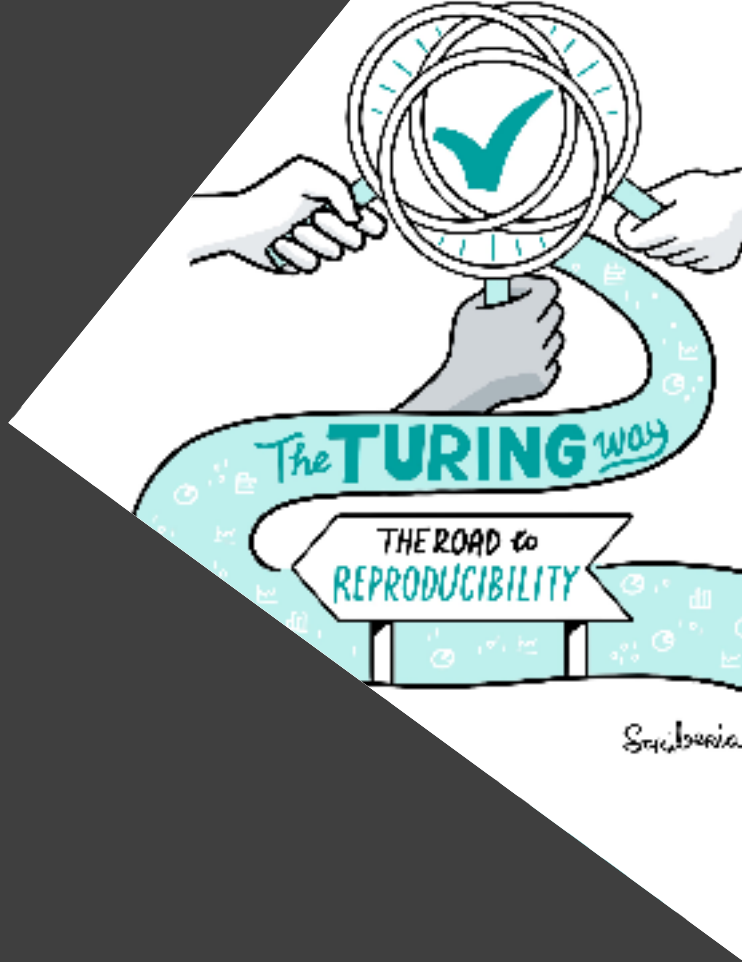
The Alan Turing Institute

binder



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

Dr Sarah Gibson

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

#TuringWay

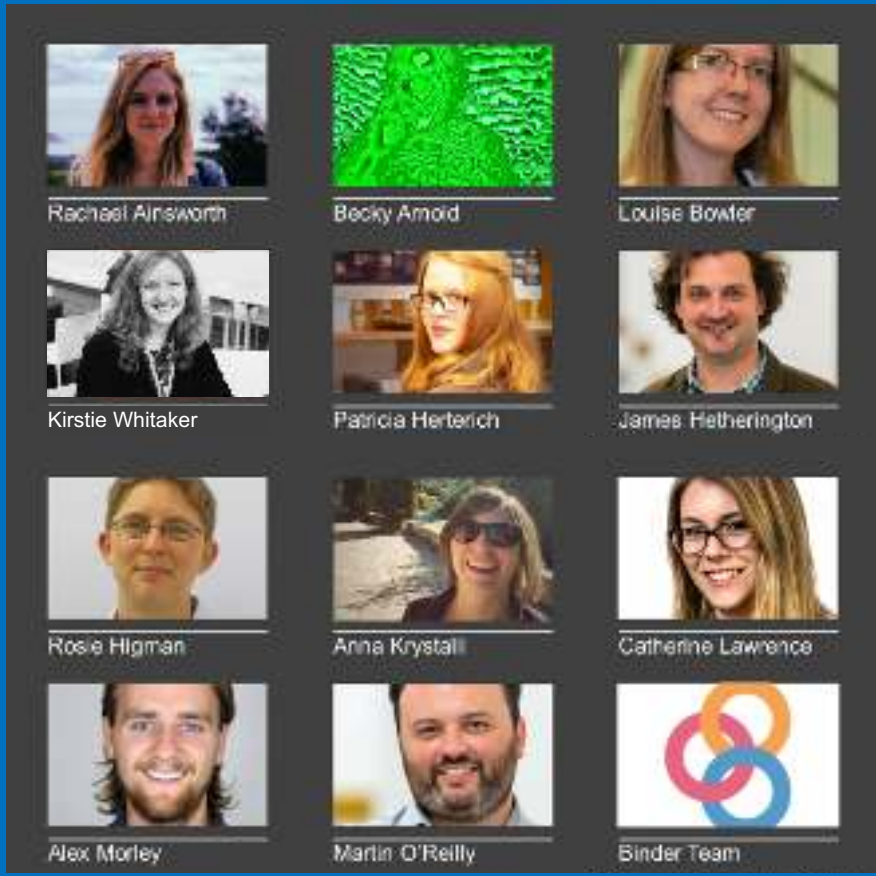
Agenda

– Talk 🗣️ (*45 mins*)
– Pizza! 🍕 (*15 mins*)
– Tutorial 👩‍💻 (*45 mins*)

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

The Turing Way is:

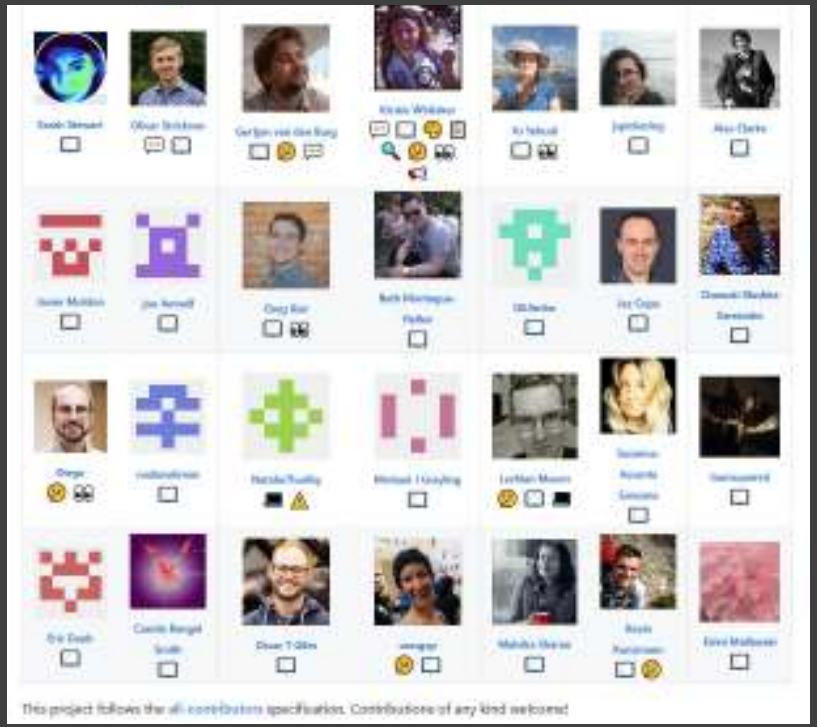
- a book
- a community
- a global collaboration
- a whole tonne of work



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

#TuringWay

Thank you to all our contributors



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

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

Academic errors have real world effects

	B	C	I	J	K	L	M
2			Real GDP growth				
3			Debt/GDP				
4	Country	Coverage	30 or less	30 to 60	60 to 90	90 or above	30 or less
26			5.1	3.0	1.5	1.7	5.5
27	Minimum		1.5	0.3	1.5	-1.8	0.8
28	Maximum		5.4	4.0	10.2	3.6	18.3
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	3.6	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	7.9	4.0	1.0	0.7	7.0
39	Italy	1951-2009	5.4	2.1	1.8	1.0	5.6
40	Ireland	1948-2009	4.4	4.5	6.0	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	n.a.	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	3.1	1.7	2.4	n.a.	5.6
46	Canada	1951-2009	1.9	3.6	4.1	n.a.	7.2
47	Belgium	1947-2009	n.a.	4.2	3.1	2.6	n.a.
48	Austria	1948-2009	5.2	3.3	-3.8	n.a.	5.7
49	Australia	1951-2009	3.2	4.9	4.0	n.a.	5.9
50							
51			4.1	2.8	2.8	AVERAGE (1,3,6,1,44)	

https://statmodeling.stat.columbia.edu/2013/04/16/memo-to-reinhart-and-rogoff-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

Academic errors have real world effects

	B	C	I	J	K	L	M
2			Real GDP growth when debt/GDP				
3							
4	Country	Coverage	30 or less	30 to 60	60 to 90	90 or above	30 or less
26			3.1	3.0	2.8	1.7	3.5
27	Minimum		1.5	0.3	1.5	-1.8	0.8
28	Maximum		5.4	4.9	10.2	3.6	10.3
29							
30	US	1946-2009	n.a.	3.4	3.3	-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.	2.9
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	7.9	4.0	1.0	0.7	7.0
39	Italy	1951-2009	5.4	2.1	1.8	1.0	5.6
40	Ireland	1948-2009	4.4	4.5	4.0	2.8	2.9
41	Greece	1970-2009	4.0	0.3	2.7	2.9	13.3
42	Germany	1946-2009	3.9	0.9	n.a.	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	3.1	1.7	2.4	n.a.	5.6
46	Canada	1951-2009	1.9	3.6	4.1	n.a.	2.2
47	Belgium	1947-2009	n.a.	4.2	3.1	2.6	n.a.
48	Austria	1948-2009	5.2	3.3	-3.8	n.a.	5.7
49	Australia	1951-2009	3.2	4.9	4.0	n.a.	3.9
50							
51			4.1	2.8	2.8	AVERAGE (1,3,5,7,..49)	

