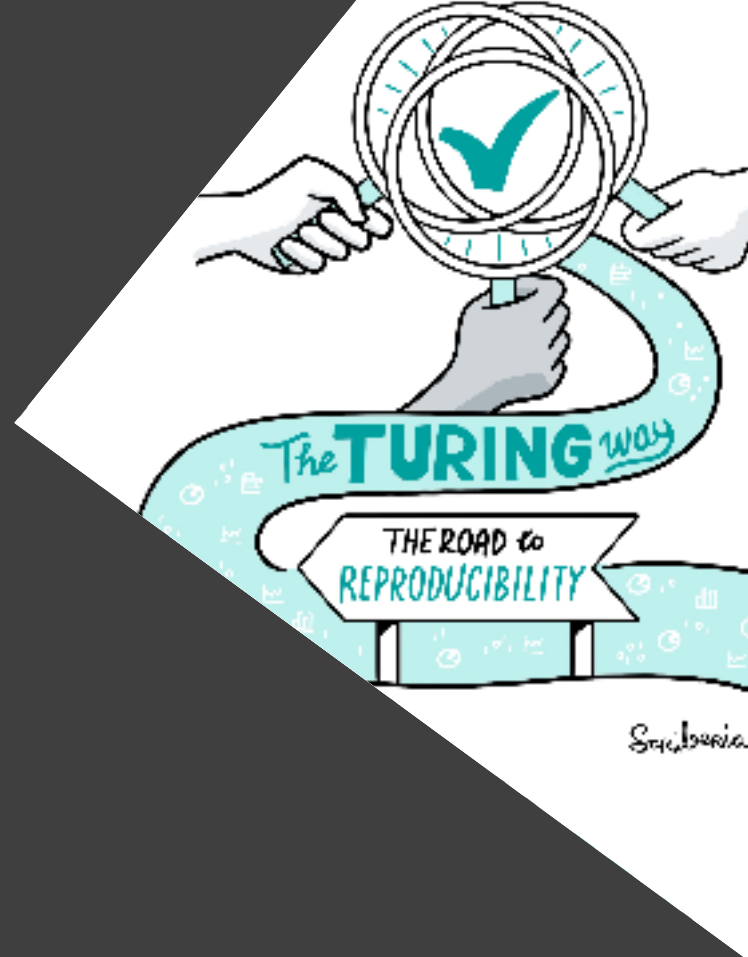


The
Alan Turing
Institute



Reproducible Research: Intro to *The Turing Way* and Binder

Dr Sarah Gibson



Agenda

- Talk 🗣️ (45 mins)
- Pizza! 🍕 (15 mins)
- Tutorial 🧑🏻💻 (45 mins)



The Turing Way is:

- a book
- a community
- a global collaboration
- a whole tonne of work



Rachael Ainsworth



Becky Arnold



Louise Bowler



Kirstie Whitaker



Patricia Herterich



James Hetherington



Rosie Higman



Anna Krystall



Catherine Lawrence



Alex Morley



Martin O'Reilly



Binder Team

Thank you to all our contributors



This project follows the all-contributors specifications. Contributions of any kind welcome!

<https://github.com/alan-turing-institute/the-turing-way#contributors>
<https://allcontributors.org/docs/en/emoji-key>
<https://doi.org/10.5281/zenodo.3628296>

#TuringWay

An Introduction to Me

- Research Software Engineer at the Turing
- *The Turing Way* developer
- Member of mybinder.org operating team
- 2020 Software Sustainability Institute Fellow



Academic errors have real world effects

	B	C	D	E	F	G	H
2			Real GDP growth				
3			relative to GDP				
4	Country	Coverage	30 or less	30 to 60	60 to 90	90 or above	50 or less
26			3.7	3.0	3.5	1.7	8.5
27	Minimum		1.8	0.3	1.3	-4.8	0.8
28	Maximum		5.4	4.6	10.2	3.6	13.3
29							
30	US	1946-2009	n.a.	3.4	3.2	-2.0	n.a.
31	UK	1946-2009	n.a.	2.4	2.5	2.4	n.a.
32	Sweden	1946-2009	3.8	2.9	2.7	n.a.	6.3
33	Spain	1946-2009	1.5	3.4	4.2	n.a.	9.9
34	Portugal	1952-2009	4.8	2.5	0.2	n.a.	7.9
35	New Zealand	1948-2009	2.5	2.9	3.9	-7.9	2.6
36	Netherlands	1956-2009	4.1	2.7	1.1	n.a.	6.4
37	Norway	1947-2009	3.4	5.1	n.a.	n.a.	5.4
38	Japan	1946-2009	7.0	4.0	1.0	0.7	7.0
39	Italy	1951-2009	5.4	2.1	1.8	1.0	5.6
40	Ireland	1918-2009	4.4	4.5	4.0	2.4	2.9
41	Greece	1970-2009	4.0	0.3	2.7	2.9	13.3
42	Germany	1946-2009	3.9	0.9	3.6	n.a.	3.2
43	France	1949-2009	4.9	2.7	3.0	n.a.	5.2
44	Finland	1946-2009	3.8	2.4	5.5	n.a.	7.0
45	Denmark	1950-2009	3.3	1.7	2.4	n.a.	5.6
46	Canada	1951-2009	1.9	3.6	4.1	n.a.	7.2
47	Hungary	1947-2009	n.a.	4.7	3.1	2.6	n.a.
48	Austria	1946-2009	5.2	3.3	-3.8	n.a.	9.7
49	Australia	1951-2009	3.2	4.6	4.0	n.a.	9.0
50							
51			4.3	2.9	2.9	=AVERAGE(L10:L14)	

<https://statmodeling.stat.columbia.edu/2013/04/16/memo-to-reinhart-and-rogoft-i-think-its-best-to-admit-your-errors-and-go-on-from-there>

<https://www.bbc.co.uk/news/magazine-22223190>

<https://doi.org/10.5281/zenodo.3628296>

#TuringWay

Academic errors have real world effects

	B	C	D	E	F	G	H
2							
3							
4	Country	Coverage	30 or less	30 to 60	60 to 90	90 or above	50 or less
26			3.7	3.0	3.5	1.7	8.5
27	Minimum		1.8	0.3	1.3	-4.8	0.8
28	Maximum		5.4	4.6	10.2	3.6	13.3
29							
30	US	1946-2009	n.a.	3.4	3.2	-2.0	n.a.
31	UK	1946-2009	n.a.	2.4	2.5	2.4	n.a.
32	Sweden	1946-2009	3.8	2.9	2.7	n.a.	6.3
33	Spain	1946-2009	1.5	3.4	4.2	n.a.	9.9
34	Portugal	1952-2009	4.8	2.5	0.3	n.a.	7.9
35	New Zealand	1948-2009	2.5	2.9	3.9	-7.9	2.6
36	Netherlands	1956-2009	4.1	2.7	1.1	n.a.	6.4
37	Norway	1947-2009	3.4	5.1	n.a.	n.a.	5.4
38	Japan	1946-2009	7.0	4.0	1.0	0.7	7.0
39	Italy	1951-2009	5.4	2.1	1.8	1.0	5.6
40	Ireland	1918-2009	4.4	4.5	4.0	2.4	2.9
41	Greece	1970-2009	4.0	0.3	2.7	2.9	13.3
42	Germany	1946-2009	3.9	0.9	3.6	n.a.	3.2
43	France	1949-2009	6.9	2.7	3.0	n.a.	5.2
44	Finland	1946-2009	3.8	-2.4	5.5	n.a.	7.0
45	Denmark	1950-2009	3.3	1.7	2.4	n.a.	5.6
46	Canada	1951-2009	1.9	3.6	4.1	n.a.	7.2
47	Hungary	1947-2009	n.a.	4.7	3.1	-2.6	n.a.
48	Austria	1948-2009	-5.1	3.3	-3.8	n.a.	9.7
49	Australia	1951-2009	3.2	4.6	4.0	n.a.	9.0
50							
51			4.1	2.9	2.9	=AVERAGE(L10:L44)	

BBC NEWS

Magazine

Reinhart, Rogoff... and Herndon: The student who caught out the profs

By Paul Fennell
BBC News

20 April 2013

This week, economists have been asked to admit that their theories, and the paper often used to make them, contain major errors. Another surprise is that these mistakes, by two named Harvard professors, were spotted by a student doing his homework.

