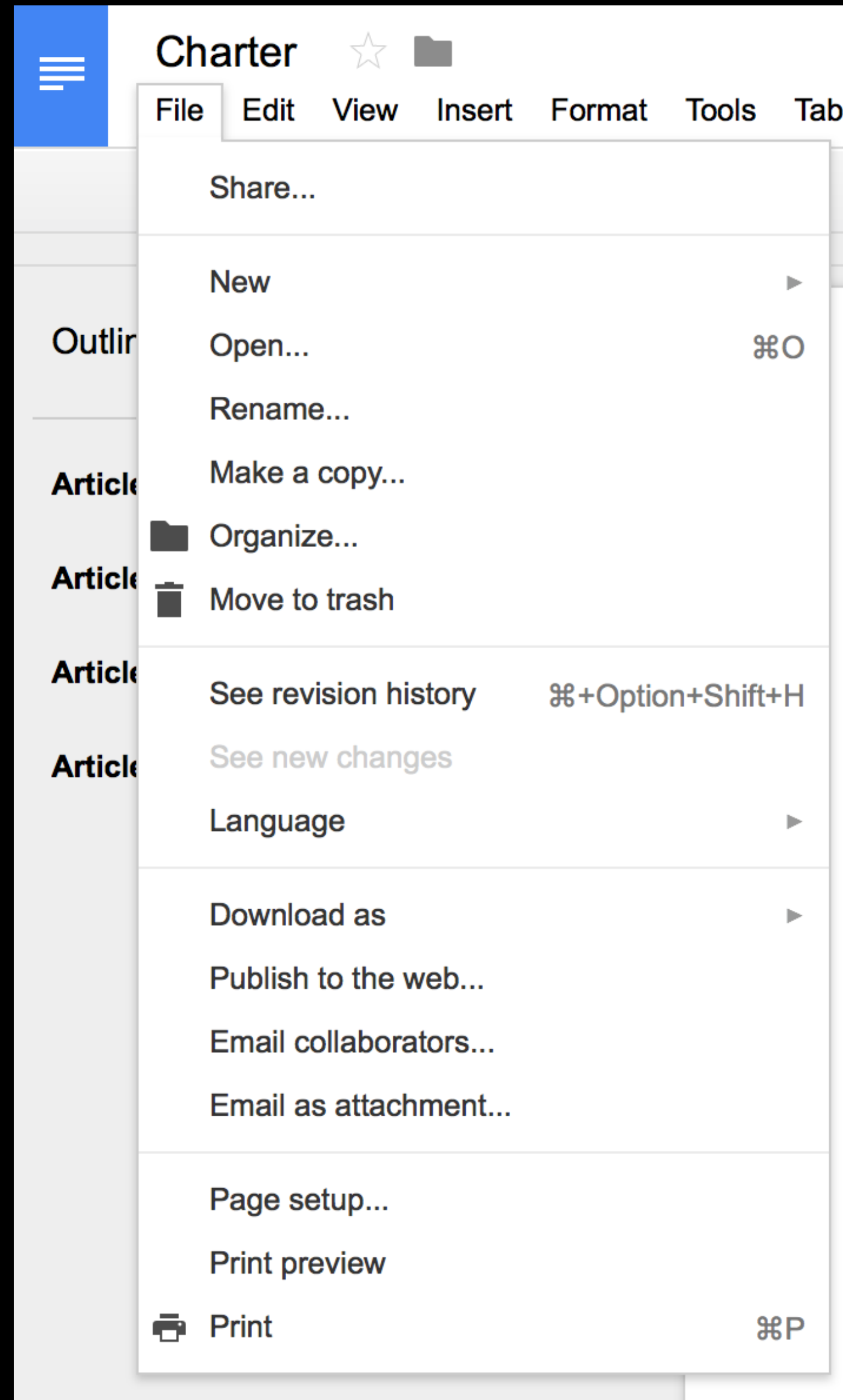
Revision history Panel:

| Revision history |
|--|
| April 19, 11:17 AM Deborah Khider |
| April 18, 2:28 PM Deborah Khider |
| April 1, 11:14 AM Deborah Khider |
| March 31, 4:19 PM Deborah Khider |
| March 31, 3:10 PM Yolanda Gil Deborah Khider |
| February 16, 10:51 AM Linked Earth |
| February 9, 12:56 PM Linked Earth |
| February 5, 11:42 AM Julien Emile-Geay |
| February 4, 4:53 PM Deborah Khider |
| February 4, 1:34 PM Deborah Khider |
| February 3, 5:24 PM Deborah Khider |
| February 3, 4:04 PM Deborah Khider |

... as well as Google apps



Version Control



... as well as Google apps



Version Control

Current Repository
datatransformation_regrid

Current Branch
main

Fetch origin
Last fetched just now

Changes 7

History New

acf1c2a4-25bd-460b-8d48-e738b907fbe8.yaml

7 changed files

✓ acf1c2a4-25bd-460b-8d48-e738b907fbe8.yaml

✓ ECMWF_regridded.nc

✓ ModelAnnotatedValues.yml

✓ RegriddingTransform.cwl

✓ RegriddingTransform_mic.yaml

✓ spec.yaml

✓ values.yaml

Summary (required)

Description

Commit to main

@@ -0,0 +1,209 @@

1 +!!python/object:modelcatalog.models.model_configuration.ModelConfiguration

2 +_author: null

3 +_citation: null

4 +_compatible_visualization_software: null

5 +_contributor: null

6 +_copyright_holder: null

7 +_date_created: null

8 +_date_published: null

9 +_description: null

10 +_doi: null

11 +_had_primary_source: null

12 +_has_assumption: null

13 +_has_build_file: null

14 +_has_causal_diagram: null

15 +_has_component_location:

16 +- https://publisher.mint.isi.edu/10PwcE/RegriddingTransform.cwl

17 +_has_constraint: null

18 +_has_contact_person: null

19 +_has_documentation: null

20 +_has_download_instructions: null

21 +_has_download_url: null

22 +_has_equation: null

23 +_has_example: null

24 +_has_executable_instructions: null

25 +_has_executable_notebook: null

26 +_has_execution_command: null

27



Use GitHub (code)



Version Control

←

Revision history

☐ Raw source

☐ Inline diff

Wed Feb 24 2021 20:22:11 GMT-0700 (MST)

Drought Period.ipynb

Text cell <QZ5Ef8bvPz9M>

[markdown]

1

2

3

4 # Drought prediction for South Sudan and Ethiopia

5

6

7 This notebook trains a CNN to predict drought as expressed as SPI

8

9

Text cell <ilCARfsLS5fC>

[markdown]

1 ## Software requirement

Code cell <2y_fuUzgRtK->

[code]

1 !pip install netCDF4

Execution output from Nov 20, 2020 11:42 AM

Stream

Collecting netCDF4

[?25l Downloading https://files.pythonhosted.org/package

[K | 4.3MB 5.1MB/s

[?25hRequirement already satisfied: numpy>=1.9 in /usr/lo

Collecting cftime

[?25l Downloading https://files.pythonhosted.org/package

[K | 296kB 30.0MB/s

[?25hInstalling collected packages: cftime, netCDF4

Successfully installed cftime-1.3.0 netCDF4-1.5.4

Code cell <UWWai2L8QGMN>

[code]

Wed Feb 24 2021 20:22:11 GMT-0700 (MST)

Drought Period.ipynb

Text cell <QZ5Ef8bvPz9M>

[markdown]

1

2

3

4 # Drought prediction for South Sudan and Ethiopia

5

6

7 This notebook trains a CNN to predict drought as expressed as SPI

8

9

Text cell <ilCARfsLS5fC>

[markdown]

1 ## Software requirement

Code cell <2y_fuUzgRtK->

[code]

1 !pip install netCDF4

Execution output from Nov 20, 2020 11:42 AM

Stream

Collecting netCDF4

[?25l Downloading https://files.pythonhosted.org/package

[K | 4.3MB 5.1MB/s

[?25hRequirement already satisfied: numpy>=1.9 in /usr/lo

Collecting cftime

[?25l Downloading https://files.pythonhosted.org/package

[K | 296kB 30.0MB/s

[?25hInstalling collected packages: cftime, netCDF4

Successfully installed cftime-1.3.0 netCDF4-1.5.4

Code cell <UWWai2L8QGMN>

[code]

☐ Only show named versions

Feb 24, 2021 8:22 PM

Deborah Khider

Don't Forget the Metadata!!!

TYPES OF METADATA

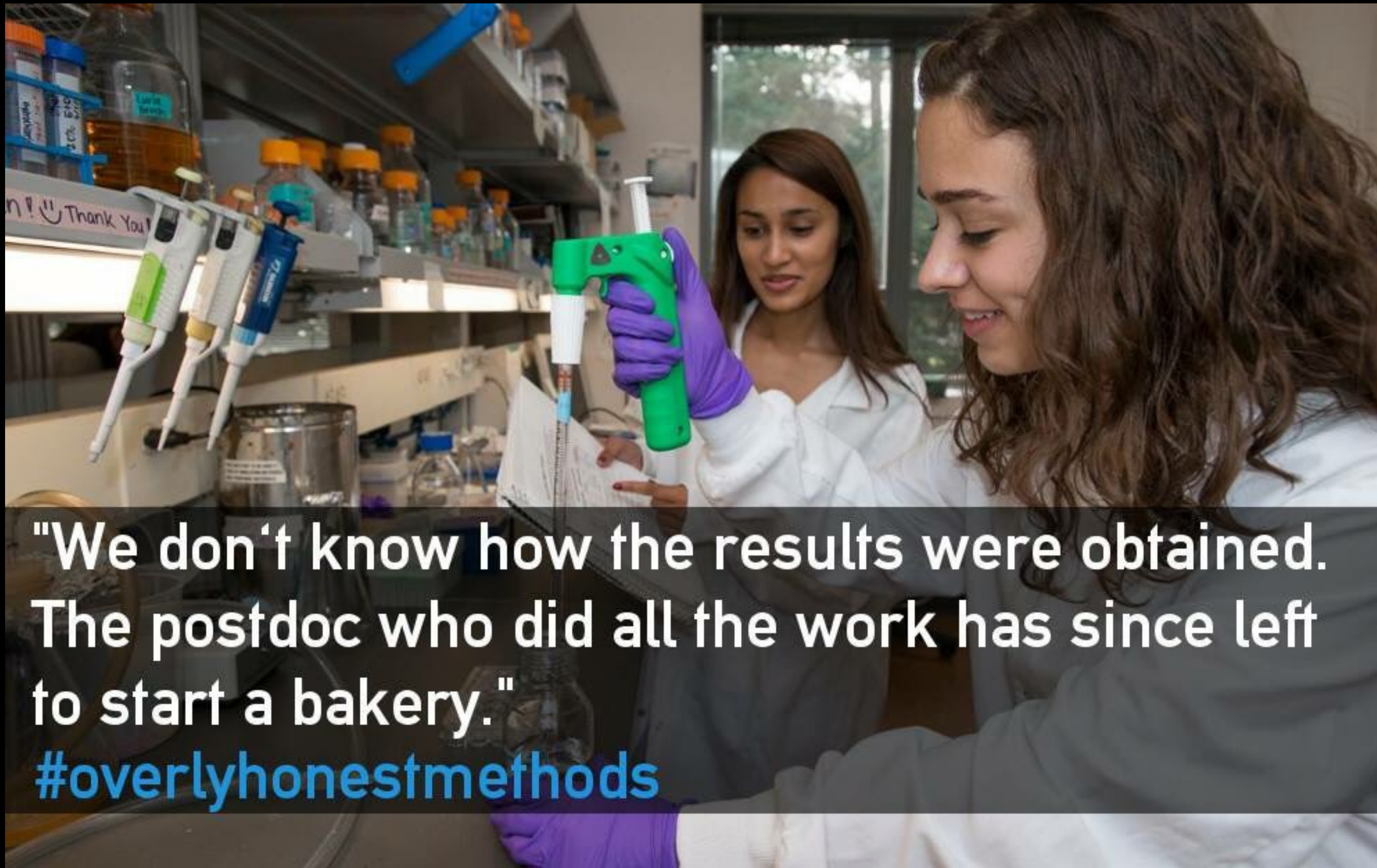
- ▶ **Descriptive Metadata:** Location, collection frequency, object, etc...
- ▶ **Data Characteristics:** Size, statistical properties,...
- ▶ **Provenance metadata:** Instrument, Method/Software, Parameters...