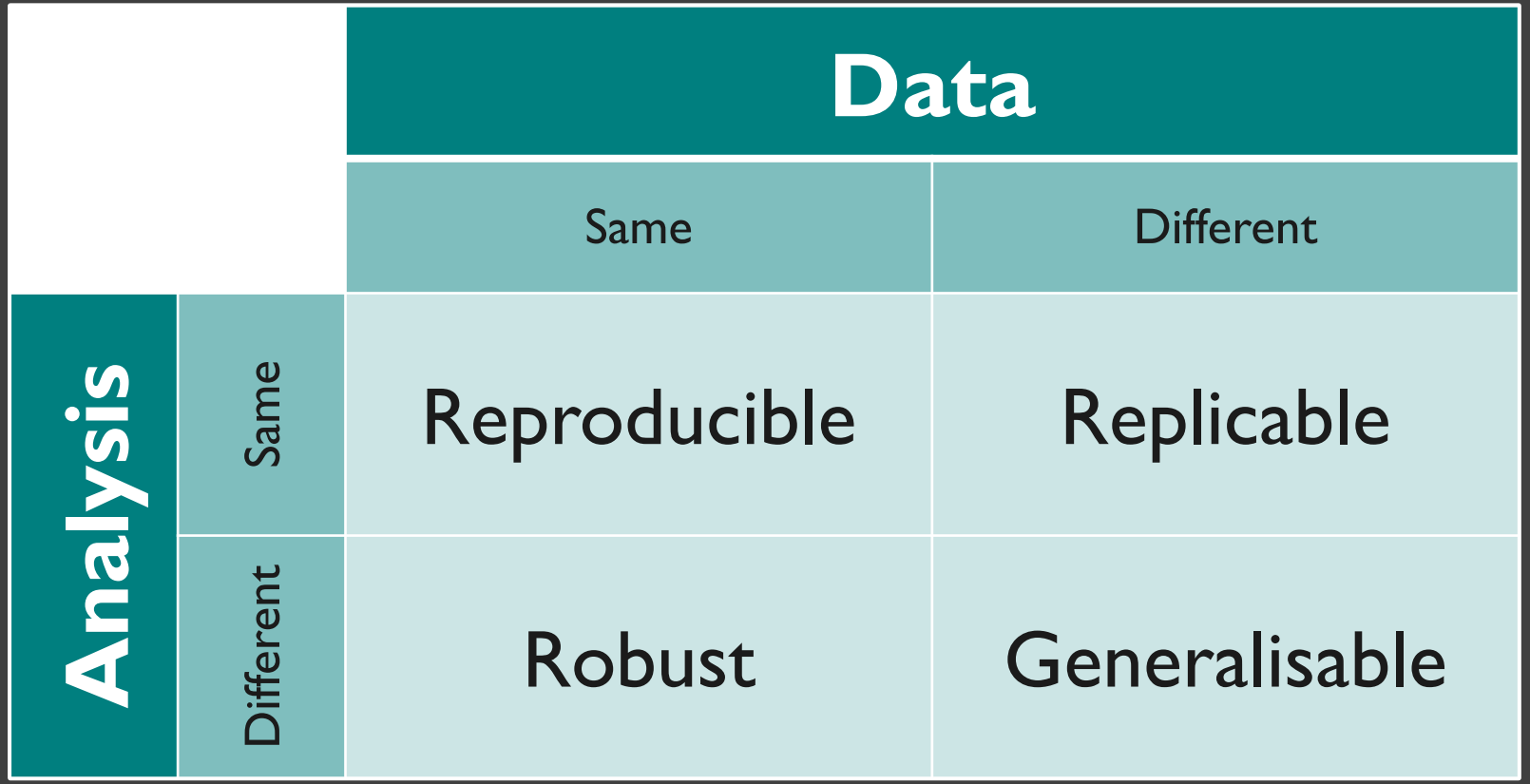
https://statmodeling.stat.columbia.edu/2013/04/16/memo-to-reinhart-and-rogoff-i-think-its-best-to-admit-your-errors-and-go-on-from-there

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

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

The humans are the hardest part of reproducibility

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

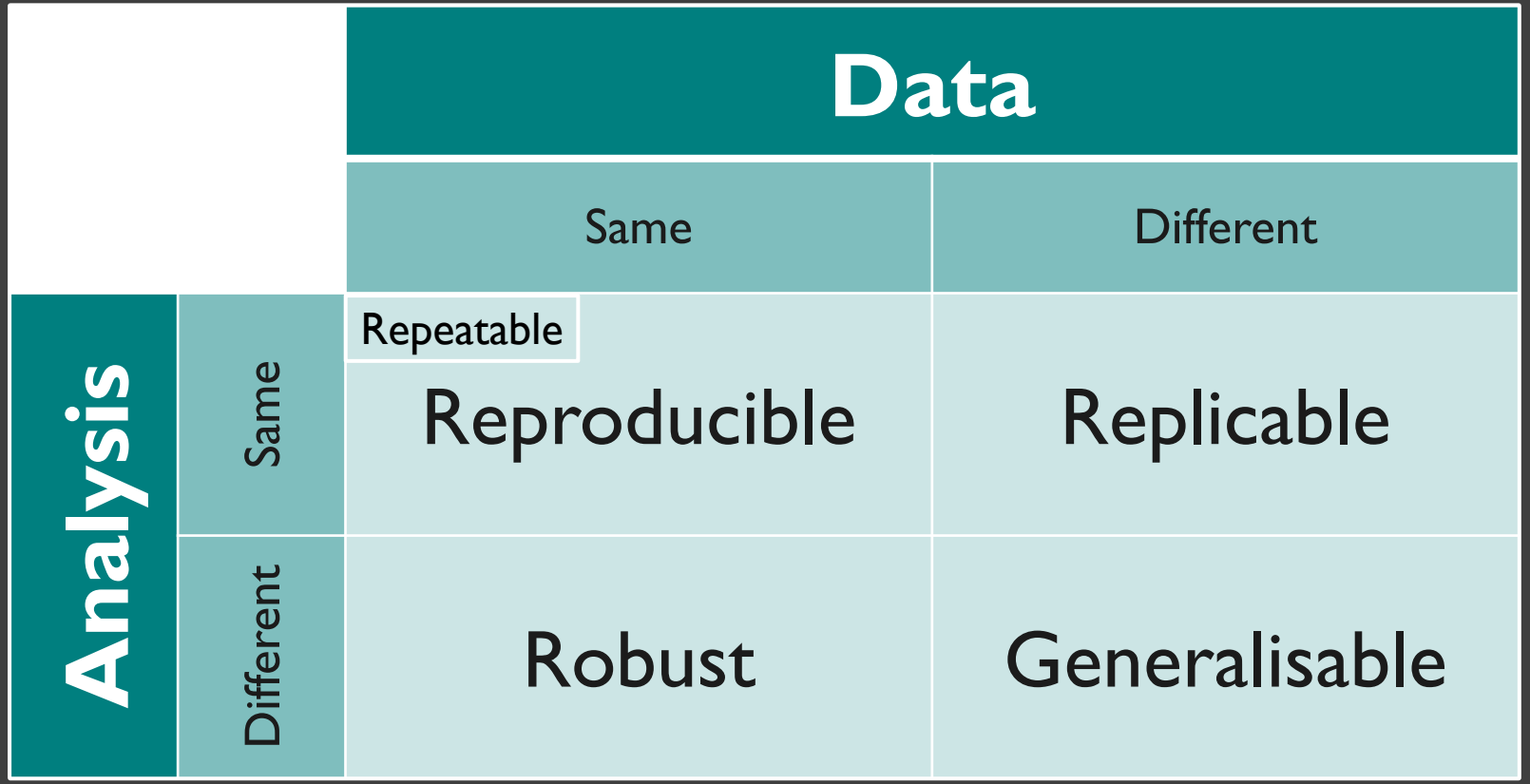


Kirstie Whitaker's talk at PyData LDN: https://youtu.be/IG3PcZ6EhiU

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

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

#TuringWay



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

Barriers to reproducible research

Is not considered for promotion

Held to higher standards than others

Publication bias towards novel findings

Requires additional skills

Plead the 5th

Support additional users

Takes time

Kirstie Whitaker's talk at PyData LDN: https://youtu.be/lG3PcZ6EhiU

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

#TuringWay

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

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

#TuringWay

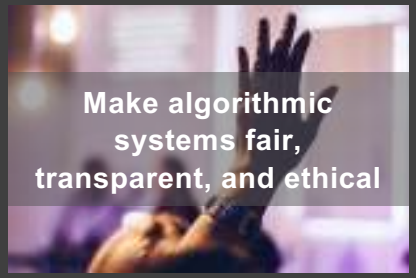
The Institute's partners and collaborators

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

#TuringWay

Challenges

Advance data science and artificial intelligence to…

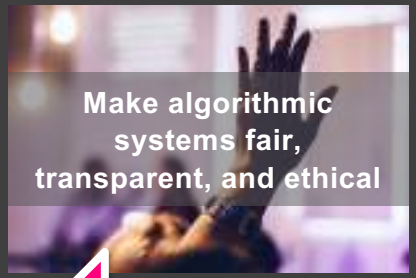


https://doi.org/10.5281/zenodo.3628296

#TuringWay

Challenges

Advance data science and artificial intelligence to…



The Turing Way

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

#TuringWay

The Turing Way

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

#TuringWay

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

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

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

Barriers to reproducible research

Is not considered for promotion

Held to higher standards than others

Publication bias towards novel findings

Requires additional skills

Plead the 5th

Support additional users

Takes time

Kirstie Whitaker's talk at PyData LDN: https://youtu.be/lG3PcZ6EhiU

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

#TuringWay

A global collaboration

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

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.”

https://rd-alliance.org/users/patricia-herterich

#TuringWay

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

moz://a

Open Leadership Principles

Understanding
You make the work accessible and clear

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

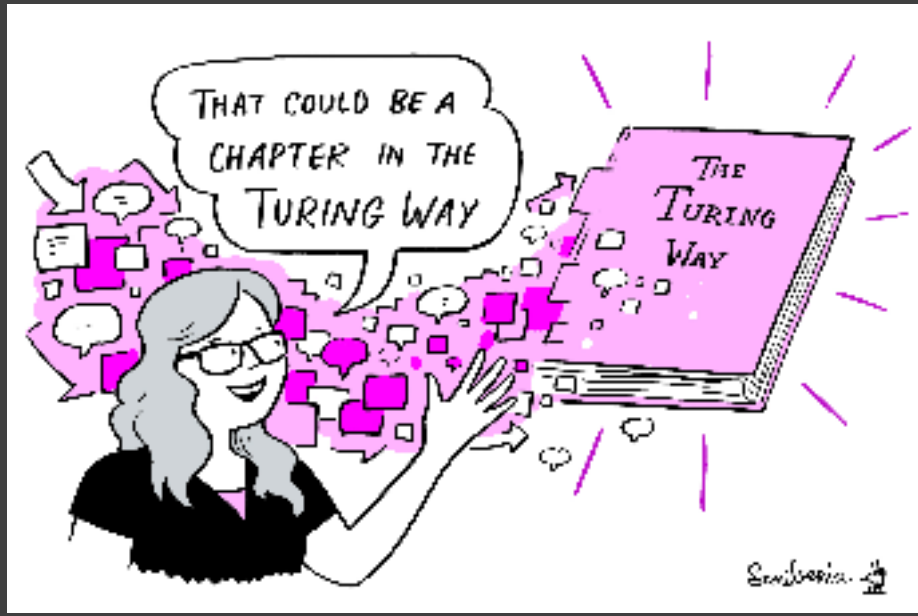
Participation & Inclusion
You build shared ownership and agency to make the work inviting and sustainable for all.