By 4 January 2013, the Harvard University professorial team of Ben Bernanke, David Colander and Andrew Ross had been asked to admit that the theories of Reinhart and Rogoff, and the paper often used to make them, contain major errors.



<https://statmodeling.stat.columbia.edu/2013/04/16/memo-to-reinhart-and-rogooff-i-think-its-best-to-admit-your-errors-and-go-on-from-there>
<https://www.bbc.co.uk/news/magazine-22223190>

<https://doi.org/10.5281/zenodo.3628296>

#TuringWay

The humans are the
hardest part of
reproducibility



		Data	
		Same	Different
Analysis	Same	Reproducible	Replicable
	Different	Robust	Generalisable

Kirstie Whitaker's talk at PyData LDN: <https://youtu.be/IG3PcZ6EhiU>

<https://the-turing-way.netlify.com/reproducibility/03/definitions.html>

<https://doi.org/10.5281/zenodo.3628296>

#TuringWay

		Data	
		Same	Different
Analysis	Same	<div>Repeatable</div> Reproducible	Replicable
	Different	Robust	Generalisable

Kirstie Whitaker's talk at PyData LDN: <https://youtu.be/IG3PcZ6EhiU>

<https://the-turing-way.netlify.com/reproducibility/03/definitions.html>

<https://doi.org/10.5281/zenodo.3628296>

#TuringWay

Is not considered
for promotion

Held to higher
standards than
others

Publication bias
towards novel
findings

Barriers to reproducible research

Requires
additional
skills

Plead the 5th

Support additional
users

Takes time

The Turing Institute

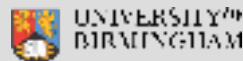
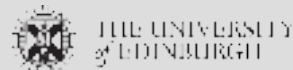
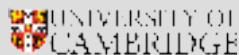


<https://www.turing.ac.uk/news/enigma-machine-goes-display-alan-turing-institute>

<https://doi.org/10.5281/zenodo.3628296>

#TuringWay

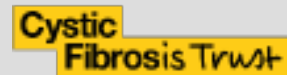
University network



The Institute's partners and collaborators

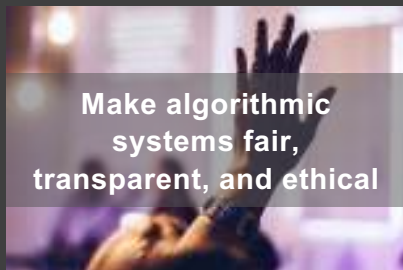


Lloyd's Register
Foundation



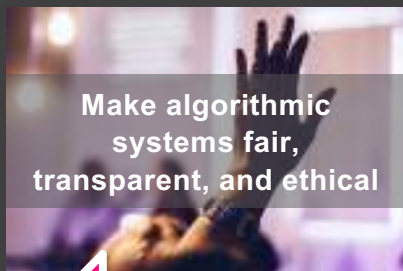
Challenges

Advance data science and artificial intelligence to...



Challenges

Advance data science and artificial intelligence to...



The Turing Way

The Turing Way



1. Introduction

2. Reproducibility

3. Open Research

4. Version Control

5. Collaborating on GitHub/GitLab

6. Research Data Management

7. Reproducible Environments

8. Testing

9. Reviewing

10. Continuous Integration

11. Reproducible Research with Make

12. Risk Assessment

Welcome to the Turing Way

The Turing Way is a lightly opinionated guide to reproducible data science.

Our goal is to provide all the information that researchers need at the start of their projects to ensure that they are easy to reproduce at the end.

This also means making sure PhD students, postdocs, PIs and funding teams know which parts of the "responsibility of reproducibility" they can affect, and what they should do to nudge data science to being more efficient, effective and understandable.

A bit more background

Reproducible research is necessary to ensure that scientific work can be trusted. Funders and publishers are beginning to require that publications include access to the underlying data and the analysis code. The goal is to ensure that all results can be independently verified and built upon in future work. This is sometimes easier said than done. Sharing these research outputs means understanding data management, library sciences, software development, and continuous integration techniques: skills that are not widely taught or expected of academic researchers and data scientists.

The Turing Way is a handbook to support students, their supervisors, funders and journal editors

<https://the-turing-way.netlify.com/introduction/introduction>

<https://doi.org/10.5281/zenodo.3628296>

#TuringWay

1. Introduction

2. Reproducibility

3. Open Research

4. Version Control

5. Collaborating on GitHub/GitLab

6. Research Data Management

7. Reproducible Environments

8. Testing

9. Reviewing

10. Continuous Integration

11. Reproducible Research with Make

12. Risk Assessment



Welcome to the Turing Way

The Turing Way is a lightly opinionated guide to reproducible data science.

Our goal is to provide all the information that researchers need at the start of their projects to ensure that they are easy to reproduce at the end.

This also means making sure PhD students, postdocs, PIs and funding teams know which parts of the "responsibility of reproducibility" they can affect, and what they should do to nudge data science to being more efficient, effective and understandable.

A bit more background

Reproducible research is necessary to ensure that scientific work can be trusted. Funders and publishers are beginning to require that publications include access to the underlying data and the analysis code. The goal is to ensure that all results can be independently verified and built upon in future work. This is sometimes easier said than done. Sharing these research outputs means understanding data management, library sciences, software development, and continuous integration techniques: skills that are not widely taught or expected of academic researchers and data scientists.

The Turing Way is a handbook to support students, their supervisors, funders and journal editors

<https://the-turing-way.netlify.com/introduction/introduction>

<https://doi.org/10.5281/zenodo.3628296>

#TuringWay

1. Introduction

2. Reproducibility

3. Open Research

4. Version Control

5. Collaborating on GitHub/GitLab

6. Research Data Management

7. Reproducible Environments

8. Testing

9. Reviewing

10. Continuous Integration

11. Reproducible Research with Make

12. Risk Assessment



Welcome to the Turing Way

The Turing Way is a lightly opinionated guide to reproducible data science.

Our goal is to provide all the information that researchers need at the start of their projects to ensure that they are easy to reproduce at the end.

This also means making sure PhD students, postdocs, PIs and funding teams know which parts of the "responsibility of reproducibility" they can affect, and what they should do to nudge data science to being more efficient, effective and understandable.

A bit more background

Reproducible research is necessary to ensure that scientific work can be trusted. Funders and publishers are beginning to require that publications include access to the underlying data and the analysis code. The goal is to ensure that all results can be independently verified and built upon in future work. This is sometimes easier said than done. Sharing these research outputs means understanding data management, library sciences, software development, and continuous integration techniques: skills that are not widely taught or expected of academic researchers and data scientists.

The Turing Way is a handbook to support students, their supervisors, funders and journal editors



Is not considered
for promotion

Held to higher
standards than
others

Publication bias
towards novel
findings

Requires
additional
skills

Barriers to reproducible research

Plead the 5th

Support additional
users

Takes time

A global collaboration



Patricia Herterich

“What really sets The Turing Way apart is HOW we’re writing the book. The focus on community, the commitment to transparency and working open right from the beginning is an exciting (and terrifying) new way of working.”