Don't Forget the Metadata!!!

USE OF METADATA

- **Facilitate** reuse by others





Don't Forget the Metadata!!!

USE OF METADATA

- ▶ **Facilitate** reuse by others
- ▶ Support **queries** on data repository

```
In [4]: import json
import requests

url = "http://wiki.linked.earth/store/ds/query"

query = """PREFIX core: <http://linked.earth/ontology#>
PREFIX wiki: <http://wiki.linked.earth/Special:URIResolver/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
SELECT distinct ?a
WHERE {
{
    ?dataset wiki:Property-3AArchiveType ?a.
}UNION
{
    ?w core:proxyArchiveType ?t.
    ?t rdfs:label ?a
}
}"""

response = requests.post(url, data = {'query': query})
res = json.loads(response.text)

print("The following archive types are available on the wiki:")
for item in res['results']['bindings']:
    print ("*" + item['a']['value'])

The following archive types are available on the wiki:
*marine sediment
*coral
*lake sediment
*glacier ice
*tree
*documents
*speleothem
*sclerosponge
*borehole
*hybrid
*bivalve
*Rock
```


Don't Forget the Metadata!!!

USE OF METADATA

- ▶ **Facilitate** reuse by others
- ▶ Support **queries** on data repository
- ▶ **Explain a data analysis** by providing context for the data





Don't Forget the Metadata!!!

USE OF METADATA

- ▶ **Facilitate** reuse by others
- ▶ Support **queries** on data repository
- ▶ **Explain a data analysis** by providing context for the data
- ▶ Enable **automated data integration**



Use

THE BIGGEST LIE I TELL
MYSELF IS "I DON'T
NEED TO WRITE THAT
DOWN I'LL REMEMBER."

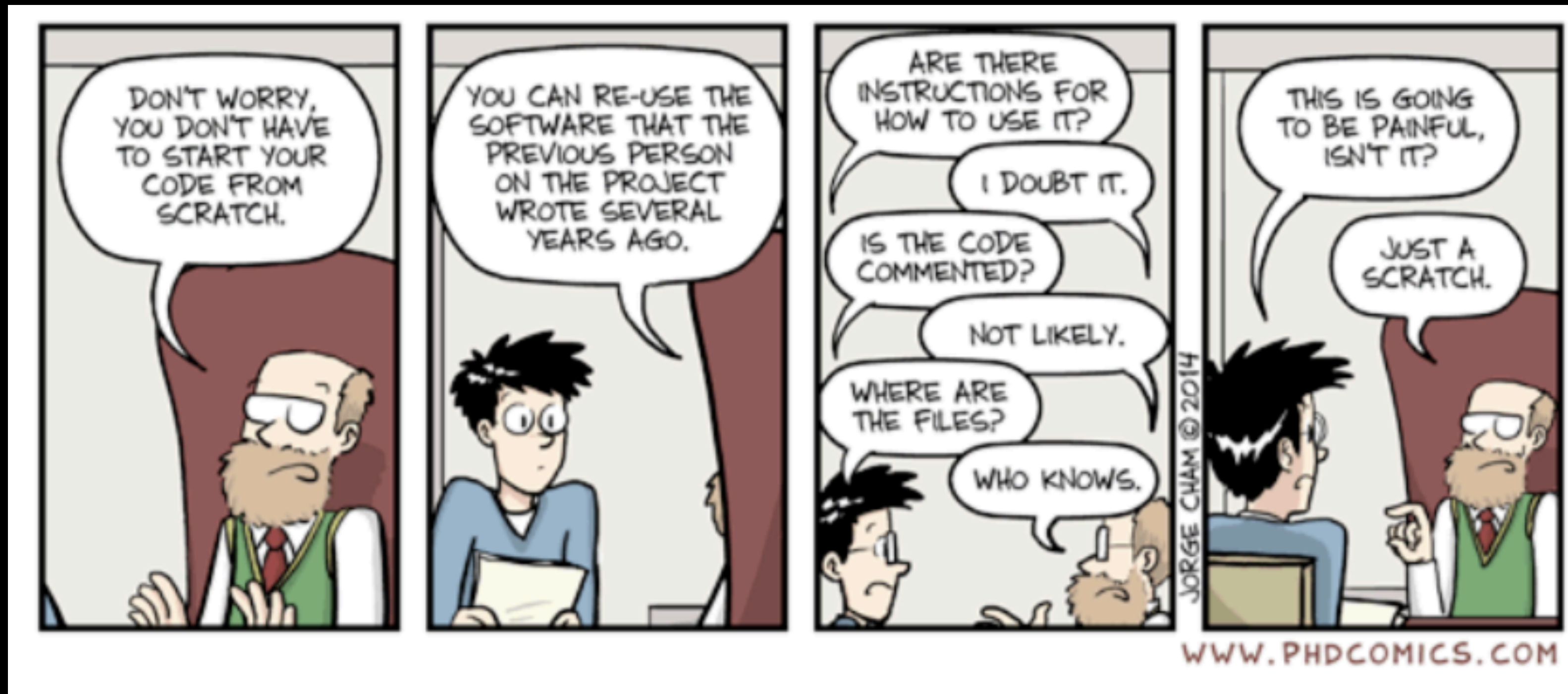
Everyone at some point in their life.



Use

WRITE IT DOWN!

- **Methods:** Laboratory, statistics, data cleaning...
- **Comment your code:** 1 line of code = 1 line of comment
- **Meeting notes**





Keep code and explanation together!



Notebook

jupyter spectrogram (autosaved)



File Edit View Insert Cell Kernel Help

Python 3

Save Add Cut Copy Paste Undo Redo Run Stop Restart Markdown CellToolbar

Simple spectral analysis

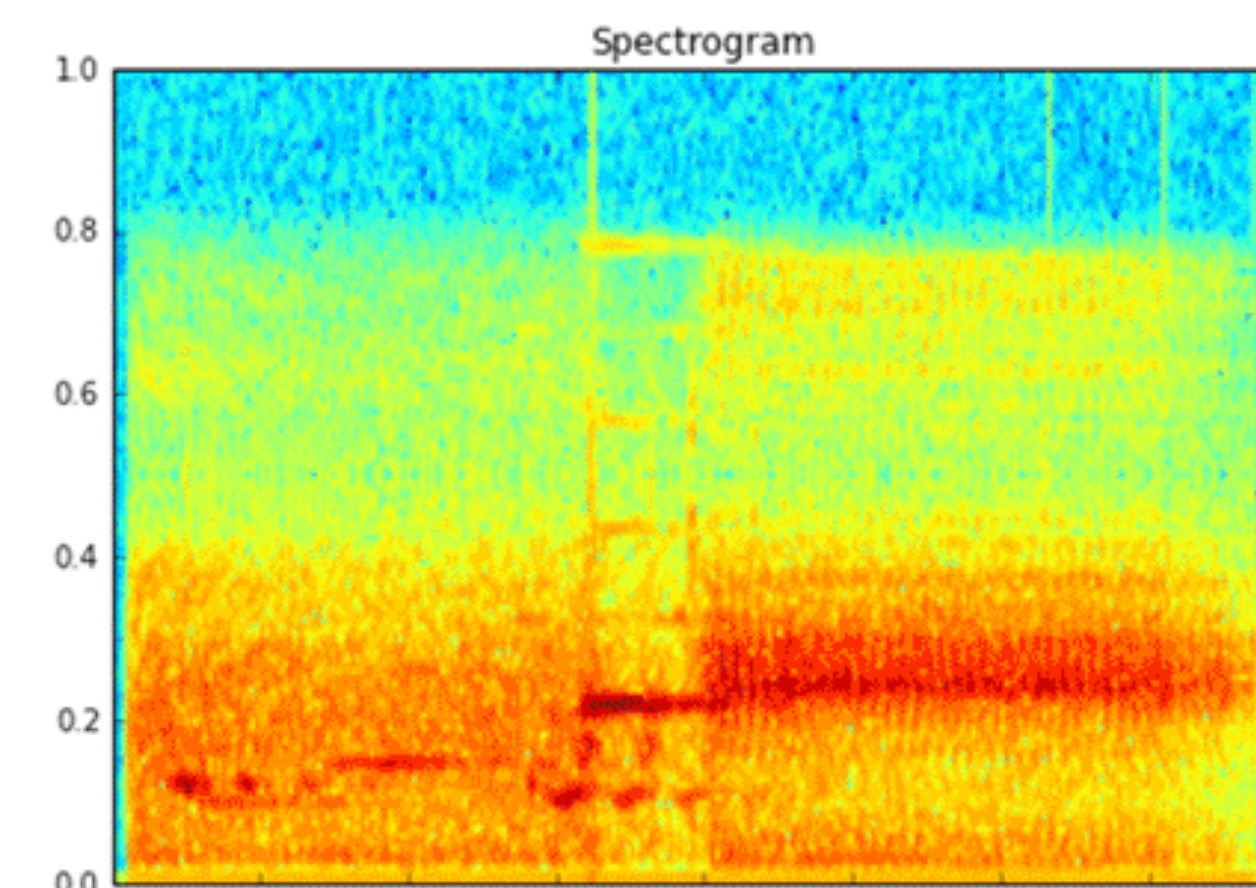
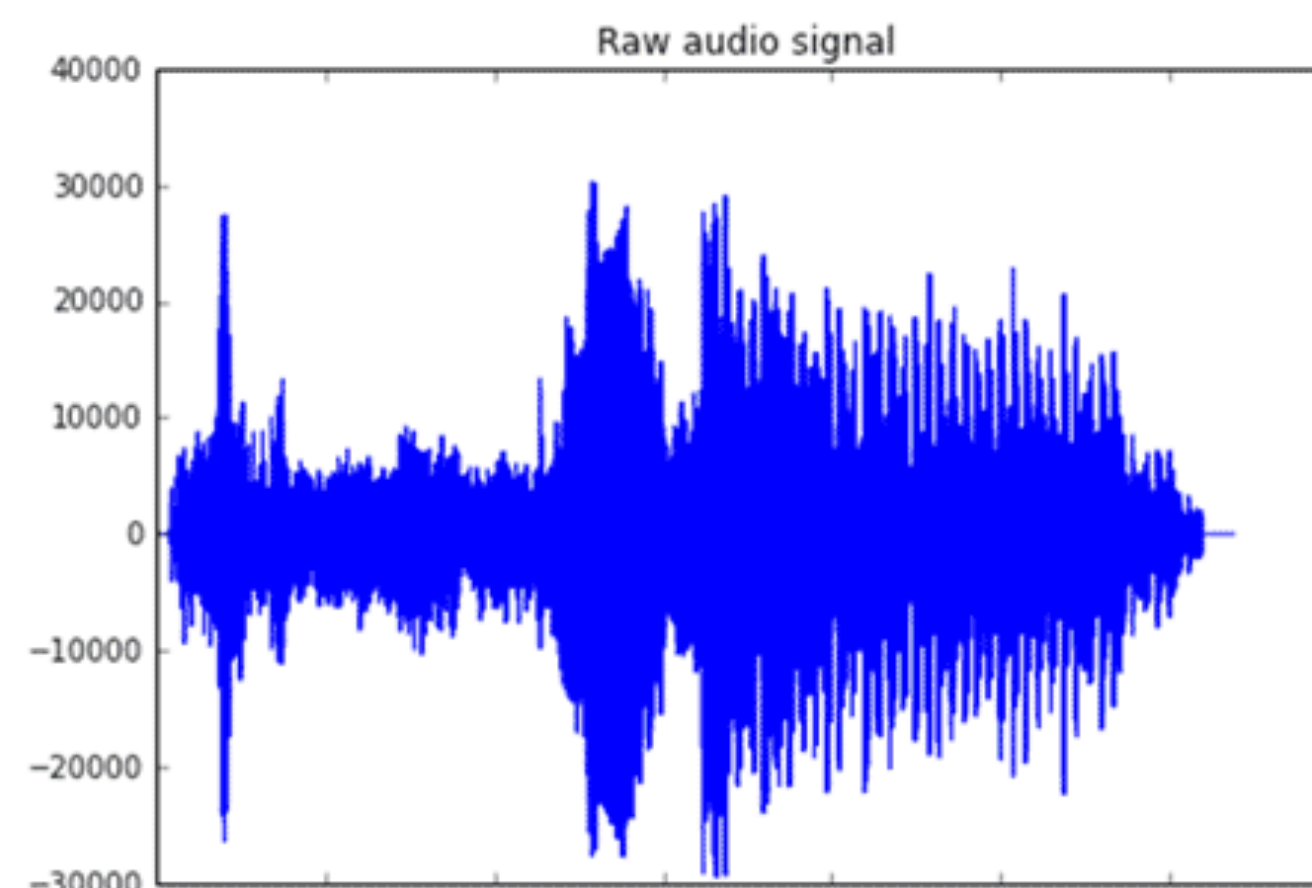
An illustration of the [Discrete Fourier Transform](#)