Read more
https://mozilla.github.io/olm-whitepaper

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

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://doi.org/10.5281/zenodo.3628296

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

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

Market Research

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

Have you ever heard…?

“Oh, it worked on my computer?”

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

Have you ever heard…?

“Oh, it worked yesterday?”

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

“Oh, it worked on my computer?”

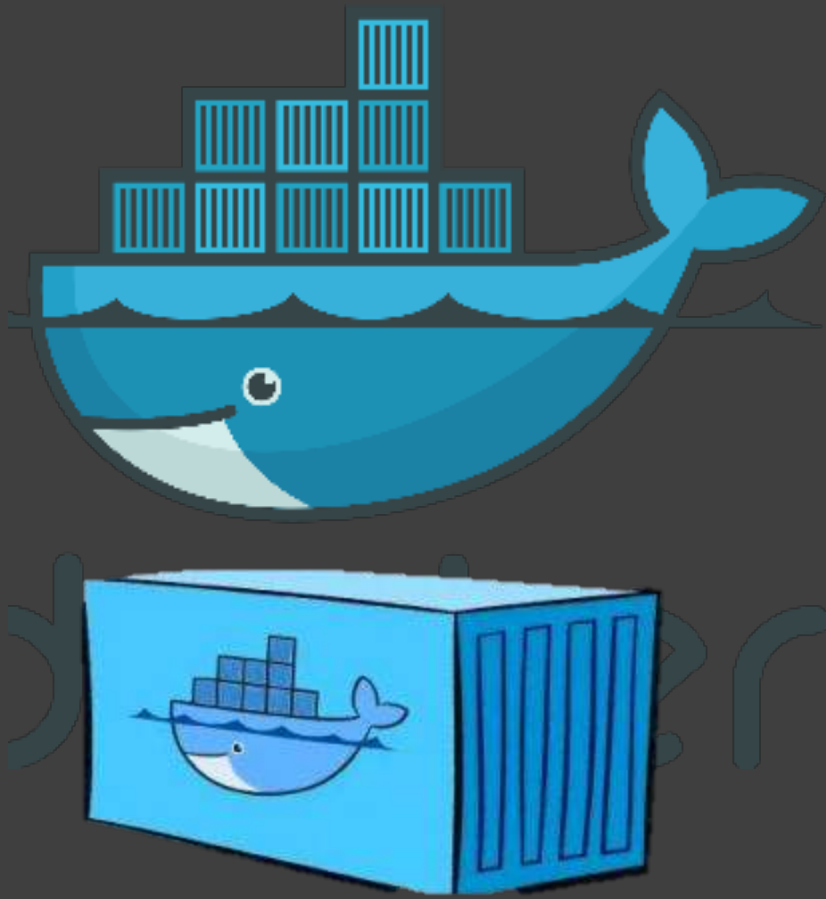
+ CI

“Oh, it worked yesterday?”

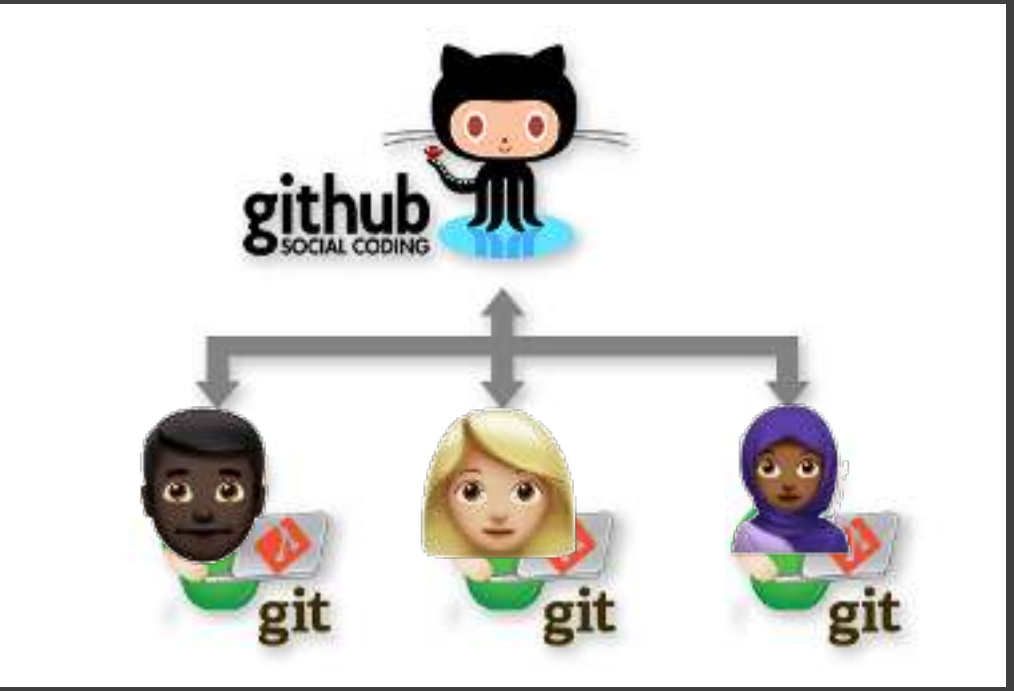
https://doi.org/10.5281/zenodo.3628296

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!