Open Leadership Principles



Understanding

You make the work accessible and clear

Read more

<https://mozilla.github.io/olm-whitepaper>



Sharing

You make the work easy to adapt, reproduce, and spread

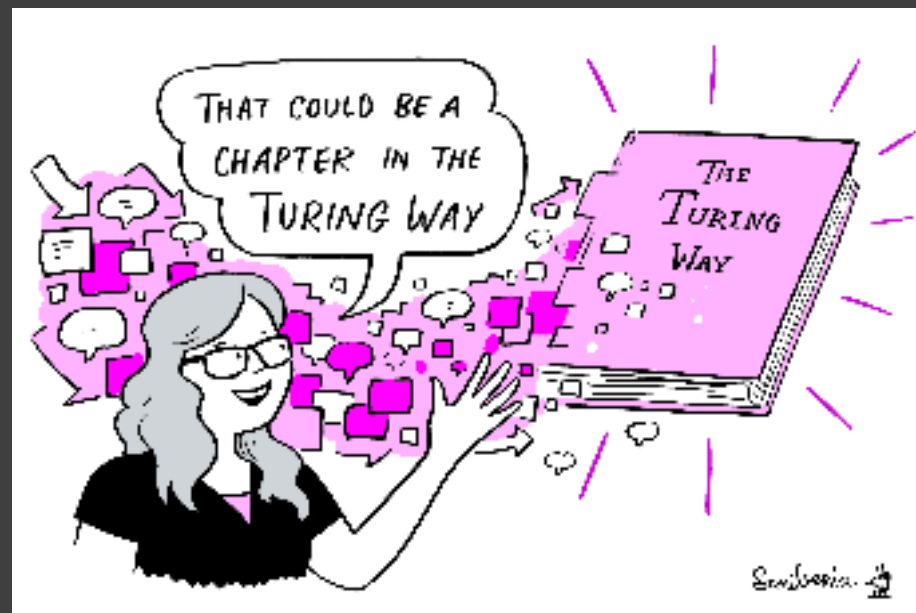


Participation & Inclusion

You build shared ownership and agency to make the work inviting and sustainable for all.

Funding extension

- Expand scope to all data science practices
 - Ethics, model selection, project management, collaborative working
- Full time community manager, contributions from Turing & beyond



[https://github.com/
alan-turing-institute/the-turing-way/
blob/master/project_management/
tps-funding-application-20190429.md](https://github.com/alan-turing-institute/the-turing-way/blob/master/project_management/tps-funding-application-20190429.md)

Metrics for success

- 20 new chapters
- 100 authors
- 200 contributors
- 1000 mailing list subscribers
- 50 first pull requests
- 20 new contributors to other open source projects



Market Research



Have you ever heard...?

*“Oh, it worked on
my computer?”*

Have you ever heard...?

*“Oh, it worked
yesterday?”*



*“Oh, it worked on
my computer?”*

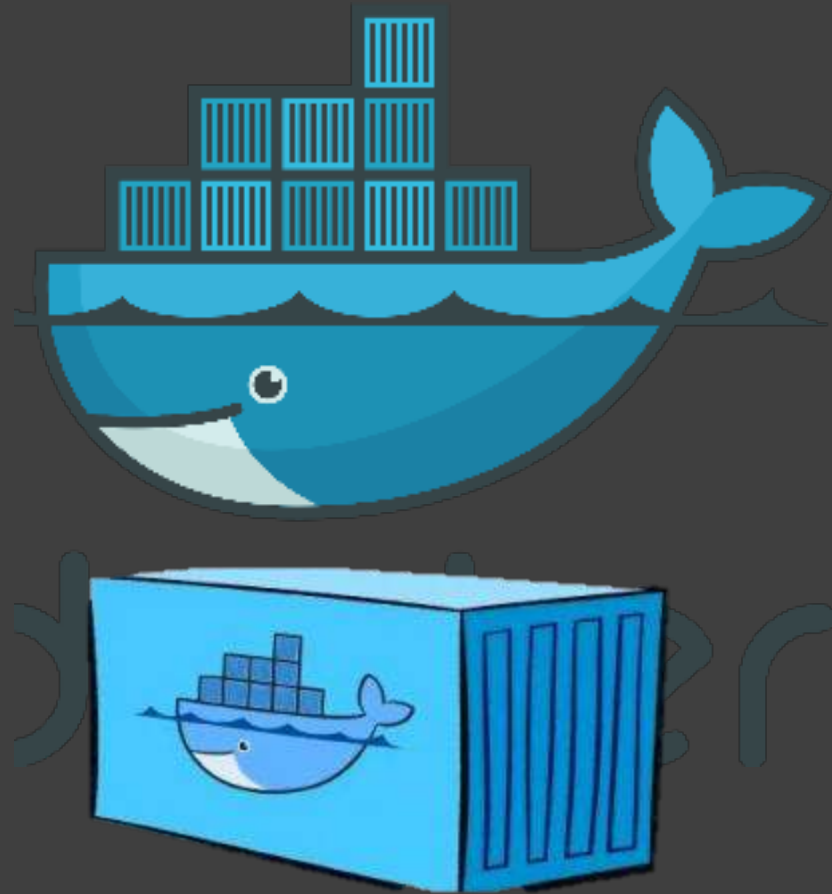


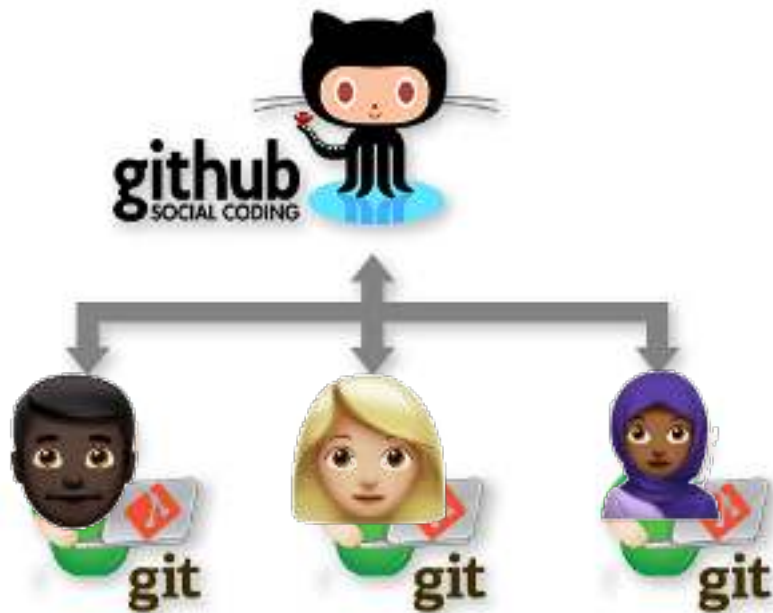
+ CI

*“Oh, it worked
yesterday?”*

What's Docker?

- A way to “containerise” software
- Bundles together infrastructure, code and data
- You don't have to worry about the different moving parts – just use the same container!