$$X_k = \sum_{n=0}^{N-1} x_n \exp\left(\frac{-2\pi i}{N} kn\right) \quad k = 0, \dots, N-1$$

```
In [2]: from scipy.io import wavfile
rate, x = wavfile.read('test_mono.wav')
```

And we can easily view it's spectral structure using matplotlib's builtin specgram routine:

```
In [5]: fig, (ax1, ax2) = plt.subplots(1,2,figsize(16,5))
ax1.plot(x); ax1.set_title('Raw audio signal')
ax2.specgram(x); ax2.set_title('Spectrogram');
```





Keep code and explanation together!



R Markdown

from R Studio

```
1 ---
2 title: "Exploring the BRFSS data"
3 output:
4   html_document:
5     fig_height: 4
6     highlight: pygments
7     theme: spacelab
8 ---
9
10 ## Setup
11
12 ### Load packages
13
14 ```{r load-packages, message = FALSE}
15 library(ggplot2)
16 library(dplyr)
17 ```
18
19 ### Load data
20 Load the BRFSS data into the workspace.
21
22 ```{r load-data}
23 load("brfss2013.RData")
```

1:1 # Exploring the BRFSS data R Markdown



Keep code and explanation together!



R Markdown

from R Studio

```
1 ---
2 title: "Exploring the BRFSS data"
3 output:
4   html_document:
5     fig_height: 4
6     highlight: pygments
7     theme: spacelab
8 ---
9
10 ## Setup
11
12 ### Load packages
13
14 ```{r load-packages, message = FALSE}
15 library(ggplot2)
16 library(dplyr)
17 ```
18
19 ### Load data
20 Load the BRFSS data into the workspace.
21
22 ```{r load-data}
23 load("brfss2013.RData")
```

1:1 # Exploring the BRFSS data R Markdown

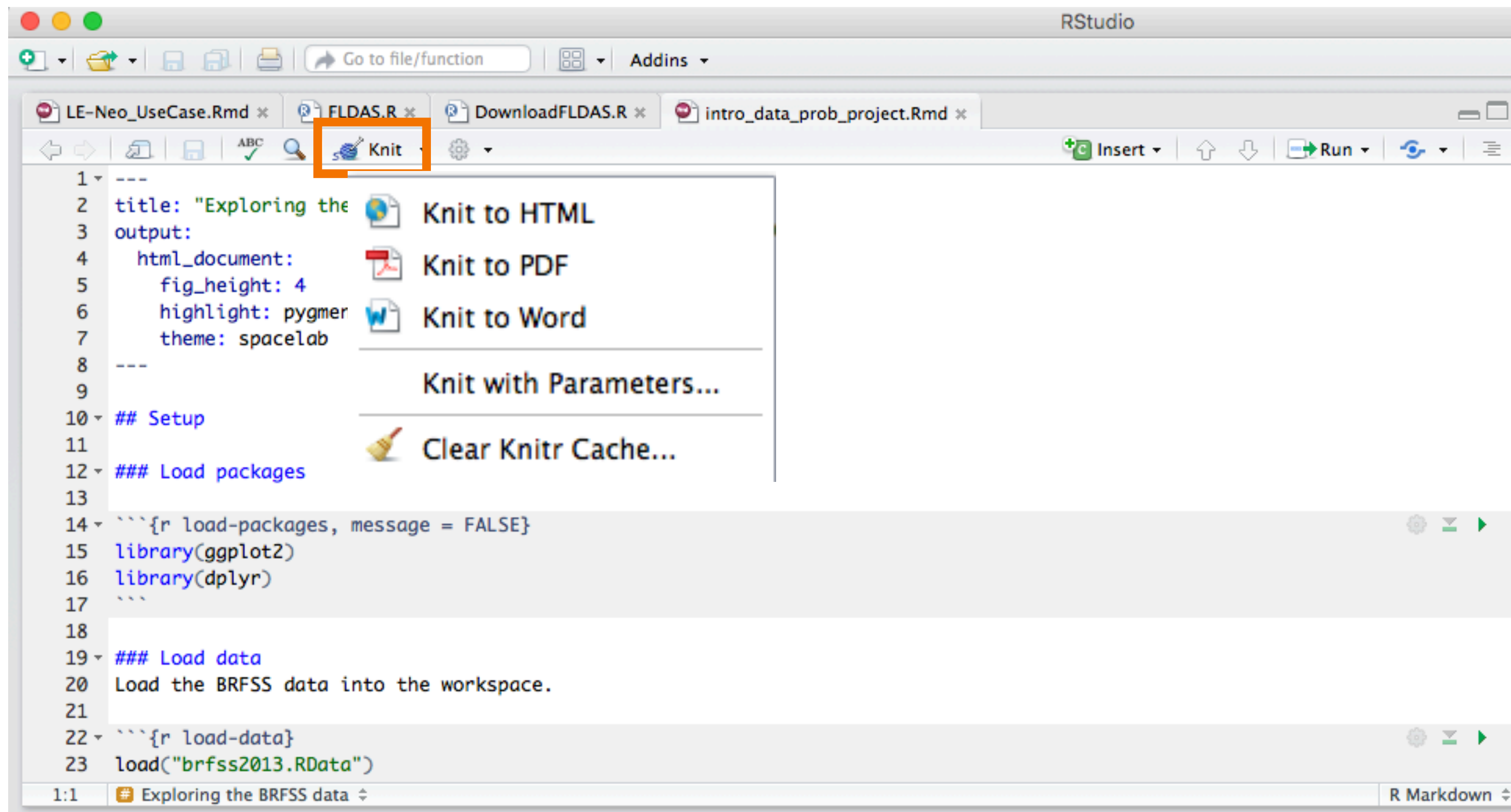


Keep code and explanation together!



R Markdown

from R Studio





Keep code and explanation together!



Notebook

R Markdown

from  Studio

Part 3: Exploratory data analysis

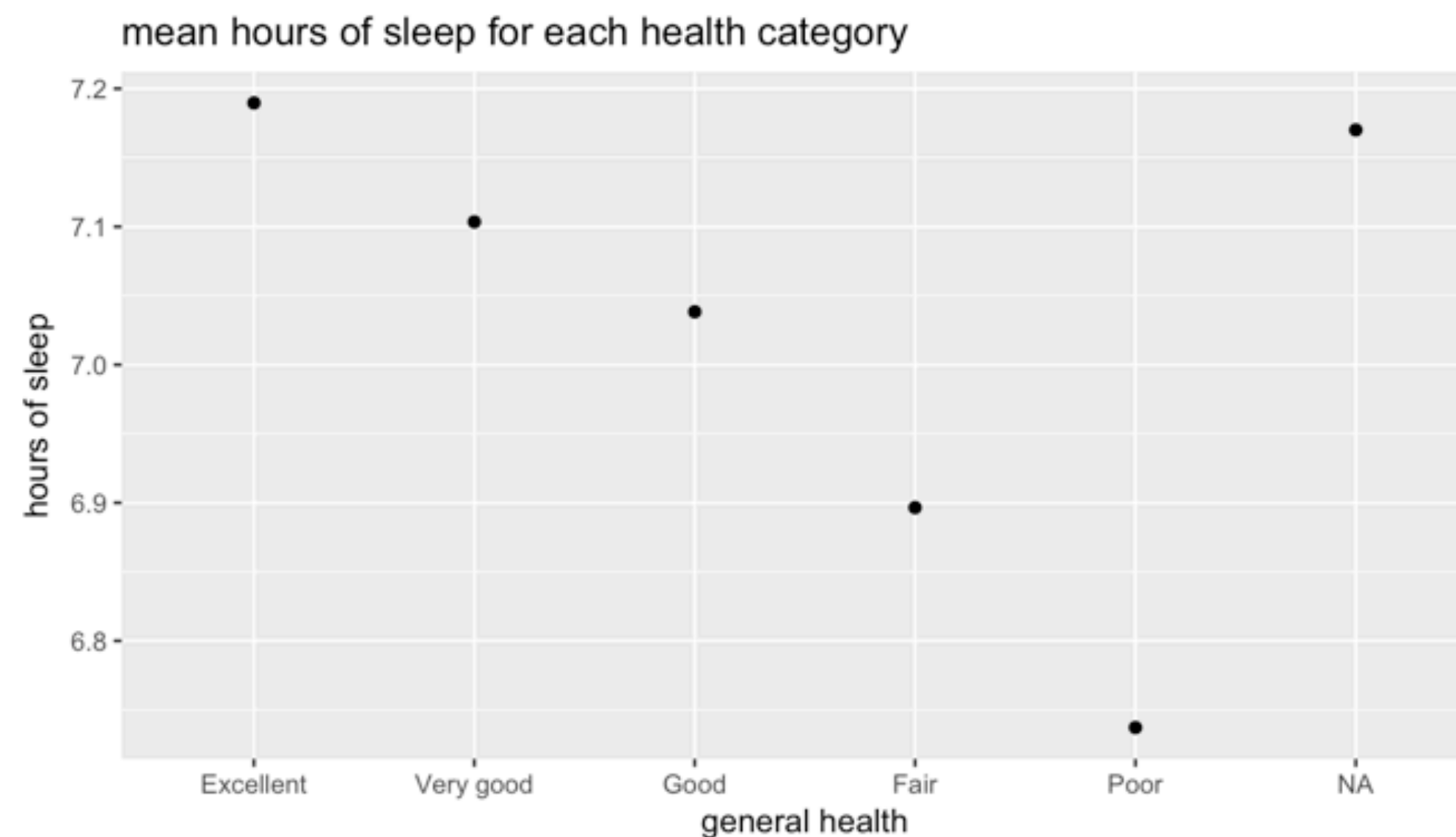
Research question 1: Relationship between sleep and general health

```
#Remove outliers (can't really sleep more than 24hours)
sleeptime <- filter(brfss2013, sleptiml<=24)
```

```
## Warning: package 'bindrcpp' was built under R version 3.2.5
```

```
# Summarize the data
healthSleep <- sleeptime %>%
  group_by(genhlth) %>%
  summarise(meanSleep = mean(sleptiml))

#Plot the general health/mean of sleep time
ggplot(healthSleep, aes(genhlth, meanSleep)) + geom_point(aes(genhlth, meanSleep)) +
  labs(title="mean hours of sleep for each health category", x = 'general health', y='hours of sleep')
```