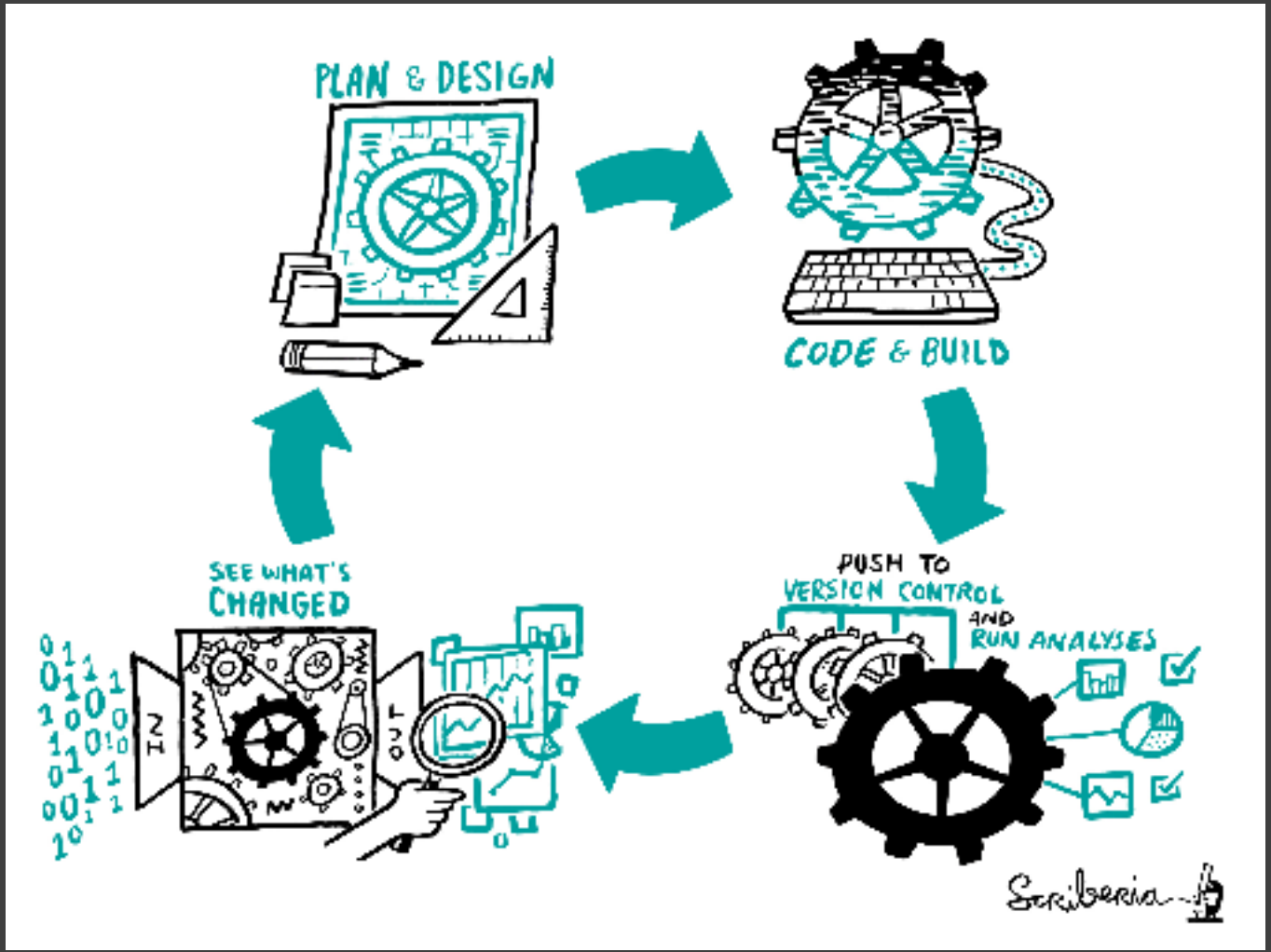
https://doi.org/10.5281/zenodo.3628296



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



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

#TuringWay

Turing Way & Binder

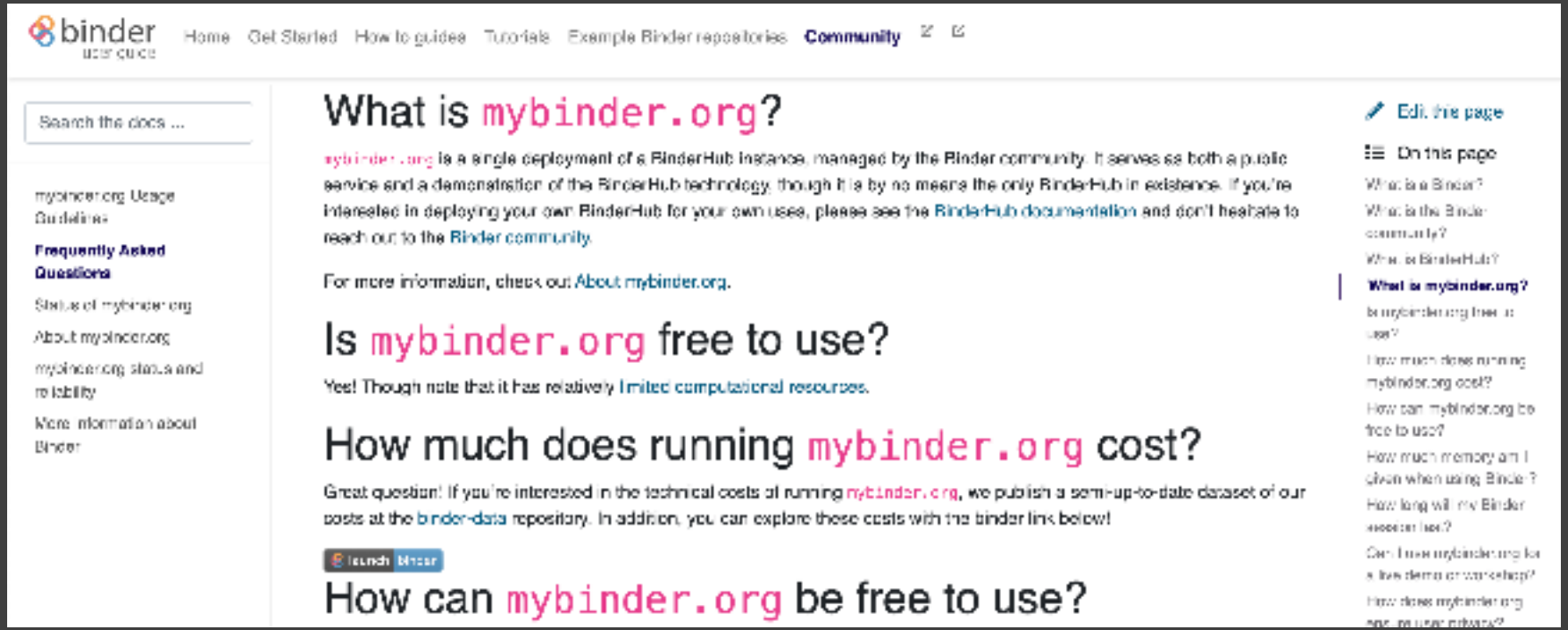
https://doi.org/10.5281/zenodo.3628296

#TuringWay

The Vocab

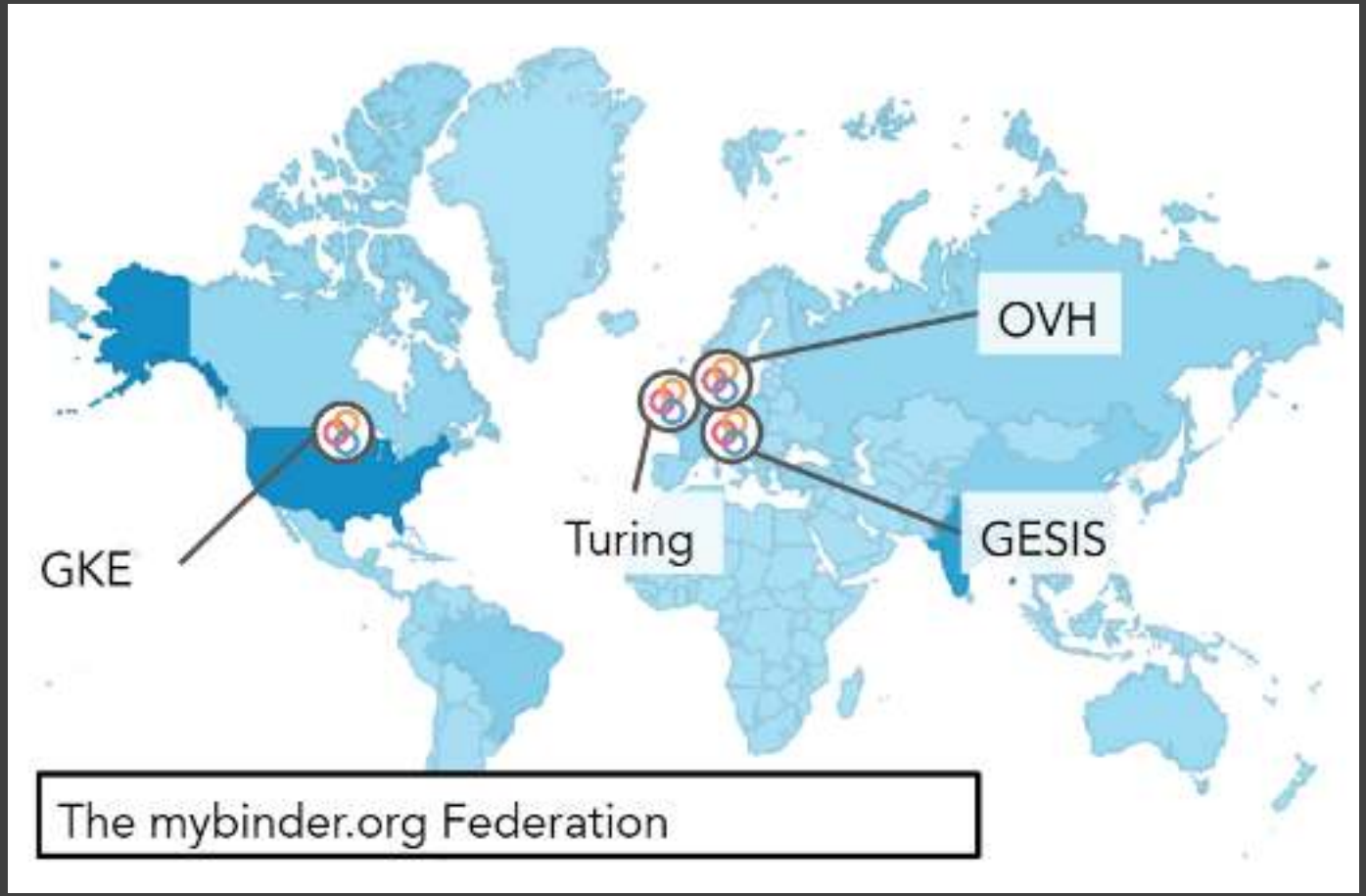
– Binder → user interface/experience
– BinderHub → computational infrastructure
– mybinder.org → public BinderHub for everyone

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



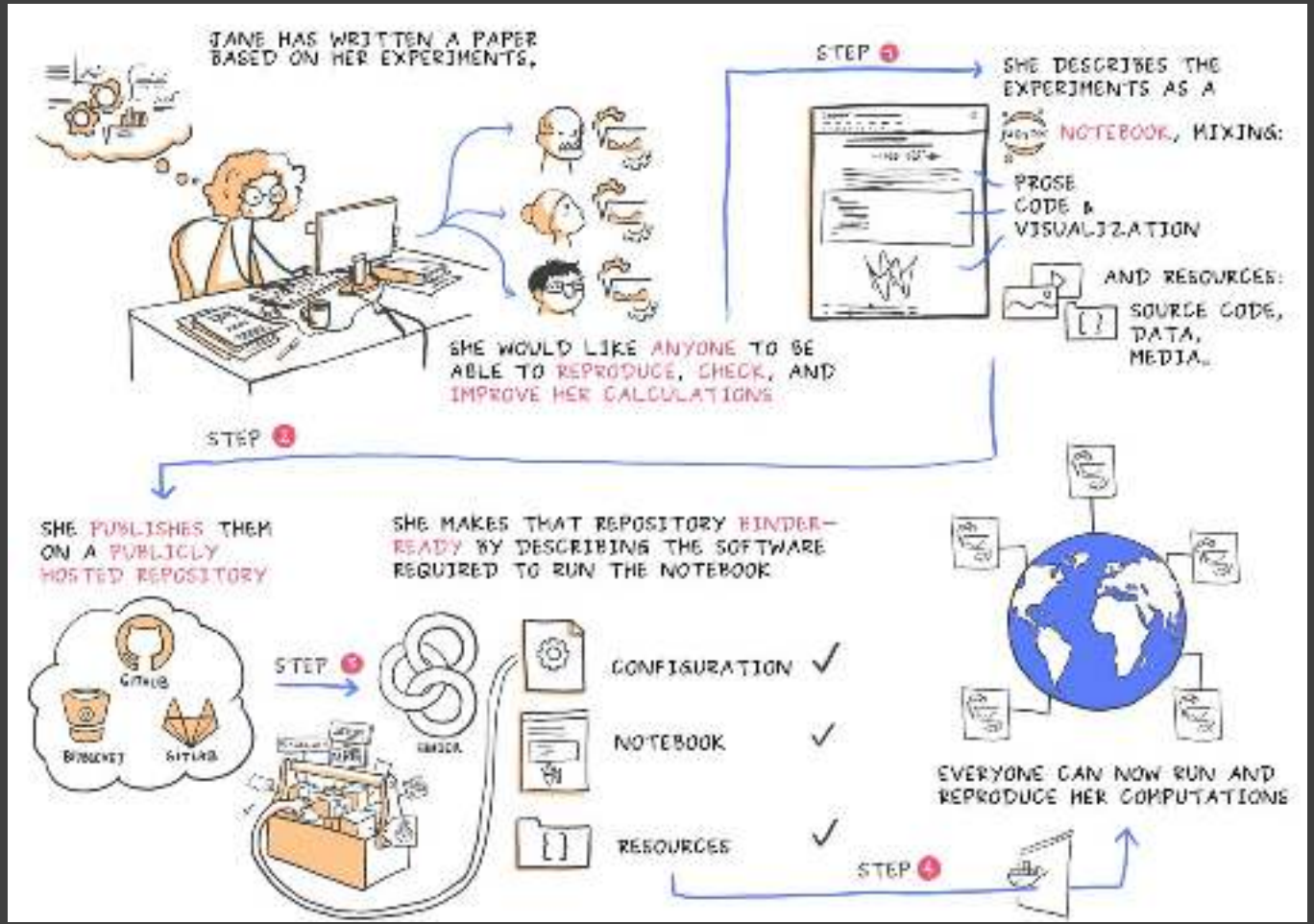
https://mybinder.readthedocs.io/en/latest/faq.html#what-is-mybinder-org