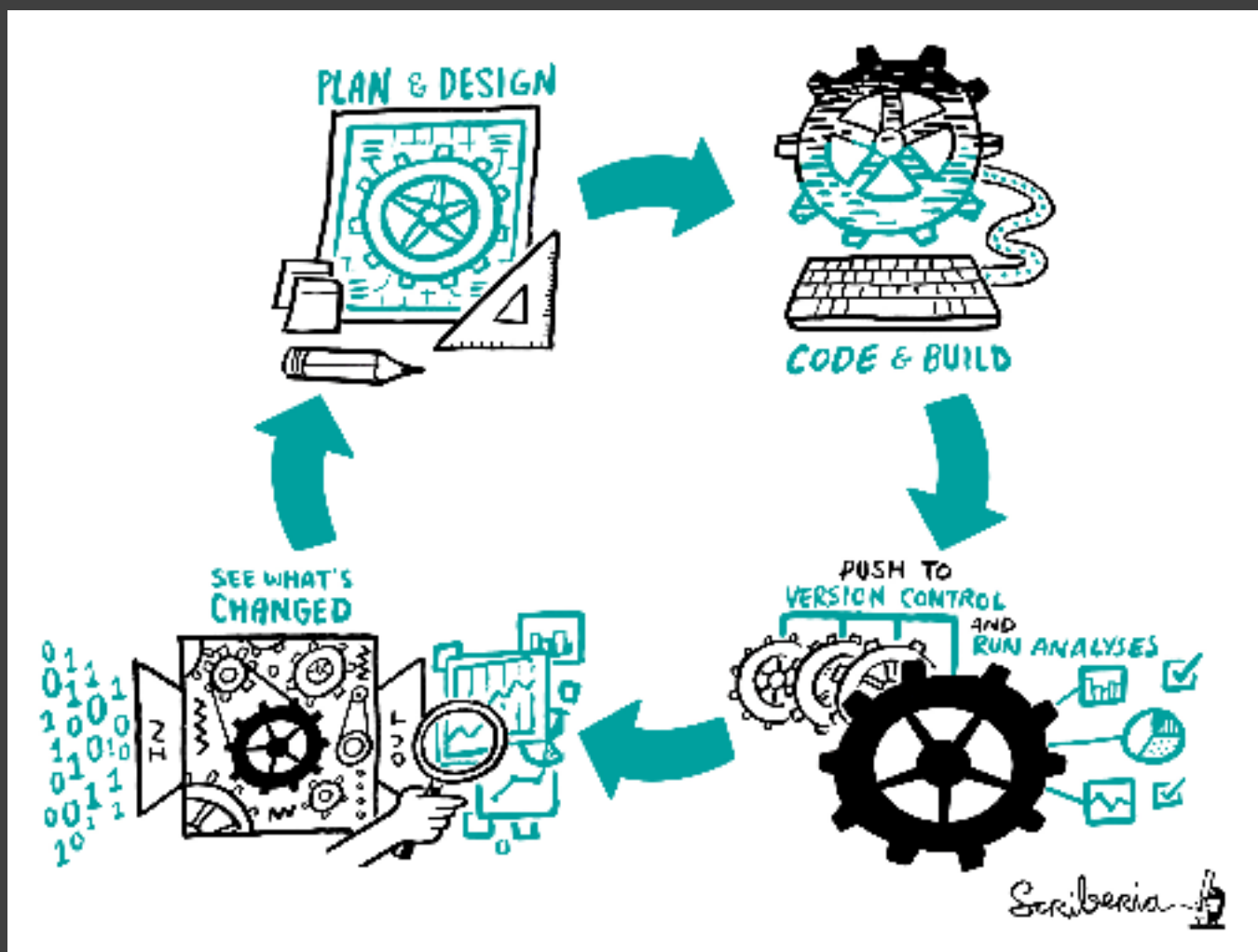


https://the-turing-way.netlify.com/version_control/version_control.html
https://the-turing-way.netlify.com/collaborating_github/collaborating_github.html
<http://phdcomics.com/comics/archive/phd101212s.gif>

<https://doi.org/10.5281/zenodo.3628296>



#TuringWay



Turing Way & Binder



The Vocab

- **Binder** → user interface/experience
- **BinderHub** → computational infrastructure
- **mybinder.org** → public BinderHub for everyone

[mybinder.org Usage Guidelines](#)[Frequently Asked Questions](#)[Status of mybinder.org](#)[About mybinder.org](#)[mybinder.org status and reliability](#)[More information about Binder](#)

What is **mybinder.org**?

mybinder.org is a single deployment of a BinderHub instance, managed by the Binder community. It serves as both a public service and a demonstration of the BinderHub technology, though it is by no means the only BinderHub in existence. If you're interested in deploying your own BinderHub for your own uses, please see the [BinderHub documentation](#) and don't hesitate to reach out to the [Binder community](#).

For more information, check out [About mybinder.org](#).

Is **mybinder.org** free to use?

Yes! Though note that it has relatively [limited computational resources](#).

How much does running **mybinder.org** cost?

Great question! If you're interested in the technical costs of running **mybinder.org**, we publish a semi-up-to-date dataset of our costs at the [binder-data](#) repository. In addition, you can explore these costs with the binder link below!

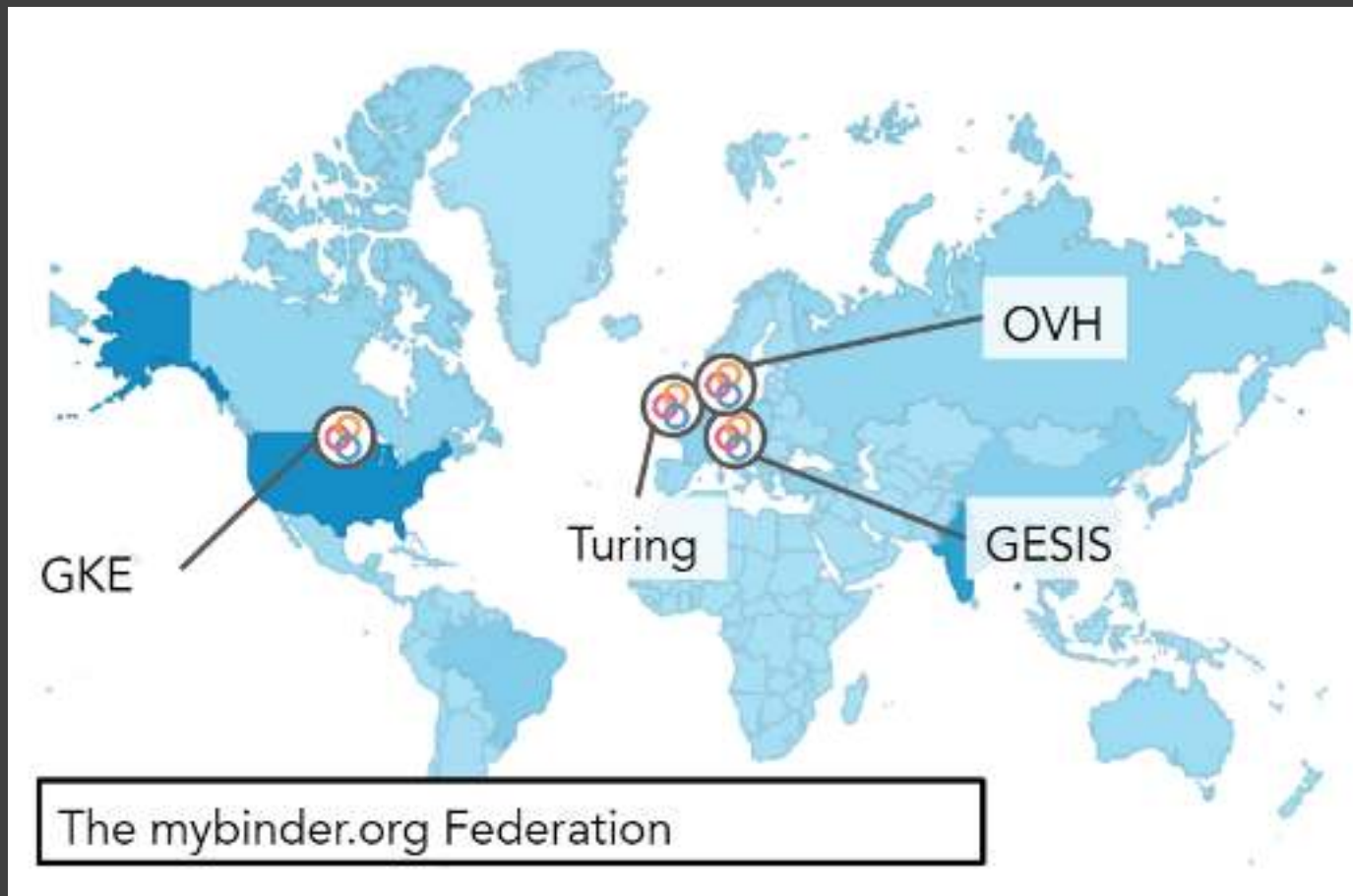
[Launch Binder](#)