From the plot above, there appears

to be a general correlation between health and the amount of sleep.



Keep code and explanation together!



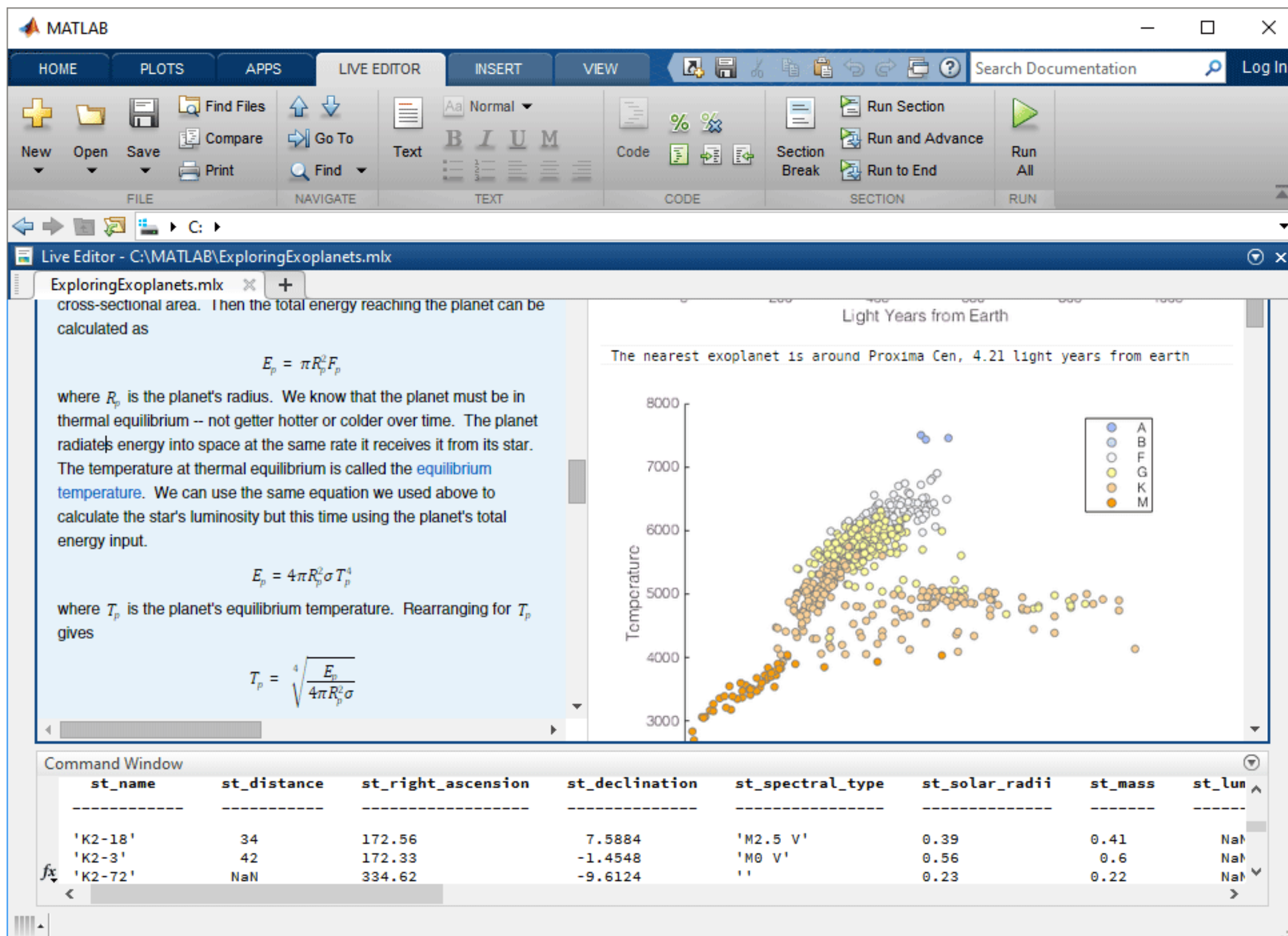
Notebook

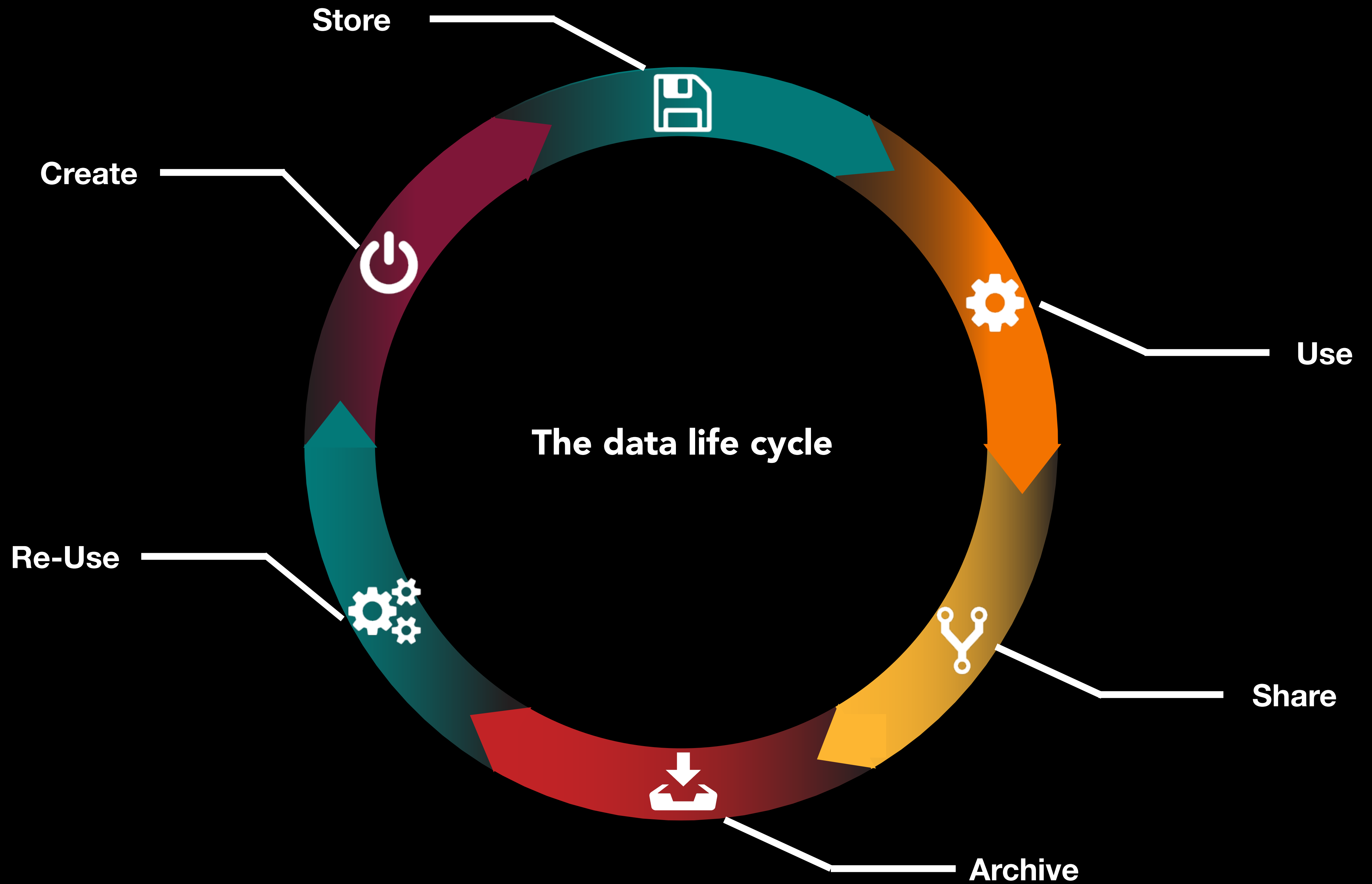
R Markdown

from R Studio



MATLAB Live Editor







Share: Physical Samples

INTERNATIONAL GEO SAMPLE NUMBER: IGSN

- ▶ Globally unique and persistent identifier for physical samples in the Earth Sciences
- ▶ To obtain a number, go to <http://www.geosamples.org/>
- ▶ Record and register quality metadata for your samples
 - ▶ At a minimum: Location, contact, access restrictions, lithology
- ▶ Use IGSNs in your publications: text, data tables,...