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



https://blog.jupyter.org/a-2019-retrospective-from-the-binder-project-57a449517362
https://doi.org/10.5281/zenodo.3628296

#TuringWay



binder

mybinder.org

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

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

#TuringWay

binder-examples / requirements

Watch 3 | Star 16 | Fork 60

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

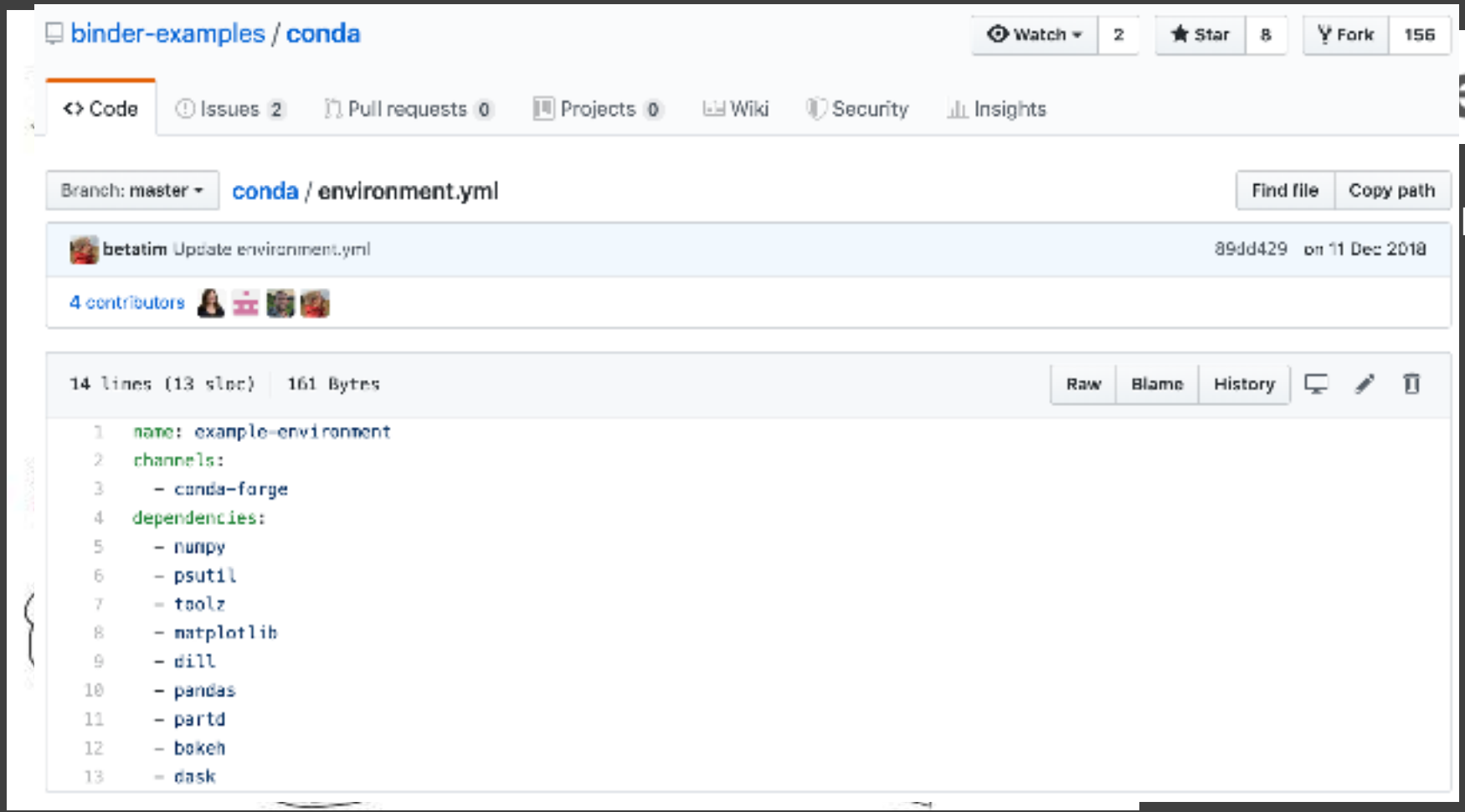
5 lines (3 sloc) 46 Bytes — Raw | Blame | History