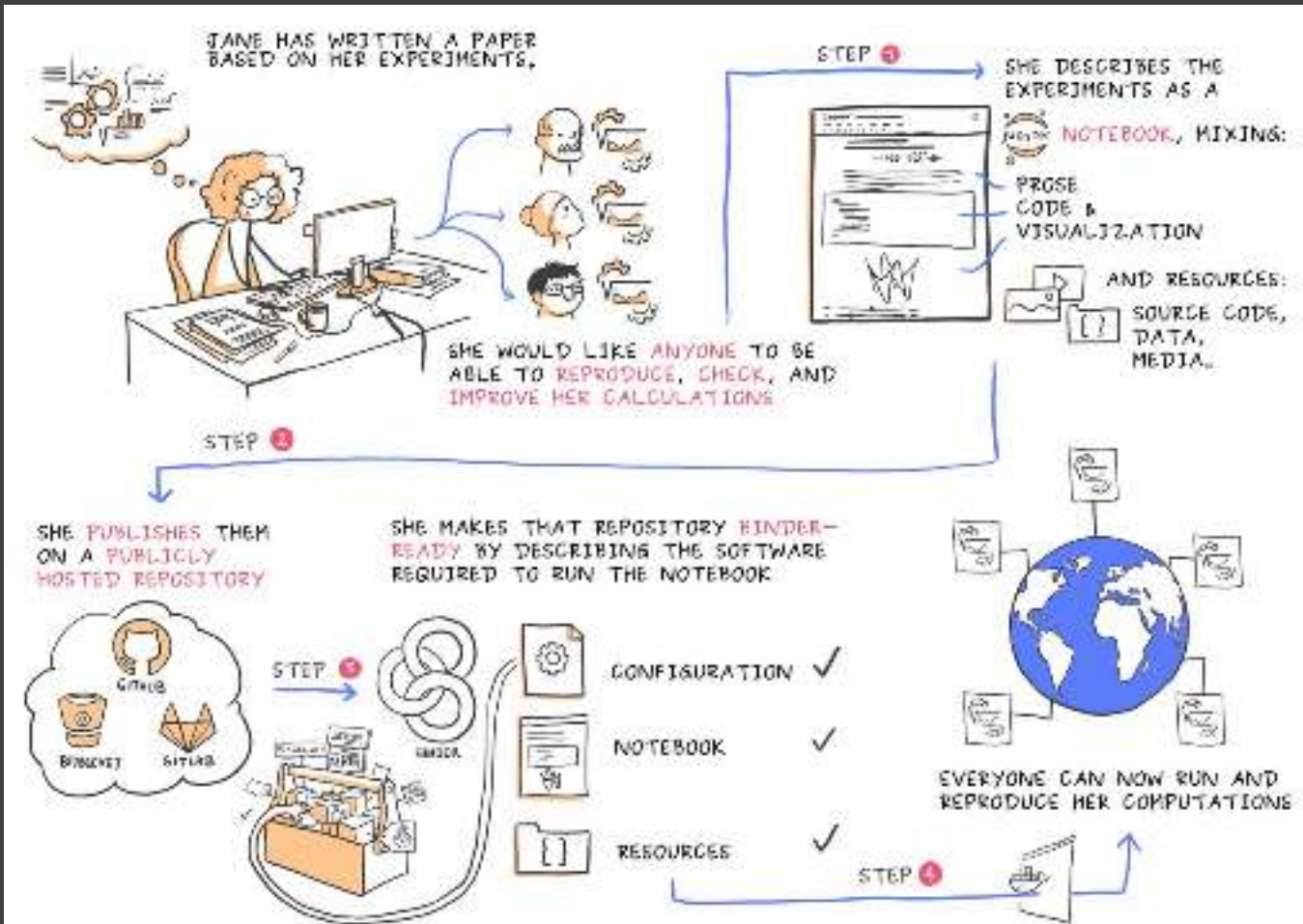
How can **mybinder.org** be free to use?

[✎ Edit this page](#)

On this page

[What is a Binder?](#)[What is the Binder community?](#)[What is BinderHub?](#)**[What is mybinder.org?](#)**[Is mybinder.org free to use?](#)[How much does running mybinder.org cost?](#)[How can mybinder.org be free to use?](#)[How much memory am I given when using Binder?](#)[How long will my Binder session last?](#)[Can I use mybinder.org for a live demo or workshop?](#)[How does mybinder.org ensure user privacy?](#)





mybinder.org

Courtesy of Juliette Taka: <https://twitter.com/mybinderteam/status/1082556317842264064>

<https://doi.org/10.5281/zenodo.3628296>

#TuringWay

<> Code

Issues 0

Pull requests 0

Projects 0

Wiki

Security

Insights

Branch: master

requirements / requirements.txt

Find file

Copy path

choldgraf Update requirements.txt

21a328d on 21 Jun

2 contributors

5 lines (3 sloc) 46 Bytes

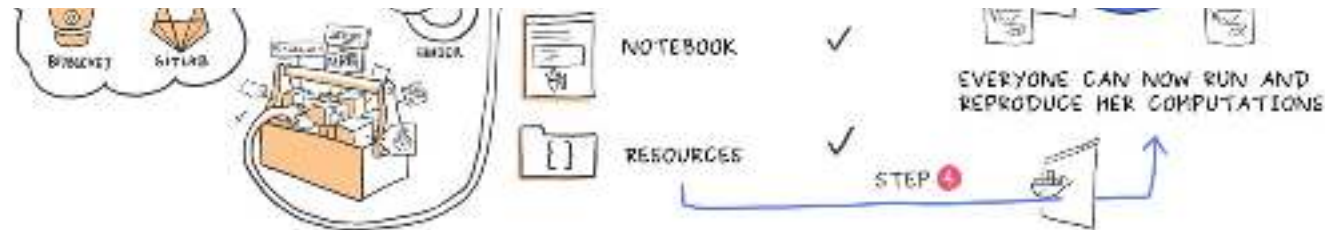
Raw

Blame

History



```
1 numpy==1.16.*
2 matplotlib==3.*
3 seaborn==0.8.1
4
```



<> Code

Issues 2

Pull requests 0

Projects 0

Wiki

Security

Insights

Branch: master

conda / environment.yml

Find file

Copy path

 betatim Update environment.yml

89dd429 on 11 Dec 2018

4 contributors



14 lines (13 sloc) 161 Bytes

Raw

Blame

History



```
1 name: example-environment
2 channels:
3   - conda-forge
4 dependencies:
5   - numpy
6   - psutil
7   - toolz
8   - matplotlib
9   - dill
10  - pandas
11  - partd
12  - bokeh
13  - dask
```

Courtesy of Juliette Taka: <https://twitter.com/mybinderteam/status/1082556317842264064><https://doi.org/10.5281/zenodo.3628296>

#TuringWay

<> Code

Issues 0

Pull requests 0

Projects 0

Wiki

Security

Insights

Branch: master

binder-r-description / DESCRIPTION

Find file

Copy path



gedankenstuecke first commit

70f8b8e on 18 Sep 2018

1 contributor

8 lines (7 sloc) 282 Bytes

Raw

Blame

History



```
1 Package: binderdescription
2 Version: 0.1
3 Date: 2018-09-18
4 Title: Binder R DESCRIPTION support
5 Description: Test that automatically building R packages works
6 Author: Bastian Greshake Tzovaras <bgresshake@googlemail.com>
7 Maintainer: Bastian Greshake Tzovaras <bgresshake@googlemail.com>
```