IGSN: DSR00050U




| | |
|----------------|--------------|
| IGSN: | DSR00050U |
| Sample Name: | VM28-328A |
| Other Name(s): | |
| Sample Type: | Core |
| Parent IGSN: | Not Provided |



Share: Data

1. Publication in a shared repository

 [figshare.com/](#)

Highly connected drug file

| | | | |
|--------------------|-----|-----|---|
| Tretinoin | 257 | 46 | Rv1155, aroG, Rv1264, mscL, thyX, gmk, glnA1, Rv1264, thyX, glnA1, trpD, leuA, blaI, ethR, Rv0223c, lipJ, Rv1264, ephG, blaI, ethR, sigC, b |
| Levothyroxine | 173 | 36 | 25 cyp130, Rv1264, lppX, gpml, ligA, nirA |
| Methotrexate | 156 | 32 | TB31.7, Rv1264, mscL, lppX, coaA, pcaA, Rv3676, fabG1, |
| 4-Hydroxytamoxifen | 115 | 115 | mmaA4, bphD, Rv1264, mscL, thyX, lppX, mmaA2, ptl |
| Estradiol | 98 | 20 | TB31.7, cyp130, aroG, Rv1264, secA1, trpD, nirA |
| Amantadine | 79 | 1 | pth, ethR, clpP, glbN, inhA, |
| Rifampin | 78 | 13 | pknD, lipJ, fabH, Rv1941, Rv3361c, Rv1264, lppX |
| Raloxifene | 75 | 18 | mmaA4, Rv1264, groEL, lppX, secA1, glmU, nusA, R |
| Propofol | 54 | 5 | mmaA4, Rv1264, thyX, lppX, secA1, serA1, Rv3529c |
| Indinavir | 51 | 14 | pepD, Rv1264, thyX, ethR, trxB2, |
| Penicillamine | 44 | 10 | pknD, pepD, fabH, Rv1941, devB, ppp, ftsZ, cyp12 |
| Daunorubicin | 44 | 12 | |
| Triclosan | 42 | 5 | |
| Darunavir | 40 | 15 | |

[Enlarge to see the rest of the document](#)

[Enlarge](#) [Download](#)

Categories

- Computational Biology

Authors


Daniel Garijo
Lei Xie
Yinliang Zhang
Yolanda Gil
Li Xie
Sarah Kinnings
Phil Bourne

Tags

[results](#) [tb-drugome](#)

License (what's this?)

CC-BY



Cite this: Garijo, Daniel; Xie, Lei; Zhang, Yinliang; Gil, Yolanda; Xie, Li; Kinnings, Sarah; Bourne, Phil (2013): Highly connected drug file. [figshare](#).
<http://dx.doi.org/10.6084/m9.figshare.776887>
Retrieved 08:56, Feb 20, 2015 (GMT)

Description

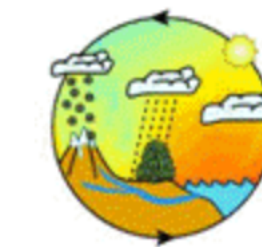
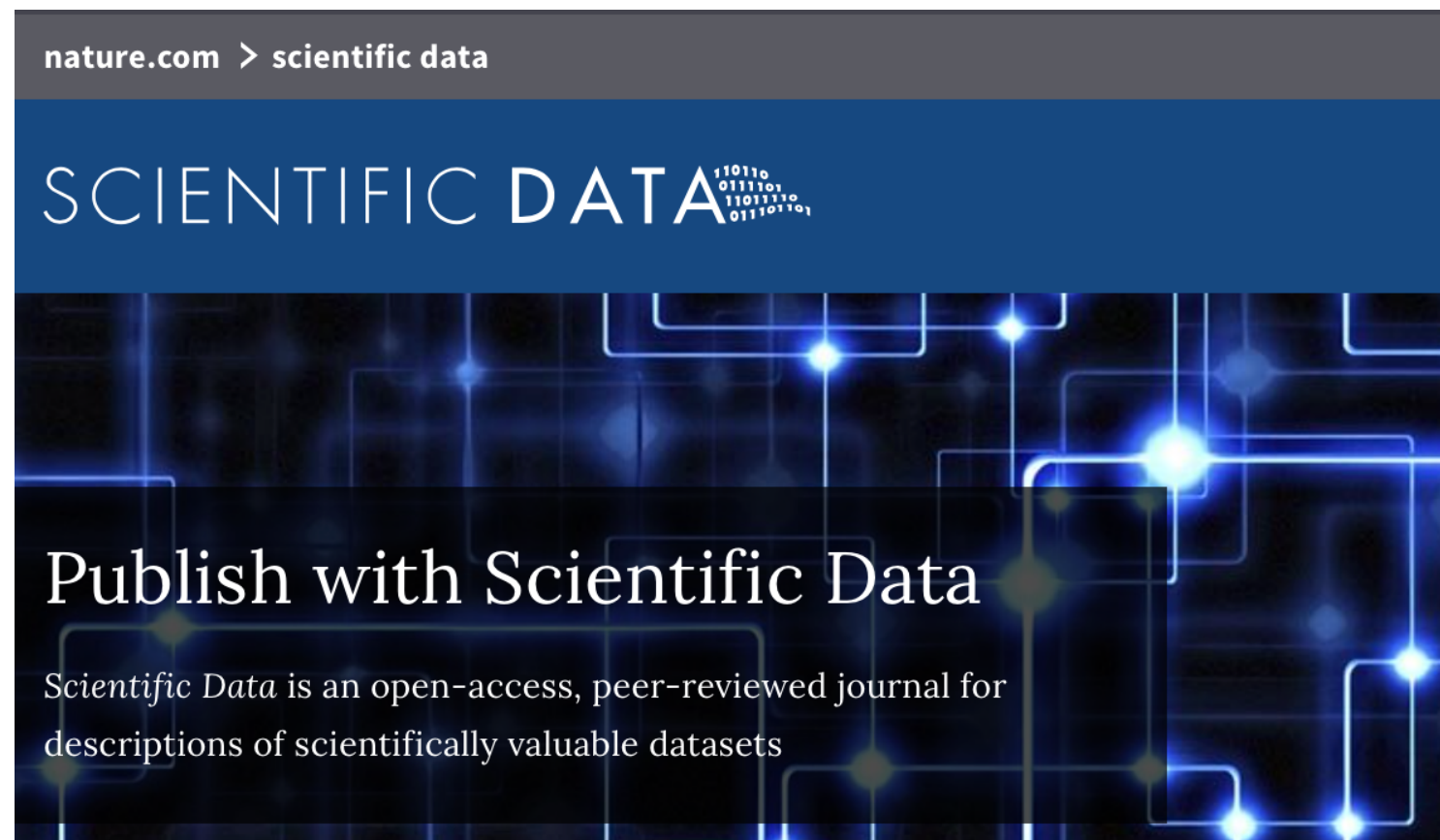
Highly connected drug file obtained as a result of the TB-Drugome Workflow.

Links

- <http://purl.org/net/tb-drugome-run>



Share: Data



Waterisotopes.org



Open
Core Data



HYDROSHARE

Magic



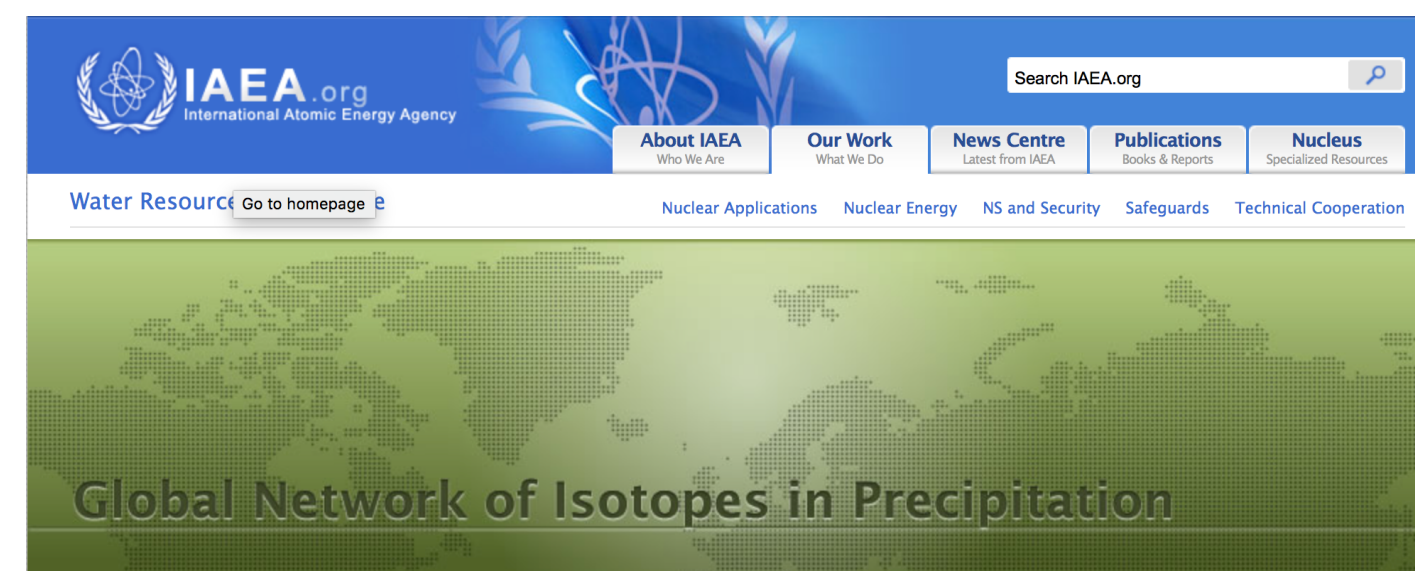
Neotoma



LinkedEarth

CSDCO

BALTO





Share: Data

figshare.com/

Highly connected drug file

| | | | |
|--------------------|-----|-----|---|
| Tretinoin | 257 | 46 | Rv1155, aroG, Rv1264, mscL, thyX, gmk, glnA1, Rv1264, thyX, glnA1, trpD, leuA, blaI, ethR, Rv0223c, lipJ, Rv1264, ephG, blaI, ethR, sigC, b |
| Levothyroxine | 173 | 36 | 25 cyp130, Rv1264, lppX, gpml, ligA, nirA |
| Methotrexate | 156 | 32 | TB31.7, Rv1264, mscL, lppX, coaA, pcaA, Rv3676, fabG1, |
| 4-Hydroxytamoxifen | 115 | 115 | mmaA4, bphD, Rv1264, mscL, thyX, lppX, mmaA2, ptl |
| Estradiol | 98 | 20 | TB31.7, cyp130, aroG, Rv1264, secA1, trpD, nirA |
| Amantadine | 79 | 1 | pth, ethR, clpP, glbN, inhA, |
| Rifampin | 78 | 13 | pknD, lipJ, fabH, Rv1941, Rv3361c, Rv1264, lppX |
| Raloxifene | 75 | 18 | mmaA4, Rv1264, groEL, lppX, secA1, glmU, nusA, R |
| Propofol | 54 | 5 | mmaA4, Rv1264, thyX, lppX, secA1, serA1, Rv3529c |
| Indinavir | 51 | 14 | pepD, Rv1264, thyX, ethR, trxB2, |
| Penicillamine | 44 | 10 | pknD, pepD, fabH, Rv1941, devB, ppp, ftsZ, cyp12 |
| Daunorubicin | 44 | 12 | |
| Triclosan | 42 | 5 | |
| Darunavir | 40 | 15 | |

[Enlarge to see the rest of the document](#)

[Enlarge](#) [Download](#)

Cite this: Garijo, Daniel; Xie, Lei; Zhang, Yinliang; Gil, Yolanda; Xie, Li; Kinnings, Sarah; Bourne, Phil (2013): Highly connected drug file. figshare.
<http://dx.doi.org/10.6084/m9.figshare.776887>
Retrieved 08:56, Feb 20, 2015 (GMT)

Description

Highly connected drug file obtained as a result of the TB-Drugome Workflow.