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

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

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

#TuringWay



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

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

#TuringWay

binder-examples / binder-r-description

Watch 1 | Star 2 | Fork 1

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 <bgreshake@googlemail.com>
7  Maintainer: Bastian Greshake Tzovaras <bgreshake@googlemail.com>
```

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

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

#TuringWay

binder-examples / r

Watch 13 | Star 71 | Fork 123

Code | 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

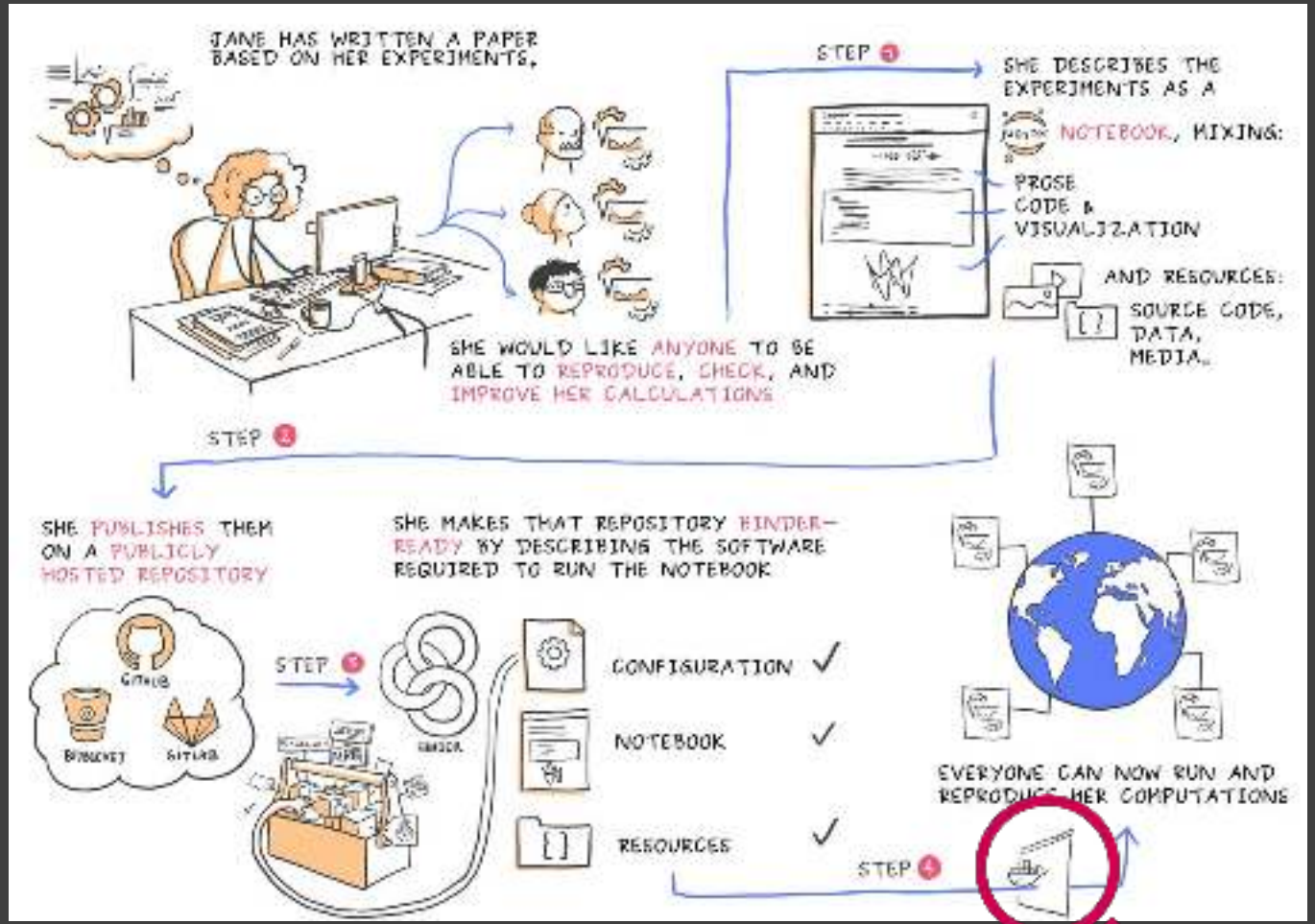
Raw | Blame | History

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

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

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

#TuringWay



binder