<> Code

① Issues 3

🔗 Pull requests 1

📁 Projects 0

📖 Wiki

🛡 Security

📊 Insights

Branch: master ▾

r / install.R

Find file

Copy path



betatim Add example Shiny app

8c01f0d on 31 May 2018

4 contributors



6 lines (5 sloc) 148 Bytes

Raw

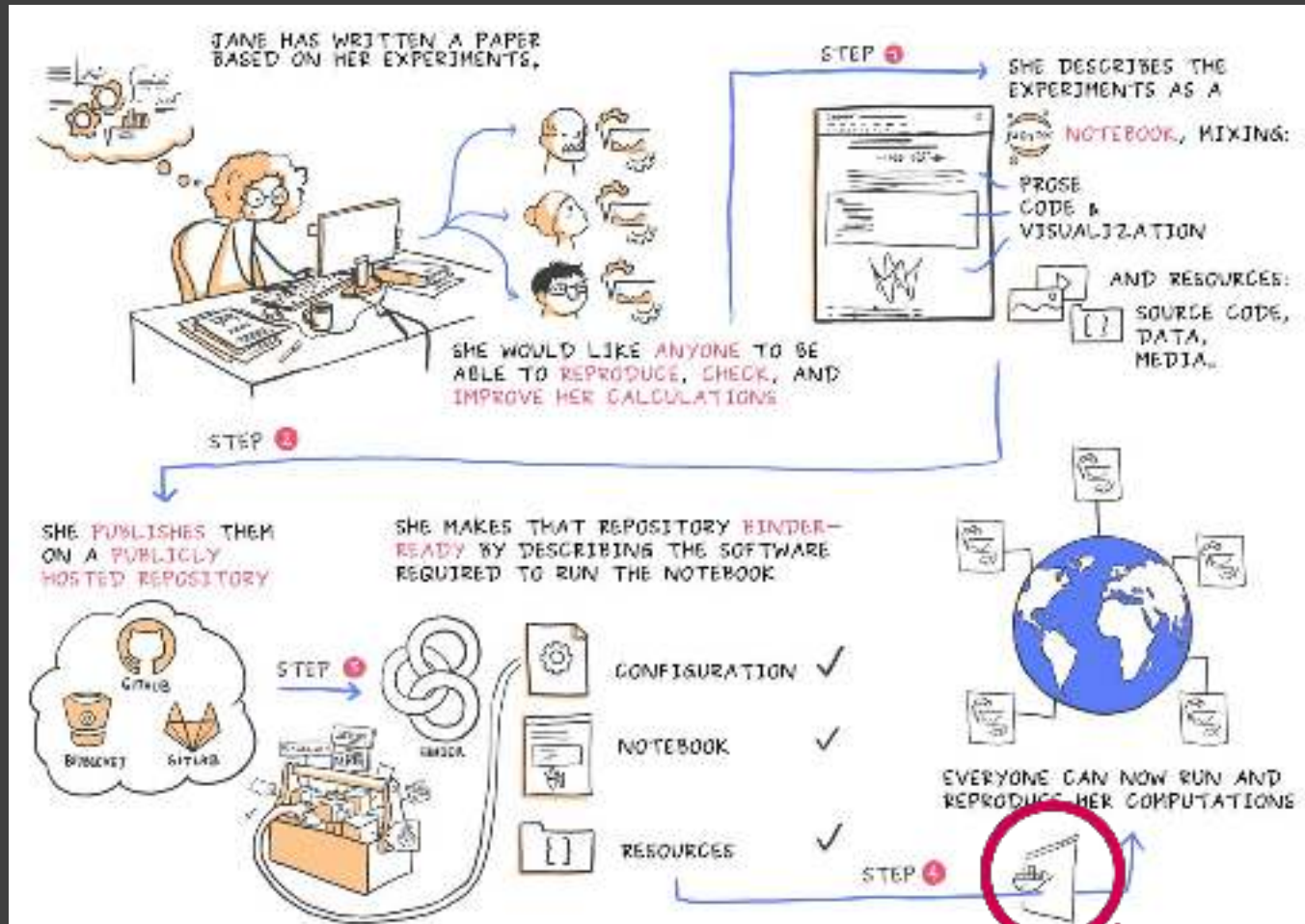
Blame

History



```
1 install.packages("tidyverse")
2 install.packages("rmarkdown")
3 install.packages("httr")
4 install.packages("shinydashboard")
5 install.packages("leaflet")
```





mybinder.org

Courtesy of Juliette Taka: <https://twitter.com/mybinderteam/status/1082556317842264064>

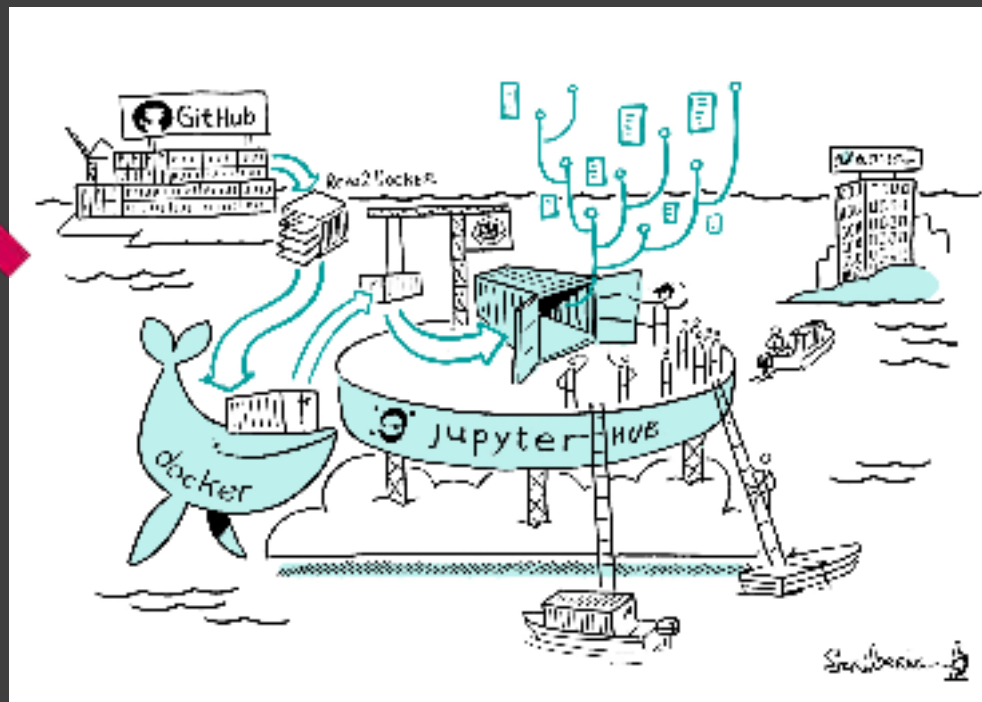
<https://doi.org/10.5281/zenodo.3628296>