Links

- <http://purl.org/net/tb-drugome-run>

Categories

- Computational Biology

Authors


Daniel Garijo
Lei Xie
Yinliang Zhang
Yolanda Gil
Li Xie
Sarah Kinnings
Phil Bourne

Tags

- results
- tb-drugome

License (what's this?)

CC-BY



1. Publication in a shared repository

2. General and domain metadata



Share: Data

GENERAL

- ▶ Dataset name/title
- ▶ Description
- ▶ Creator(s)
- ▶ Publication date
- ▶ License
- ▶ Publisher/contact
- ▶ Version
- ▶ Resource type
- ▶ Location of the data

DOMAIN SPECIFIC

- ▶ Categories
- ▶ Keywords/tags
- ▶ Related Links
- ▶ A data repository in a given discipline may request metadata using accepted standards



Share: Data

GENERAL

- ▶ Dataset name/title
- ▶ Description
- ▶ Creator(s)
- ▶ Publication date
- ▶ License
- ▶ Publisher/contact
- ▶ Version
- ▶ Resource type
- ▶ Location of the data

DOMAIN SPECIFIC

- ▶ Categories
- ▶ Keywords/tags
- ▶ Related Links
- ▶ A data repository in a given discipline may request metadata using accepted standards



Share: Data

Choose a License

Creative Commons Corporation creativecommons.org/choose/

YG WINGS WINGS-Portal ODS DII EC ECC ISD ISI

creative commons About Licenses Public Domain Support CC Projects News

License Features

Your choices on this panel will update the other panels on this page.

Allow adaptations of your work to be shared?

☒ Yes ☐ No



☐ Yes, as long as others share alike

Allow commercial uses of your work?


☒ Yes ☐ No

Selected License

Attribution 4.0 International

This is a Free Culture License!



Help others attribute you!

This part is optional, but filling it out will add machine-readable metadata to the suggested HTML!

Title of work

Attribute work to name

Attribute work to URL


Source work URL

More permissions URL

Format of work

License mark

Have a web page?



This work is licensed under a [Creative Commons Attribution 4.0 International License](http://creativecommons.org/licenses/by/4.0/).

Copy this code to let your visitors know!

```
<a rel="license"
href="http://creativecommons.org/licenses/by/4.0/">
</a><br />This work is licensed under a <a rel="license"
href="http://creativecommons.org/licenses/by/4.0/">Creativ
a Commons Attribution 4.0 International License</a>
```

☒ Normal Icon ☐ Compact Icon

Recommended: CC-BY and CC0



Attribution CC BY

This license lets others distribute, remix, tweak, and build upon your work, even commercially, as long as they credit you for the original creation. This is the most accommodating of licenses offered. Recommended for maximum dissemination and use of licensed materials.

CC0 (datasets)



CC0 can be particularly important for the sharing of data and databases, since it otherwise may be unclear whether highly factual data and databases are restricted by copyright or other rights.

Databases may contain facts that, in and of themselves, are not protected by copyright law.

CC0 is recommended for data and databases and is used by hundreds of organizations. It is especially recommended for scientific data. Although CC0 doesn't legally require users of the data to cite the source, it does not take away the moral responsibility to give attribution, as is common in scientific research.

<http://creativecommons.org/licenses/>



Share: Data

figshare.com/

Highly connected drug file

| | | | |
|--------------------|-----|-----|---|
| Tretinoin | 257 | 46 | Rv1155, aroG, Rv1264, mscL, thyX, gmk, glnA1, Rv1264, thyX, glnA1, trpD, leuA, blaI, ethR, Rv0223c, lipJ, Rv1264, ephG, blaI, ethR, sigC, b |
| Levothyroxine | 173 | 36 | 25 cyp130, Rv1264, lppX, gpml, ligA, nirA |
| Methotrexate | 156 | 32 | TB31.7, Rv1264, mscL, lppX, coaA, pcaA, Rv3676, fabG1, |
| 4-Hydroxytamoxifen | 115 | 115 | mmaA4, bphD, Rv1264, mscL, thyX, lppX, mmaA2, ptl |
| Estradiol | 98 | 20 | TB31.7, cyp130, aroG, Rv1264, secA1, trpD, nirA |
| Amantadine | 79 | 1 | pth, ethR, clpP, glbN, inhA, |
| Rifampin | 78 | 13 | pknD, lipJ, fabH, Rv1941, Rv3361c, Rv1264, lppX |
| Raloxifene | 75 | 18 | mmaA4, Rv1264, groEL, lppX, secA1, glmU, nusA, R |
| Propofol | 54 | 5 | mmaA4, Rv1264, thyX, lppX, secA1, serA1, Rv3529c |
| Indinavir | 51 | 14 | pepD, Rv1264, thyX, ethR, trxB2, |
| Penicillamine | 44 | 10 | pknD, pepD, fabH, Rv1941, devB, ppp, ftsZ, cyp12! |
| Daunorubicin | 44 | 12 | |
| Triclosan | 42 | 5 | |
| Darunavir | 40 | 15 | |

[Enlarge to see the rest of the document](#)

[Enlarge](#) [Download](#)

Categories

- Computational Biology

Authors

Daniel Garijo
Lei Xie
Yinliang Zhang
Yolanda Gil
Li Xie
Sarah Kinnings
Phil Bourne

Tags

- results
- tb-drugome

License [\(what's this?\)](#)

CC-BY


Cite this: Garijo, Daniel; Xie, Lei; Zhang, Yinliang; Gil, Yolanda; Xie, Li; Kinnings, Sarah; Bourne, Phil (2013): Highly connected drug file. figshare. <http://dx.doi.org/10.6084/m9.figshare.776887>
Retrieved 08:56, Feb 20, 2015 (GMT)

Description

Highly connected drug file obtained as a result of the TB-Drugome Workflow.

Links

- <http://purl.org/net/tb-drugome-run>



1. Publication in a shared repository

2. General and domain metadata

3. Accessibility of data (domain and machine)



Share: Data

figshare.com/

Highly connected drug file

| | | | |
|--------------------|-----|-----|---|
| Tretinoin | 257 | 46 | Rv1155, aroG, Rv1264, mscL, thyX, gmk, glnA1, Rv1264, thyX, glnA1, trpD, leuA, blaI, ethR, Rv0223c, lipJ, Rv1264, ephG, blaI, ethR, sigC, b |
| Levothyroxine | 173 | 36 | 25 cyp130, Rv1264, lppX, gpml, ligA, nirA |
| Methotrexate | 156 | 32 | TB31.7, Rv1264, mscL, lppX, coaA, pcaA, Rv3676, fabG1, |
| 4-Hydroxytamoxifen | 115 | 115 | mmaA4, bphD, Rv1264, mscL, thyX, lppX, mmaA2, ptl |
| Estradiol | 98 | 20 | TB31.7, cyp130, aroG, Rv1264, secA1, trpD, nirA |
| Amantadine | 79 | 1 | pth, ethR, clpP, glbN, inhA, |
| Rifampin | 78 | 13 | pknD, lipJ, fabH, Rv1941, Rv3361c, Rv1264, lppX |
| Raloxifene | 75 | 18 | mmaA4, Rv1264, groEL, lppX, secA1, glmU, nusA, R |
| Propofol | 54 | 5 | mmaA4, Rv1264, thyX, lppX, secA1, serA1, Rv3529c |
| Indinavir | 51 | 14 | pepD, Rv1264, thyX, ethR, trxB2, |
| Penicillamine | 44 | 10 | pknD, pepD, fabH, Rv1941, devB, ppp, ftsZ, cyp12! |
| Daunorubicin | 44 | 12 | |
| Triclosan | 42 | 5 | |
| Darunavir | 40 | 15 | |

[Enlarge to see the rest of the document](#)

[Enlarge](#) [Download](#)

[results](#) [tb-drugome](#)

Cite this: Garijo, Daniel; Xie, Lei; Zhang, Yinliang; Gil, Yolanda; Xie, Li; Kinnings, Sarah; Bourne, Phil (2013): Highly connected drug file. figshare. <http://dx.doi.org/10.6084/m9.figshare.776887> Retrieved 08:56, Feb 20, 2015 (GMT)

Description

Highly connected drug file obtained as a result of the TB-Drugome Workflow.

Links

- <http://purl.org/net/tb-drugome-run>

Categories


- Computational Biology

Authors

Daniel Garijo
Lei Xie
Yinliang Zhang
Yolanda Gil
Li Xie
Sarah Kinnings
Phil Bourne

Tags

License (what's this?)
CC-BY



1. Publication in a shared repository

2. General and domain metadata

3. Accessibility of data (domain and machine)

4. Unique Persistent Identifier



Share: Data

figshare.com/

Highly connected drug file

| | | | |
|--------------------|-----|----|--|
| Tretinoin | 257 | 46 | Rv1155, aroG, Rv1264, mscL, thyX, gmk, glnA1, Rv1264, thyX, glnA1, trpD, leuA, blaI, ethR, Rv0223c, lipJ, Rv1264, ephG, blaI, ethR, sigC, b, cyp130, Rv1264, lppX, gpml, ligA, nirA, TB31.7, Rv1264, mscL, lppX, coaA, pcaA, Rv3676, fabG1, mmaA4, bphD, Rv1264, mscL, thyX, lppX, mmaA2, ptl, TB31.7, cyp130, aroG, Rv1264, secA1, trpD, nirA, pth, ethR, clpP, glbN, inhA, pknD, lipJ, fabH, Rv1941, Rv3361c, Rv1264, lppX, mmaA4, Rv1264, groEL, lppX, secA1, glmU, nusA, Rv1264, thyX, lppX, secA1, serA1, Rv3529c, pepD, Rv1264, thyX, ethR, trxB2, pknD, pepD, fabH, Rv1941, devB, ppp, ftsZ, cyp12! |
| Levothyroxine | 173 | 36 | |
| Methotrexate | 156 | 32 | |
| 4-Hydroxytamoxifen | 115 | 25 | |
| Estradiol | 98 | 20 | |
| Amantadine | 79 | 1 | |
| Rifampin | 78 | 13 | |
| Raloxifene | 75 | 18 | |
| Propofol | 54 | 5 | |
| Indinavir | 51 | 14 | |
| Penicillamine | 44 | 10 | |
| Daunorubicin | 44 | 12 | |
| Triclosan | 42 | 5 | |
| Darunavir | 40 | 15 | |

[Enlarge to see the rest of the document](#)

[Enlarge](#) [Download](#)

[results](#) [tb-drugome](#)

Cite this: Garijo, Daniel; Xie, Lei; Zhang, Yinliang; Gil, Yolanda; Xie, Li; Kinnings, Sarah; Bourne, Phil (2013): Highly connected drug file. figshare. <http://dx.doi.org/10.6084/m9.figshare.776887> Retrieved 08:56, Feb 20, 2015 (GMT)

Description

Highly connected drug file obtained as a result of the TB-Drugome Workflow.