mybinder.org

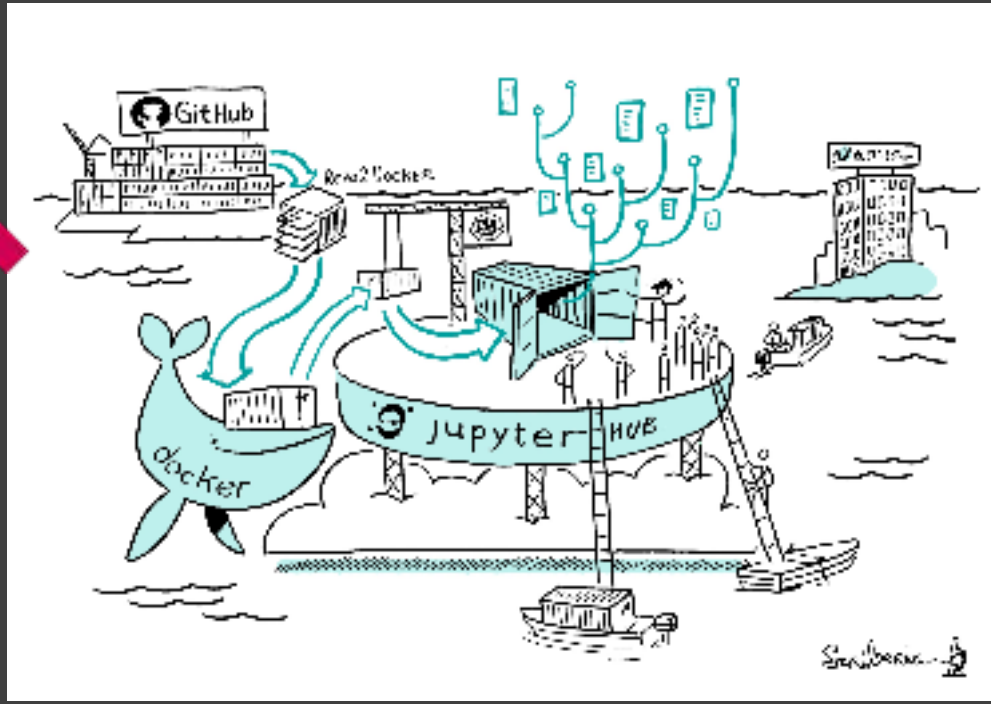
Courtesy of Juliette Taka: https://twitter.com/mybinderteam/status/1082556317842264064

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

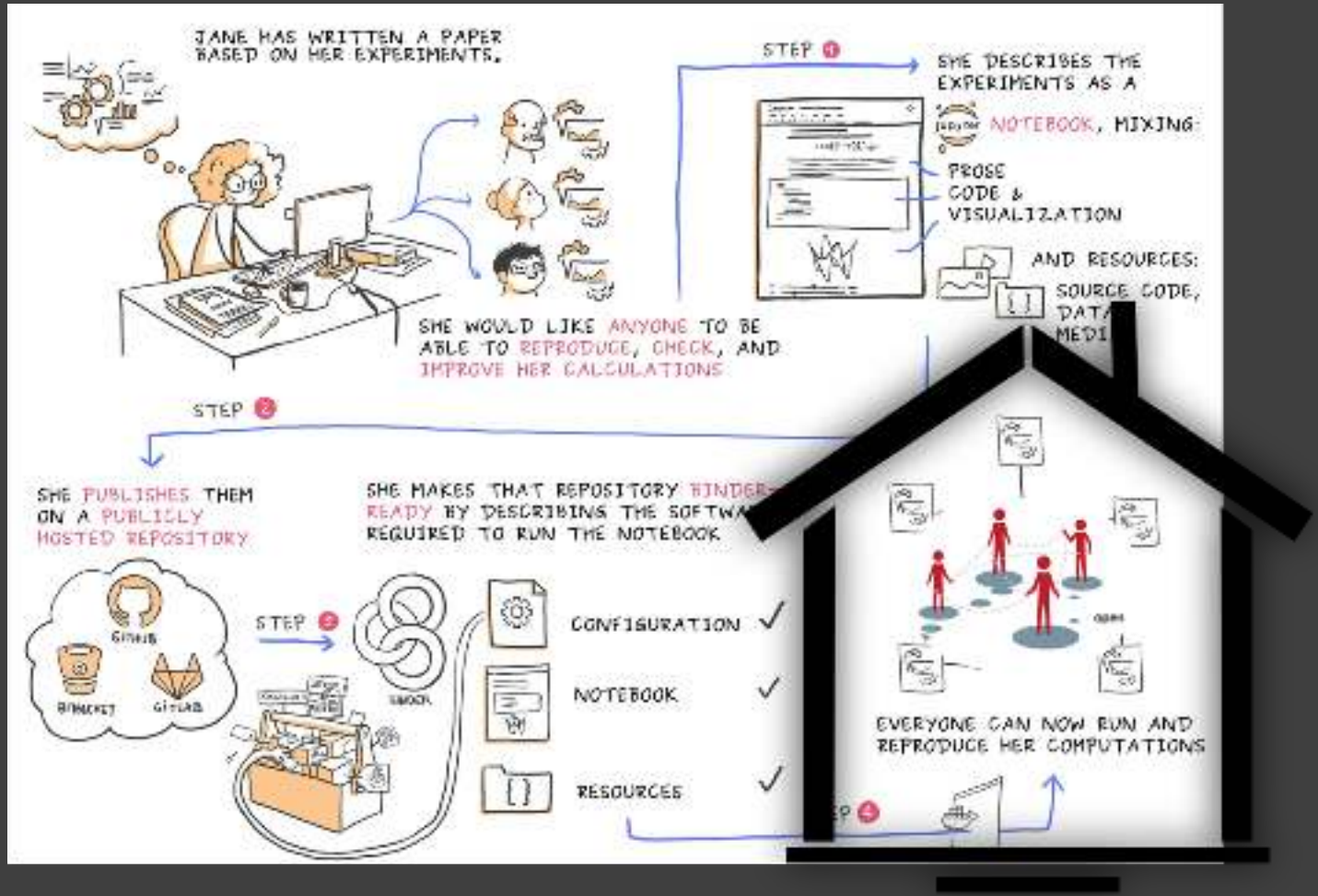
#TuringWay

BinderHub

– Collection of tools working in harmony

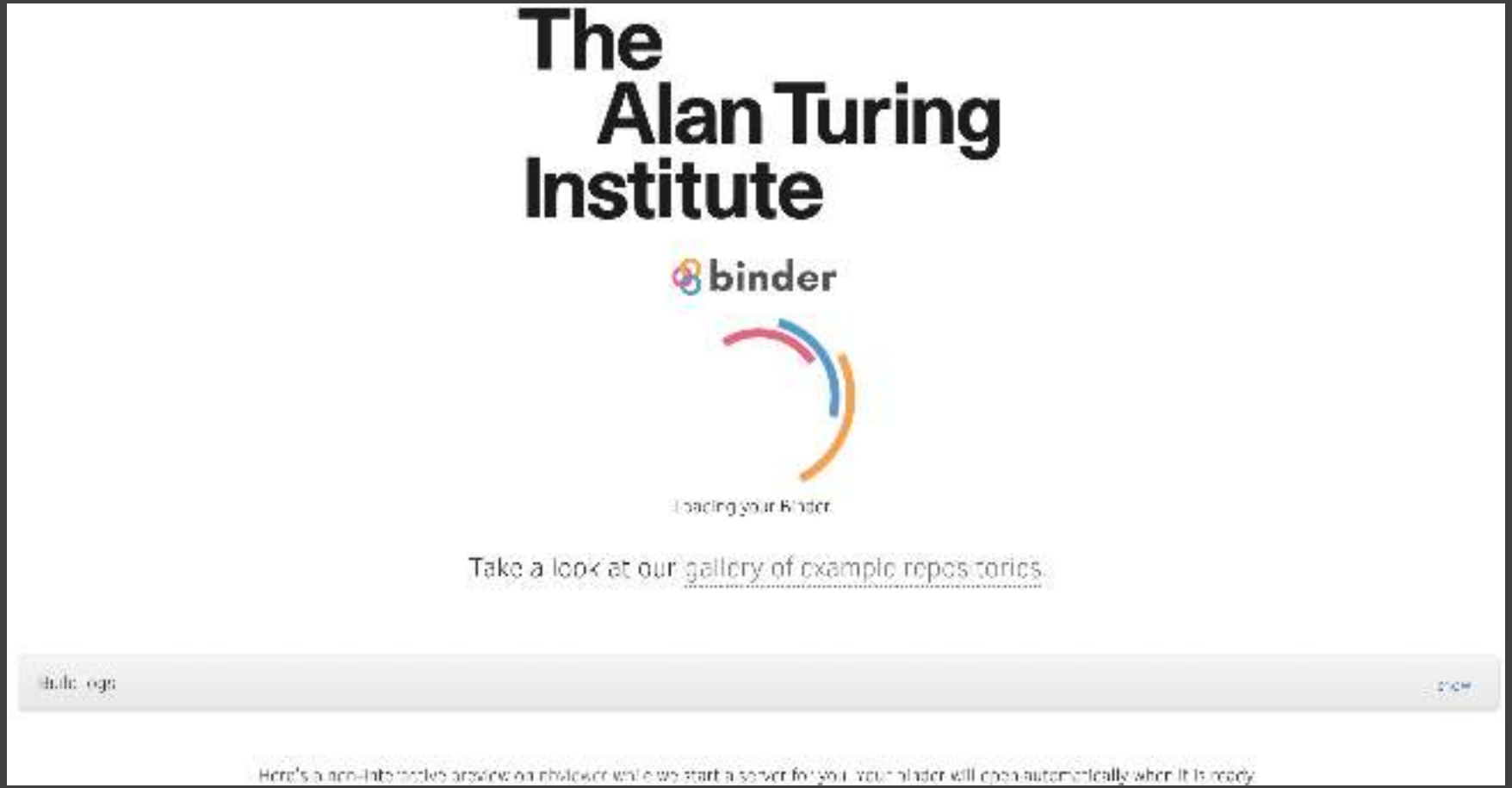


https://doi.org/10.5281/zenodo.3628296



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

#TuringWay



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

#TuringWay

Kirstie Whitaker

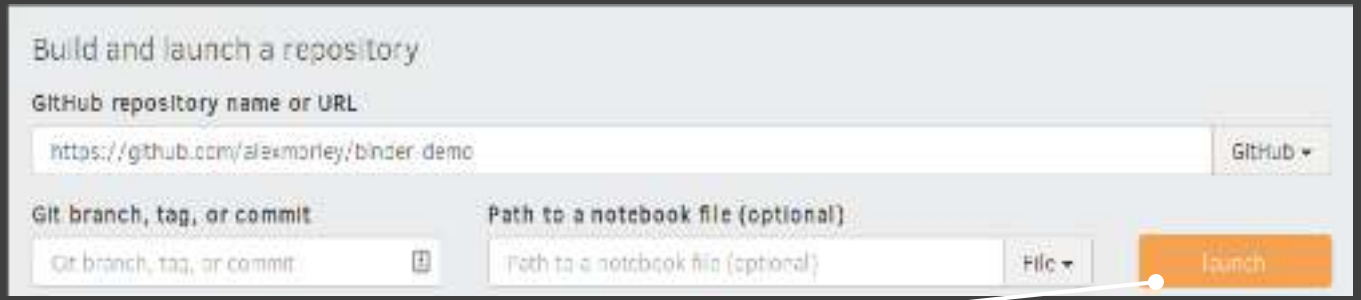
- Check analysis on her phone
- Share the responsibility with busy PIs
- Requires version control, capturing environment and new build for each change

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

~~Magic!~~ Technology

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

BinderHub

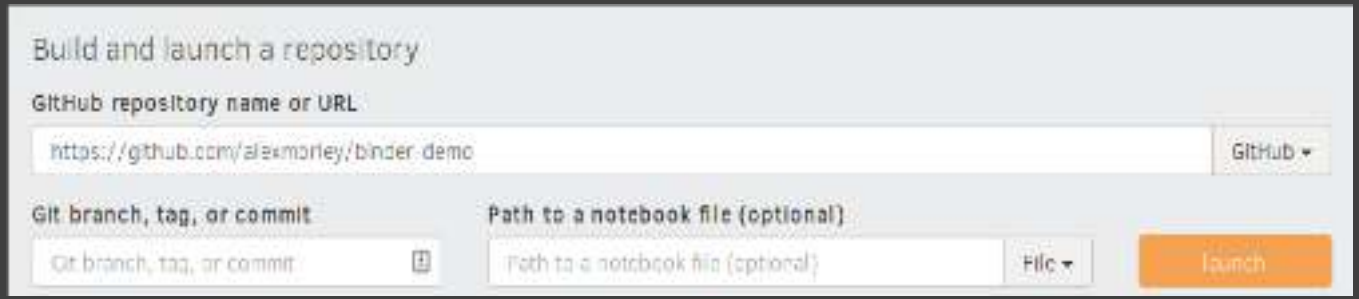


Clone GitHub Repo

1

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