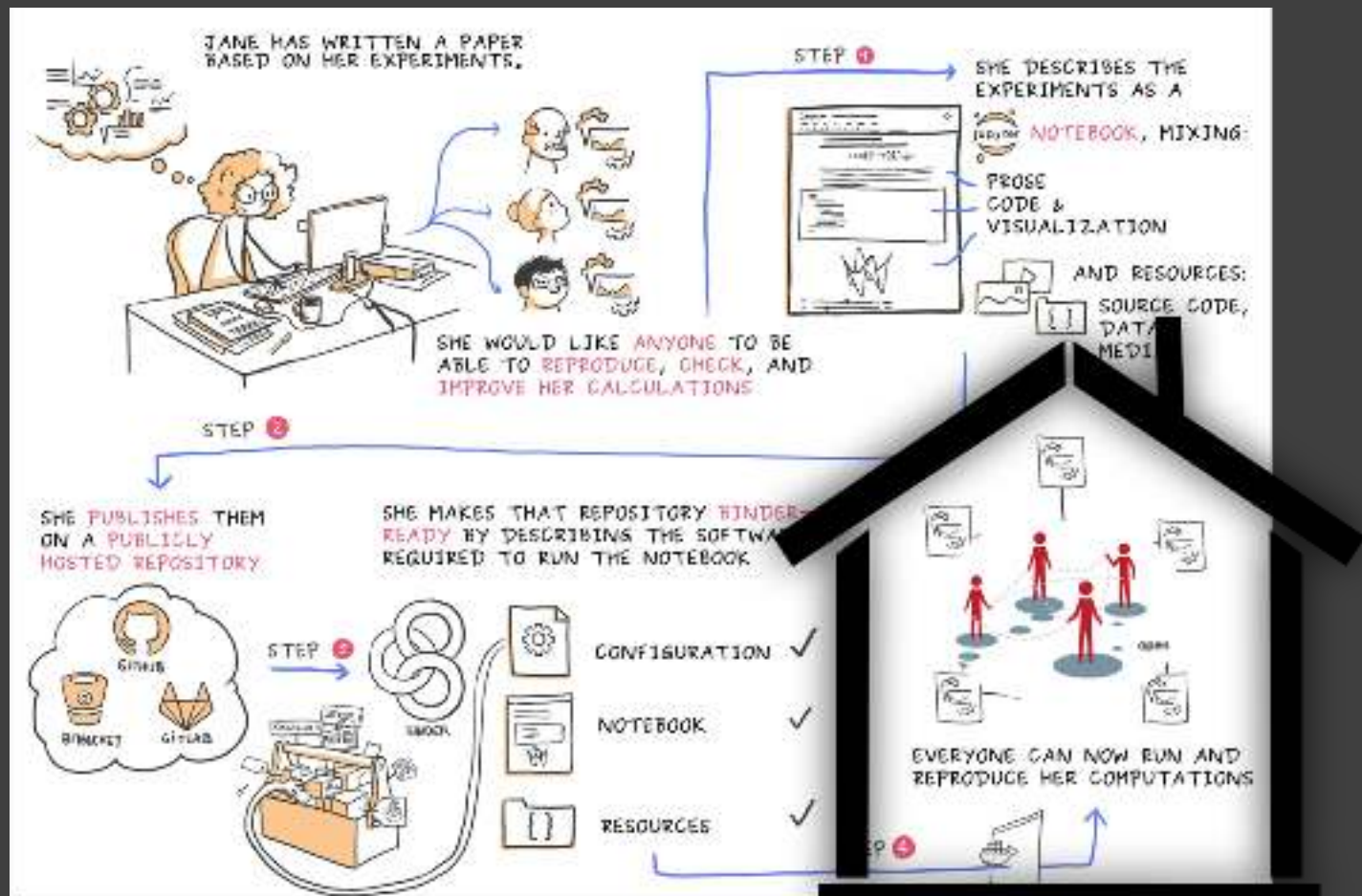
#TuringWay

BinderHub



- Collection of tools working in harmony





The Alan Turing Institute

 binder



Loading your Binder.

Take a look at our [gallery of example repositories](#).

data.egg

0/100%

Here's a app-into-pocket browser on a physical device while we start a server for you. Your binder will open automatically when it is ready.

Kirstie Whitaker

- Check analysis on her phone
- Share the responsibility with busy PIs
- Requires version control, capturing environment and new build for each change



~~Magic!~~ Technology



BinderHub

Build and launch a repository

GitHub repository name or URL

GitHub ▾

Git branch, tag, or commit

ⓘ

Path to a notebook file (optional)

File ▾

Clone GitHub Repo

1



BinderHub

Build and launch a repository

GitHub repository name or URL

GitHub ▾

Git branch, tag, or commit

ⓘ

Path to a notebook file (optional)

File ▾

1 Clone GitHub Repo



2

Build image
according to
instructions
contained within the
repo

BinderHub

Build and launch a repository

GitHub repository name or URL

GitHub ▾

Git branch, tag, or commit

ⓘ

Path to a notebook file (optional)

File ▾



1 Clone GitHub Repo

2

Build image
according to
instructions
contained within the
repo

3

Execute image

BinderHub

Build and launch a repository

GitHub repository name or URL

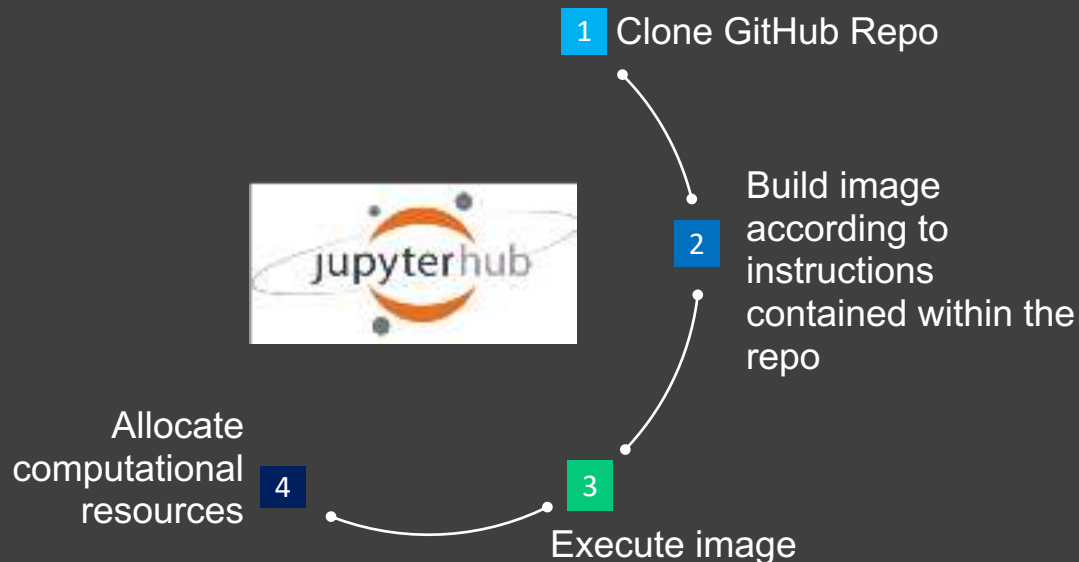
GitHub ▾

Git branch, tag, or commit

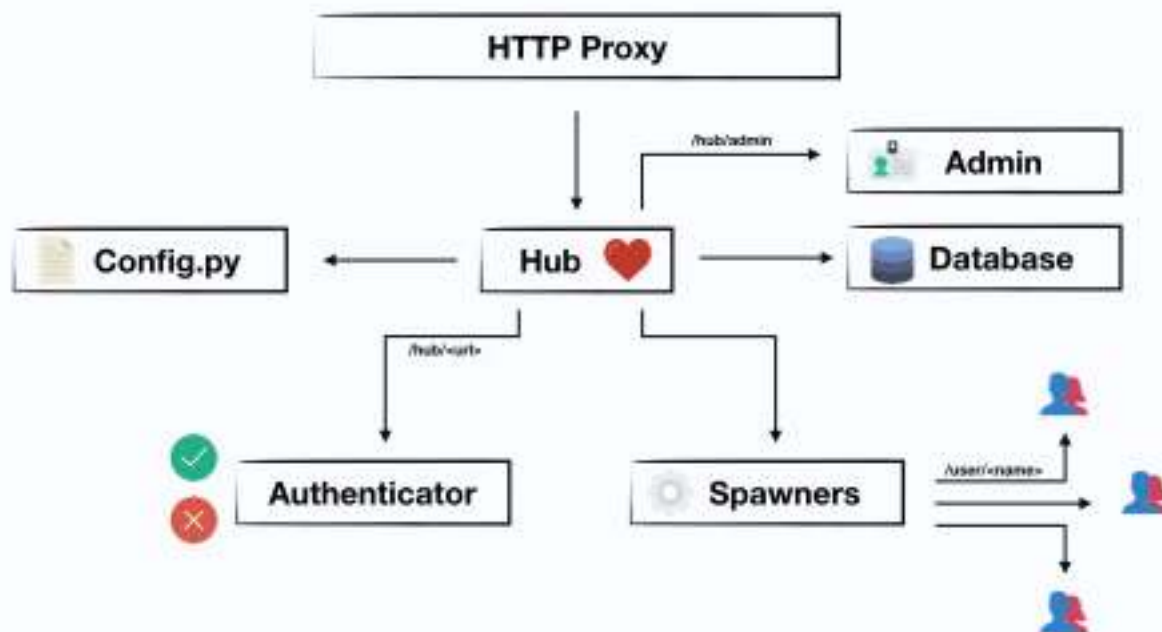
ⓘ

Path to a notebook file (optional)

File ▾



What is a JupyterHub?



All icons were obtained from Flaticon (<https://www.flaticon.com/packs/essential-collection>)

resources

Execute image

JupyterHub is a way to help your humans use your computers. With notebooks!