Links

- <http://purl.org/net/tb-drugome-run>

Categories


- Computational Biology

Authors

Daniel Garijo
Lei Xie
Yinliang Zhang
Yolanda Gil
Li Xie
Sarah Kinnings
Phil Bourne

Tags

License (what's this?)
CC-BY



1. Publication in a shared repository

2. General and domain metadata

3. Accessibility of data (domain and machine)

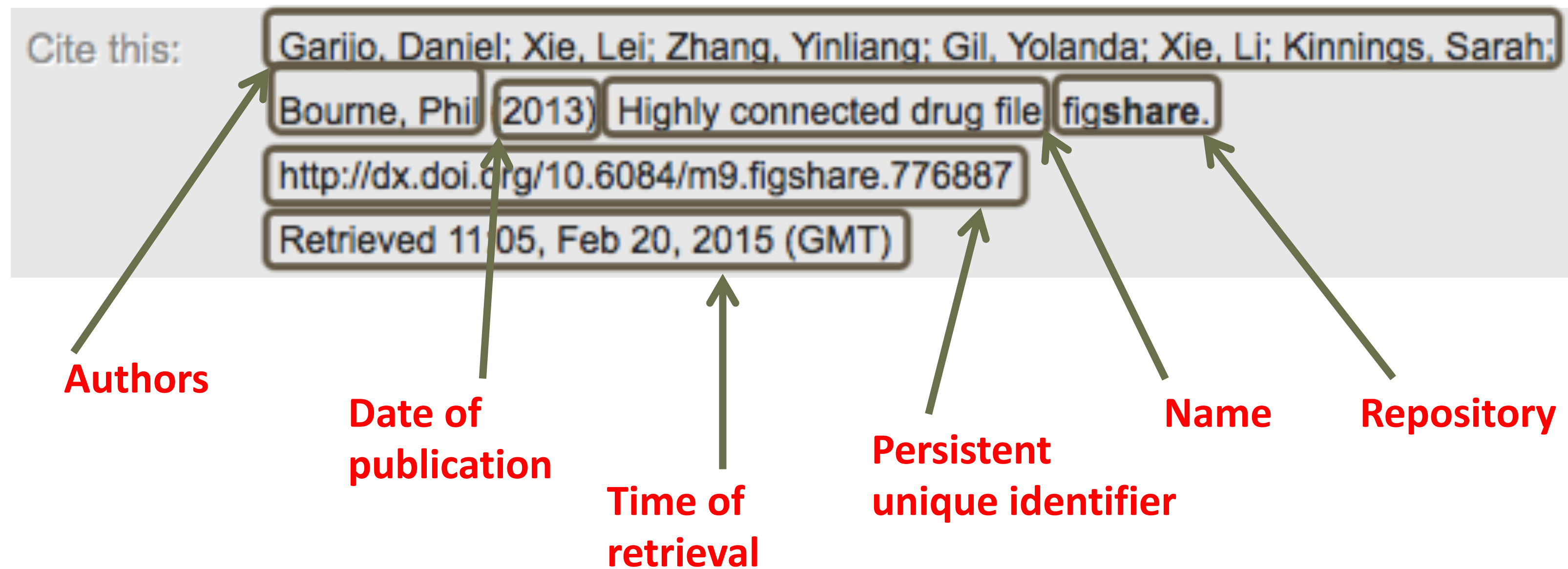
4. Unique Persistent Identifier

4. Citation preference



Share: Data

Citing Data: Data repositories and journals often specify how to cite data





WHY IS SCIENTIFIC SOFTWARE NOT SHARED?

- ▶ “No one would use my code if I shared it”
- ▶ “My code is really bad”
- ▶ “My code is not ready to be shared”
- ▶ “Sharing my software will take a lot of time”
- ▶ “I won’t get anything out of sharing my software”
- ▶ “I’ve shared software before, bad things happened”
- ▶ “I work for the government”
- ▶ “I want to commercialize my software”
- ▶ “I don’t want anyone to commercialize my software”
- ▶ “I don’t know where to start”



WHY IS SCIENTIFIC SOFTWARE NOT SHARED?

- ▶ “No one would use my code if I shared it”
- ▶ “My code is really bad”
- ▶ “My code is not ready to be shared”
- ▶ “Sharing my software will take a lot of time”
- ▶ “I won’t get anything out of sharing my software”
- ▶ “I’ve shared software before, bad things happened”
- ▶ “I work for the government”
- ▶ “I want to commercialize my software”
- ▶ “I don’t want anyone to commercialize my software”
- ▶ “I don’t know where to start”





BEST PRACTICES FOR MAKING SOFTWARE AVAILABLE

- ▶ Accessible from a public location
 - ▶ Software repository



Share: Code

BEST PRACTICES FOR MAKING SOFTWARE AVAILABLE

- ▶ Accessible from a public location
 - ▶ Software repository





BEST PRACTICES FOR MAKING SOFTWARE AVAILABLE

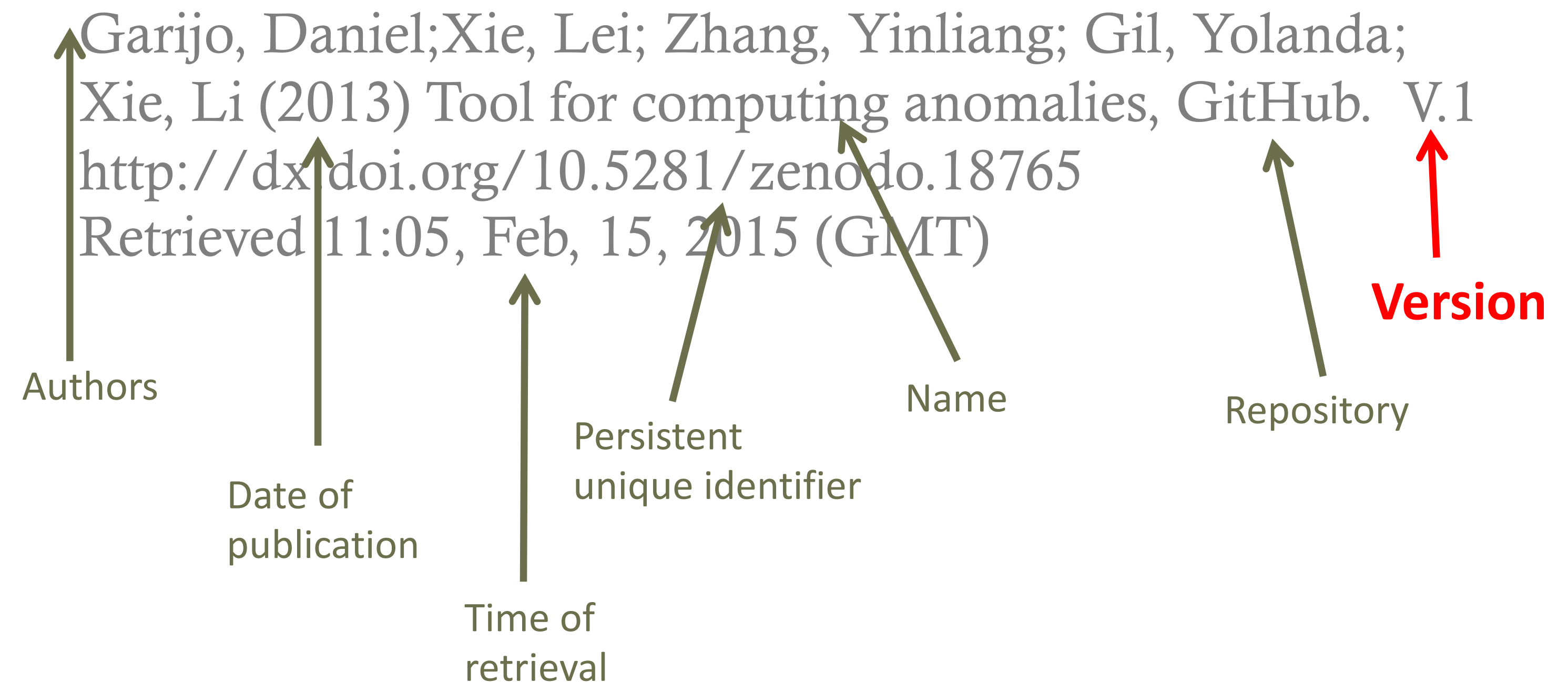
- ▶ Accessible from a public location
- ▶ License
 - ▶ Open source license: reduce constraints and enable software developers to make their source code available to the public
 - ▶ “Copyleft” license (GNU General Public License)
 - ▶ “Permissive” license



Share: Code

BEST PRACTICES FOR MAKING SOFTWARE AVAILABLE

- ▶ Accessible from a public location
- ▶ License
- ▶ Citation



<https://zenodo.org>



Share: Code

BEST PRACTICES FOR MAKING SOFTWARE AVAILABLE



- ▶ Accessible from a public location
- ▶ License
- ▶ Citation
- ▶ Executable via a zero-install environment (in the cloud)

Turn a Git repo into a collection of interactive notebooks

How it works

1

Enter your repository information

Provide in the above form a URL or a GitHub repository that contains Jupyter notebooks, as well as a branch, tag, or commit hash. Launch will build your Binder repository. If you specify a path to a notebook file, the notebook will be opened in your browser after building.

2

We build a Docker image of your repository

Binder will search for a dependency file, such as requirements.txt or environment.yml, in the repository's root directory ([more details on more complex dependencies in documentation](#)). The dependency files will be used to build a Docker image. If an image has already been built for the given repository, it will not be rebuilt. If a new commit has been made, the image will automatically be rebuilt.