BinderHub

Build and launch a repository

GitHub repository name or URL

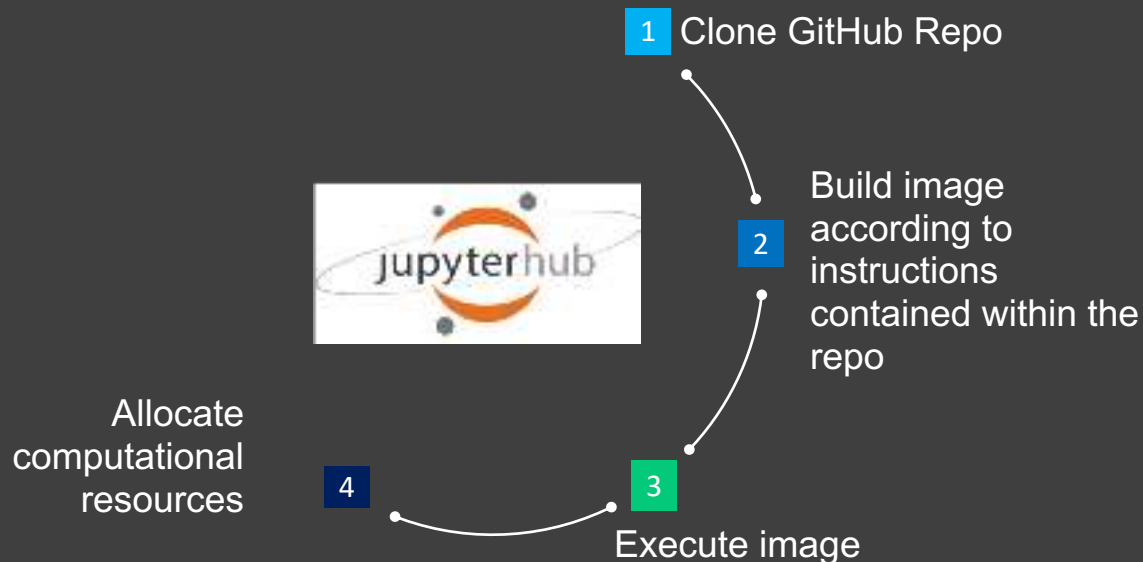
GitHub ▾

Git branch, tag, or commit

ⓘ

Path to a notebook file (optional)

File ▾



BinderHub

Build and launch a repository

GitHub repository name or URL

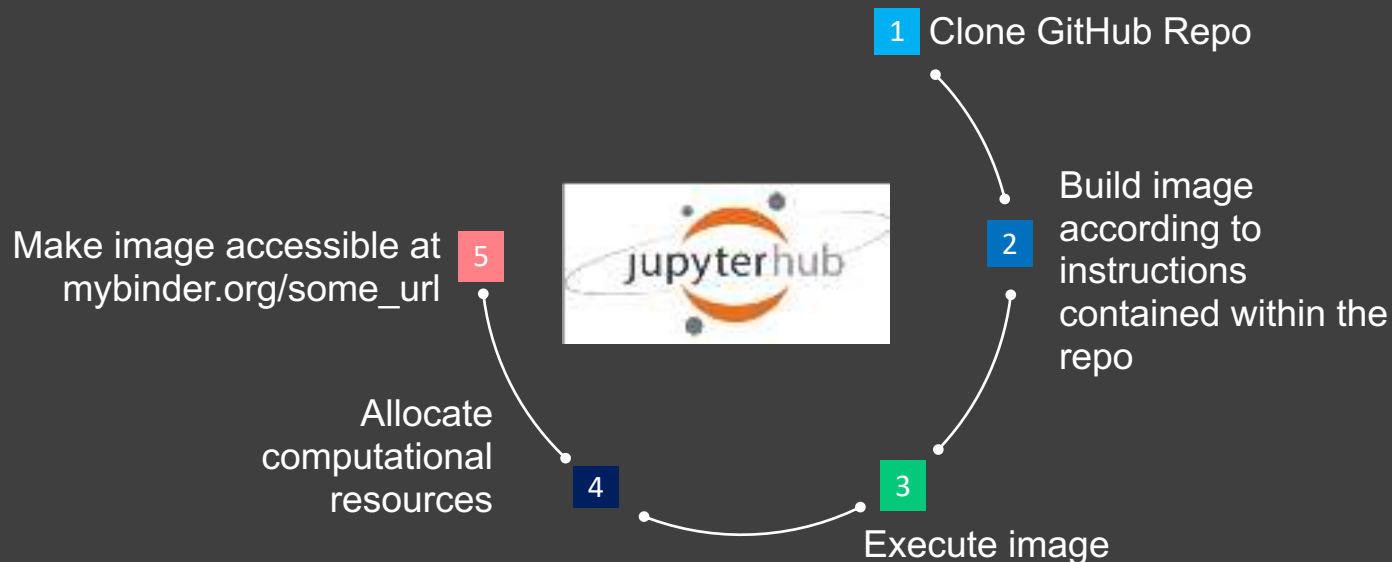
GitHub ▾

Git branch, tag, or commit

ⓘ

Path to a notebook file (optional)

File ▾



BinderHub

Build and launch a repository

GitHub repository name or URL

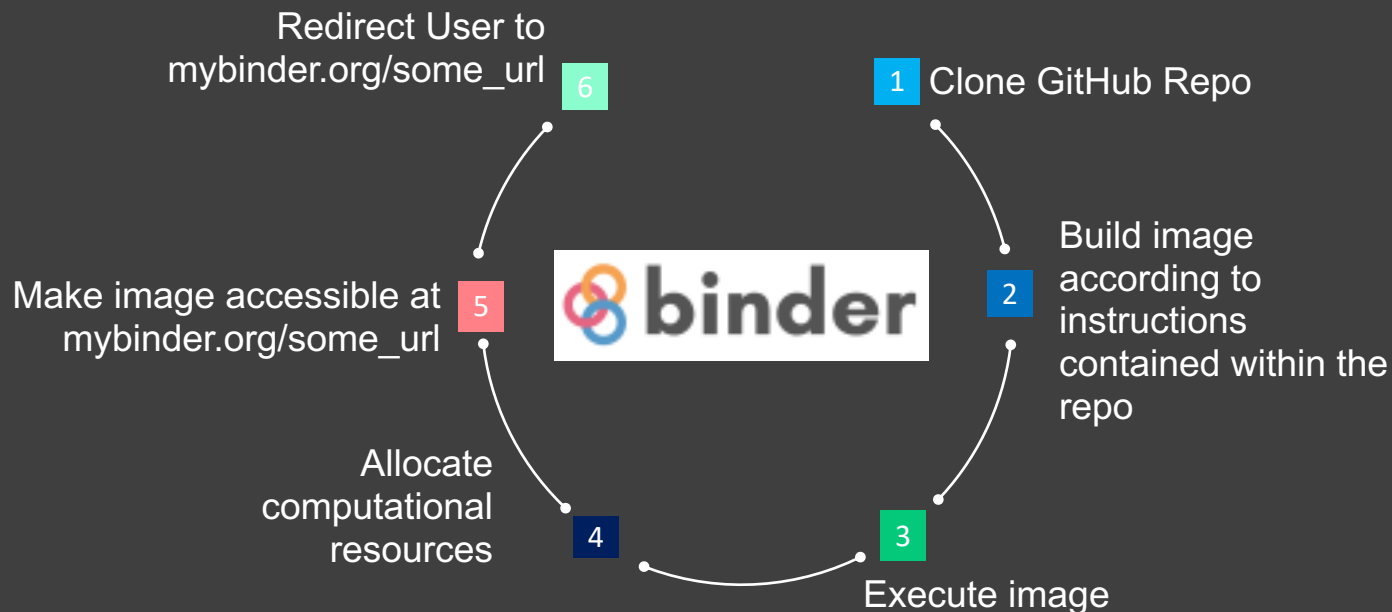
GitHub ▾

Git branch, tag, or commit

ⓘ

Path to a notebook file (optional)

File ▾





Thank you to our current (and future!) contributors



This project follows the all-contributors specification. Contributions of any kind welcome!

<https://github.com/alan-turing-institute/the-turing-way#contributors>
<https://allcontributors.org/docs/en/emoji-key>
<https://doi.org/10.5281/zenodo.3628296>

#TuringWay

Thank you

The
Alan Turing
Institute



Tutorial: <https://bit.ly/zero-to-binder-tutorial>

- Book: <https://the-turing-way.netlify.com>
- Newsletter: <https://tinyletter.com/TuringWay>
- GitHub: <https://github.com/alan-turing-institute/the-turing-way>
- Chat: <https://gitter.im/alan-turing-institute/the-turing-way>
- This work was supported by The UKRI Strategic Priorities Fund under the EPSRC Grant EP/T001569/1, particularly the "Tools, Practices and Systems" theme within that grant, and by The Alan Turing Institute under the EPSRC grant EP/N510129/1
- Original artwork by Scriberia: <https://doi.org/10.5281/zenodo.3332807>