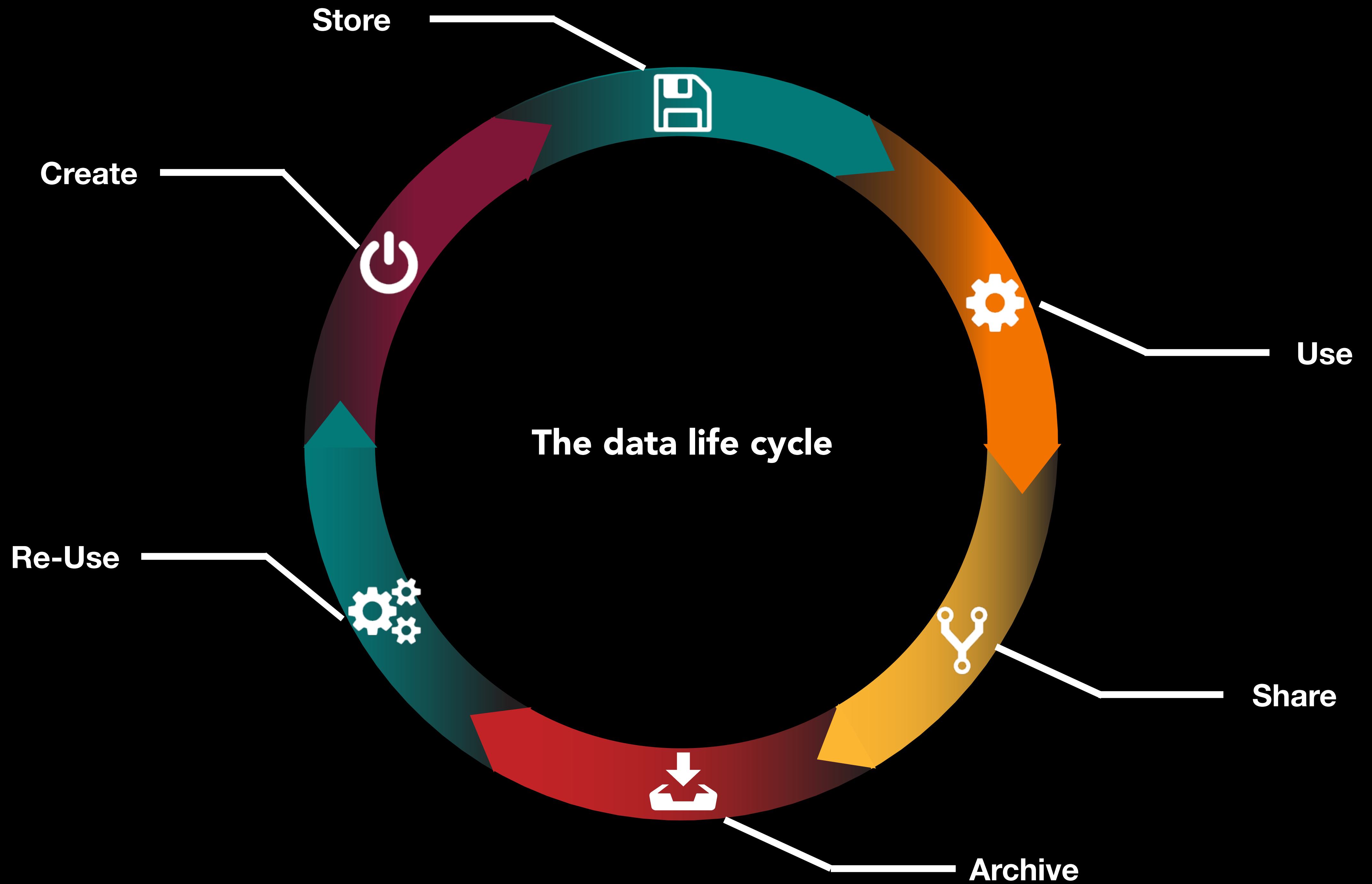
3

Interact with your notebooks in a live environment!

A [JupyterHub](#) server will host your repository's contents. We offer you a reusable link and badge to your live repository that you can easily share with others.

<https://mybinder.org>

<https://github.com/LinkedEarth/paleoHackathon>



Project management with GitHub

LinkedEarth / **Pyleoclim_util**

Unwatch ▾

10

Star

37

Fork

13

- <> Code
- Issues 18
- Pull requests
- Discussions
- Actions
- Projects 2
- Wiki
- Security
- Insights
- Settings

| <input type="checkbox"/> | 🕒 18 Open ✓ 93 Closed | Author ▾ | Label ▾ | Projects ▾ | Milestones ▾ | Assignee ▾ | Sort ▾ |
|--------------------------|--|----------|---------|------------|--------------|------------|--------|
| <input type="checkbox"/> | <div><div>🕒</div><div>Allow user to change archive type in lipd object</div><div>enhancement help wanted low priority</div></div> <div>#159 opened 13 days ago by Myunghee13</div> | | | | | | |
| <input type="checkbox"/> | <div><div>🕒</div><div>Halloween theme</div><div>low priority</div></div> <div>#157 opened on Jul 5 by CommonClimate</div> | | | | | | |
| <input type="checkbox"/> | <div><div>🕒</div><div>MultipleSeries plots</div><div>enhancement</div></div> <div>#154 opened on May 14 by Mattriks ➡ Pyleoclim paper</div> | | | | | | 3 |
| <input type="checkbox"/> | <div><div>🕒</div><div>cwt wavelet: amplitude scaling</div></div> <div>#153 opened on Apr 29 by Mattriks ➡ Pyleoclim paper</div> | | | | | | 3 |

Project management with GitHub

LinkedEarth / Pyleoclim_util

Unwatch ▾

10

Star

37

Fork

13

- <> Code
- ⦿ Issues 18
- 🔗 Pull requests
- 💬 Discussions
- ▶ Actions
- 📁 Projects 2
- 📖 Wiki
- 🛡 Security
- 📈 Insights
- ⚙ Settings


Allow user to change archive type in lipd object #159

Edit

New issue

🕒 Open

Myunghee13 opened this issue 13 days ago · 0 comments


 Myunghee13 commented 13 days ago

Is your feature request related to a problem? Please describe.

in the case of archive type is unknown or guessed, default will be other but needs to change this if more information is available.

Describe the solution you'd like

write a method in the lipd class to overwrite archive type in the lipd

 khider added enhancement help wanted low priority labels 13 days ago

Assignees

No one—assign yourself

Labels

enhancement help wanted low priority

Projects

📁 Development status

Awaiting triage ▾

Milestone

No milestone

Linked pull requests


Successfully merging a pull request may close this issue.

None yet

Notifications

Customize

🔔 Unsubscribe




Write

Preview

H B I ≡ <> 🔗 ≡ ≡ ☑ @ ↗ ↶ ▾

Leave a comment

Attach files by dragging & dropping, selecting or pasting them. 

👍 Close issue

Comment

Project management with GitHub

LinkedEarth / Pyleoclim_util

Unwatch ▾

10

Star

37

Fork

13

- <> Code
- ⦿ Issues 18
- 🔗 Pull requests
- 💬 Discussions
- ▶ Actions
- 📁 Projects 2
- 📖 Wiki
- 🛡 Security
- 📈 Insights
- ⚙ Settings


Allow user to change archive type in lipd object #159

Edit

New issue

🟢 Open

 Myunghee13 opened this issue 13 days ago · 0 comments


 Myunghee13 commented 13 days ago

Is your feature request related to a problem? Please describe.

in the case of archive type is unknown or guessed, default will be other but needs to change this if more information is available.

Describe the solution you'd like

write a method in the lipd class to overwrite archive type in the lipd

 khider added enhancement help wanted low priority labels 13 days ago


Assignees

No one—assign yourself

Labels

enhancement help wanted low priority

Projects

 **Development status**

Awaiting triage ▾

Milestone

No milestone

Linked pull requests


Successfully merging a pull request may close this issue.










None yet

Notifications


Customize

Unsubscribe

 Write Preview

H B I         

Leave a comment

Attach files by dragging & dropping, selecting or pasting them. 

🔍 Close issue

Comment

Project management with GitHub

LinkedEarth / Pyleoclim_util

Unwatch

10

Star

37

Fork

13

<> Code

Issues 18

Pull requests

Discussions

Actions

Projects 2

Wiki

Security

Insights

Settings

LinkedEarth / Pyleoclim_util

Unwatch

10

Star

37

Fork

13

<> Code

Issues 18

Pull requests

Discussions

Actions

Projects 2

Wiki

Security

Insights

Settings

Development status

Updated 10 days ago

Filter cards

+ Add cards(1 new)

Fullscreen

Menu

3 In progress

cwt outputs don't compare to wwz

#135 opened by khider

bug enhancement

Pyleoclim paper

Create guidelines for contributing to pyleoclim

#110 opened by khider

documentation low priority

Pyleoclim paper

Lipd unit tests

#158 opened by khider

testing

Pyleoclim paper

5 To do (next)

Need method arguments in Series.summary_plot()

#84 opened by fzh2e

enhancement

Pyleoclim paper

Age model plot

#2 opened by CommonClimate

Pyleoclim paper

Wiki integration - Notebook from Daniel

#50 opened by khider

enhancement

Pyleoclim paper

MultipleSeries plots

#154 opened by Matriks

enhancement

9 To do

Research-grade example for MSSA

#85 opened by khider

documentation enhancement

Enable MSSA

#129 opened by khider

enhancement

Gap detection function -> split

#81 opened by khider

enhancement

A shortcut preprocessing method: Series.preprocess()

#107 opened by fzh2e

enhancement low priority

Handle ensembles for cross-wavelet

40 Done

ReadTheDocs does not index methods

#146 opened by CommonClimate

documentation

reenable mapNearRecord from previous versions of Pyleoclim

#90 opened by khider

Pyleoclim paper

Scree plot method

#136 opened by CommonClimate

enhancement

Pyleoclim paper

Lomb-Scargle method return error for a specific series

#143 opened by khider

bug

Display a menu

Project management with GitHub

LinkedEarth / Pyleoclim_util

Unwatch ▾

10

Star 37

Fork 13

- <> Code
- Issues 18
- Pull requests
- Discussions
- Actions
- Projects 2
- Wiki
- Security
- Insights
- Settings


Allow user to change archive type in lipd object #159

Edit

New issue

Open

Myunghee13 opened this issue 13 days ago · 0 comments


 Myunghee13 commented 13 days ago

Is your feature request related to a problem? Please describe.

in the case of archive type is unknown or guessed, default will be other but needs to change this if more information is available.

Describe the solution you'd like

write a method in the lipd class to overwrite archive type in the lipd

 khider added enhancement help wanted low priority labels 13 days ago

Write

Preview

H B I ≡ <> 🔗 ≡ ≡ ☑ @ ↗ ↶

Leave a comment

Attach files by dragging & dropping, selecting or pasting them.

Close issue

Comment

Assignees

No one—assign yourself

Labels

enhancement help wanted low priority

Projects

Development

Waiting triage ▾

Milestone

No milestone

Linked pull requests

Successfully merged pull request may close this issue.

None yet

Notifications

Customize

Unsubscribe

Project management with GitHub

LinkedEarth / Pyleoclim_util

Unwatch ▾

10

Star

37

Fork

13

- <> Code
- Issues 18
- Pull requests
- Discussions
- Actions
- Projects 2
- Wiki
- Security
- Insights
- Settings

Labels

Milestones

New milestone

2 Open ✓ 4 Closed

Sort ▾

Pyleoclim paper

⚠ Past due by 3 months ⌚ Last updated about 3 hours ago

Submission of the Pyleoclim package to PP.



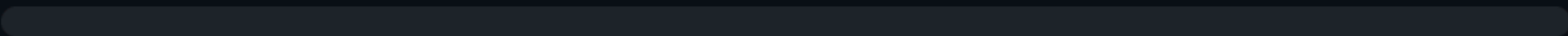
62% complete 6 open 10 closed

Edit Close Delete

Support functionalities for Big Paleo Data

📅 Due by August 31, 2021 ⌚ Last updated 5 months ago

Functions needed to be incorporated into Pyleoclim to support analysis tools needed for Big Paleo Data work.



0% complete 1 open 0 closed

Edit Close Delete

References

- Geoscience Paper of the Future:
 - Gil, Y., & . (Ed .) .. (2016, April 17). The Geoscience Paper of the Future: OntoSoft Training (Version 9). figshare. <https://doi.org/10.6084/m9.figshare.1586773.v9>

<http://www.scientificpaperofthefuture.org/gpf/>

Slides Availability

doi: 10.6084/m9.figshare.6510305

<https://figshare.com/s/999787b6f9f6416266b1>

License: CC BY 4.0