BinderHub



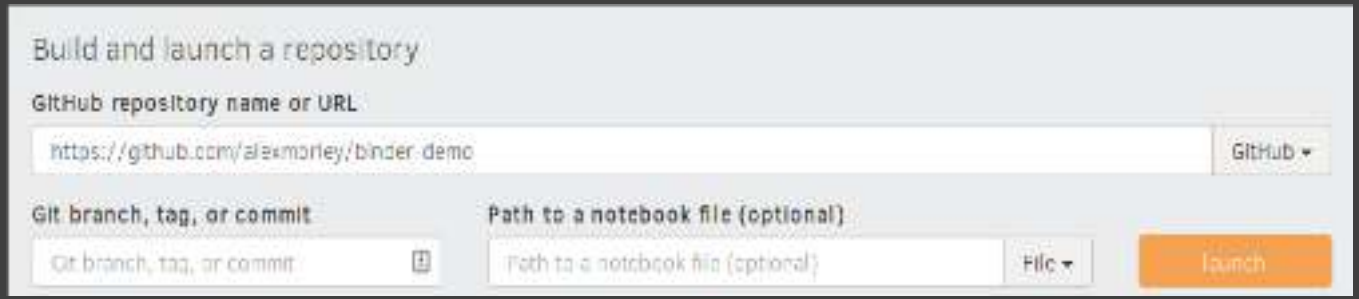
1 Clone GitHub Repo

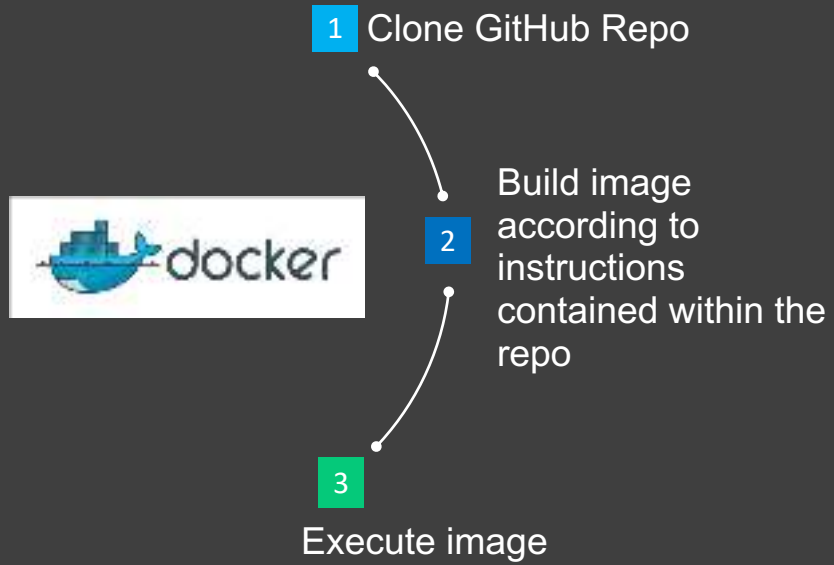
repo2docker

2 Build image according to instructions contained within the repo

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

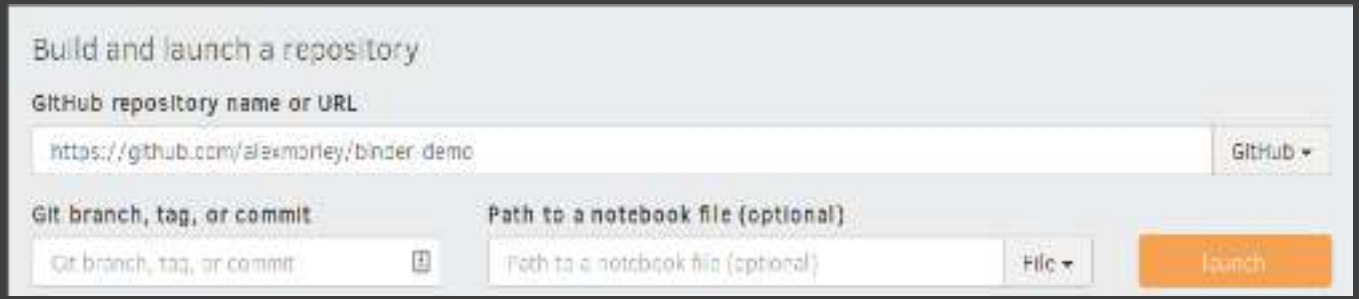
BinderHub

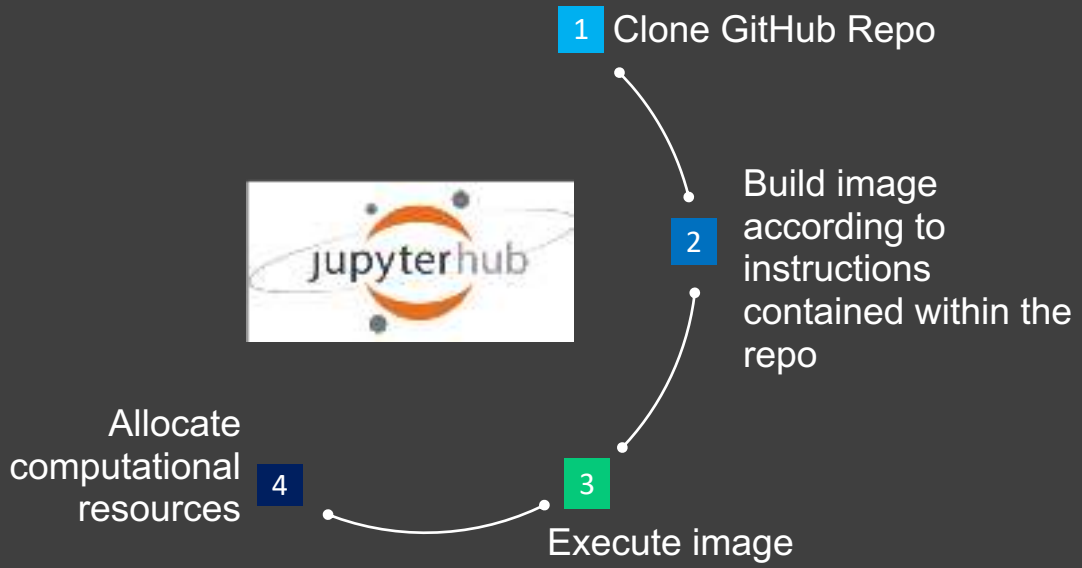




https://doi.org/10.5281/zenodo.3628296

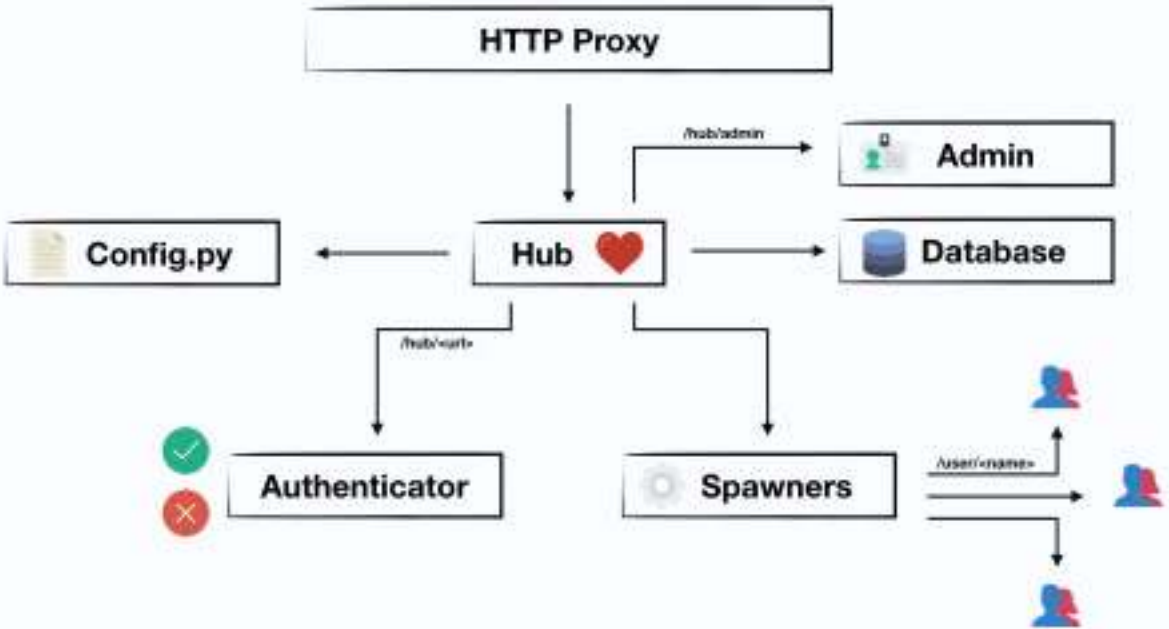
BinderHub





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

What is a JupyterHub?



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

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

hin the

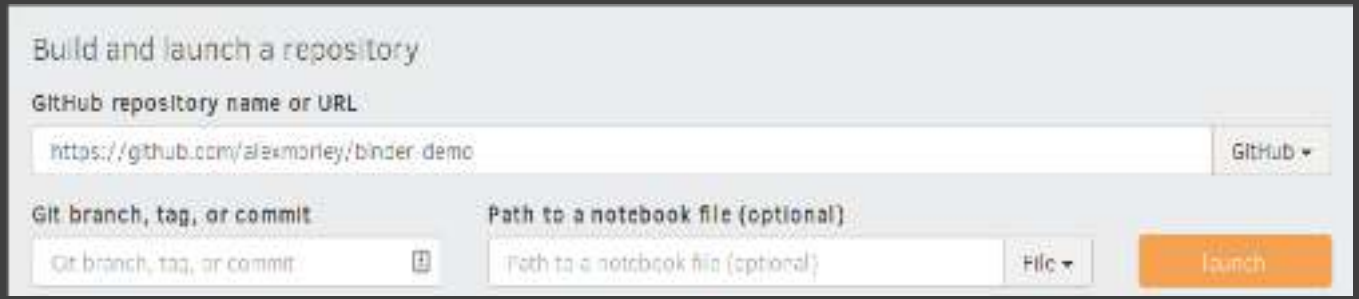
resources

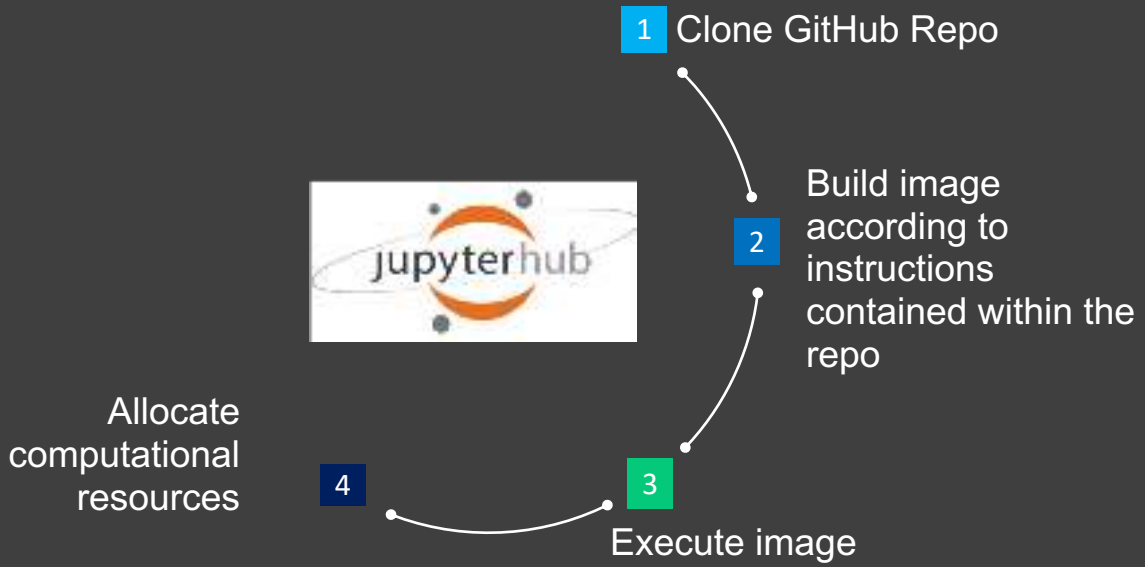
Execute image

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

#TuringWay

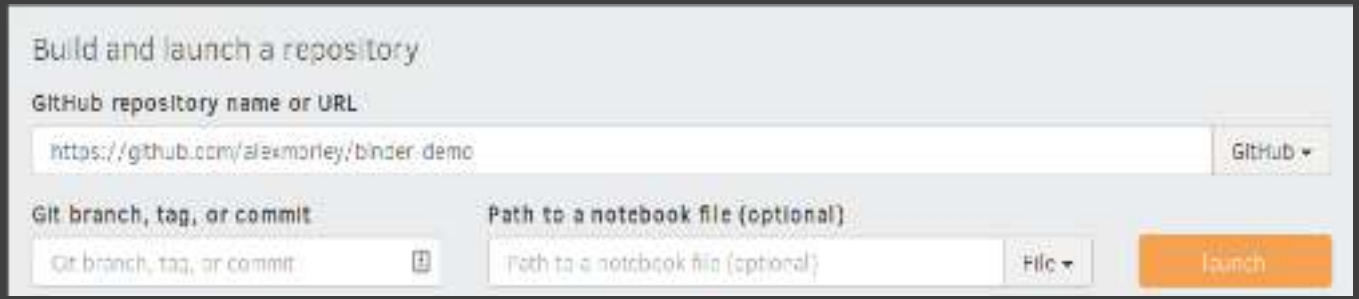
BinderHub

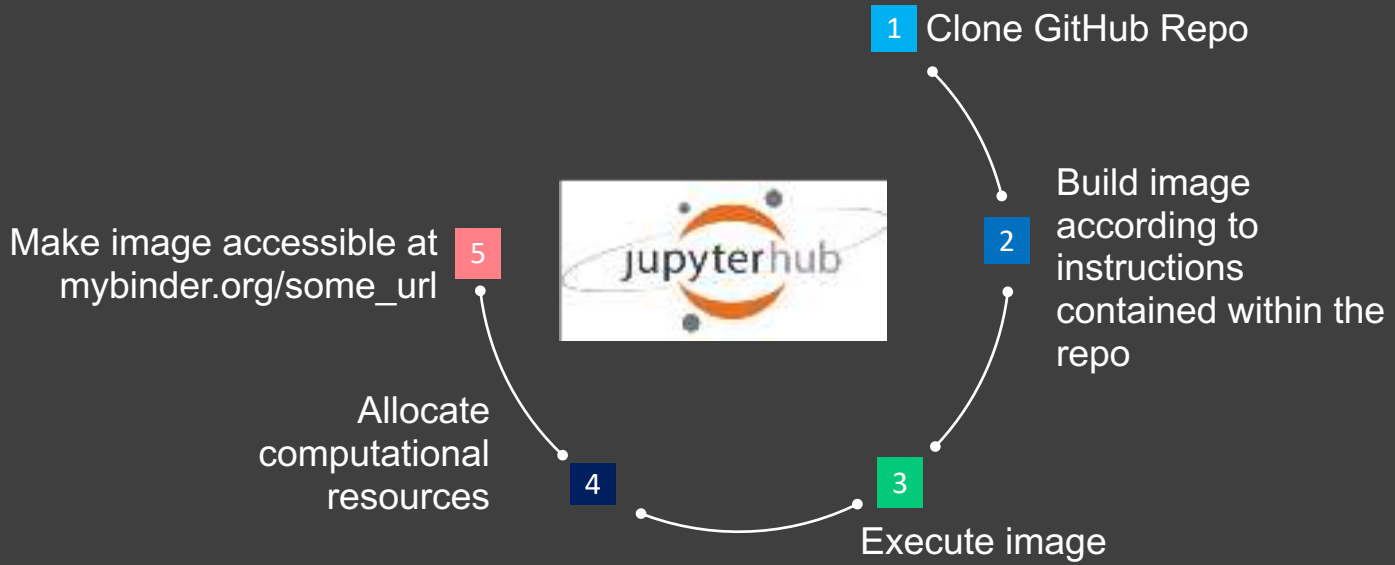




https://doi.org/10.5281/zenodo.3628296

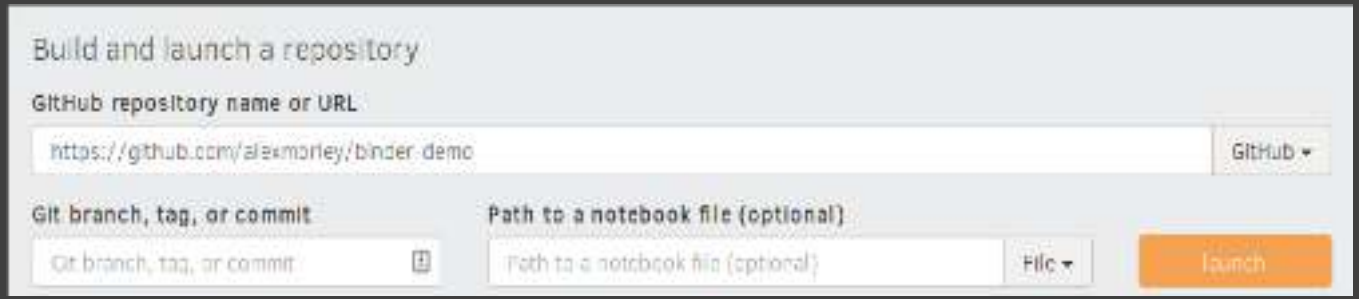
BinderHub

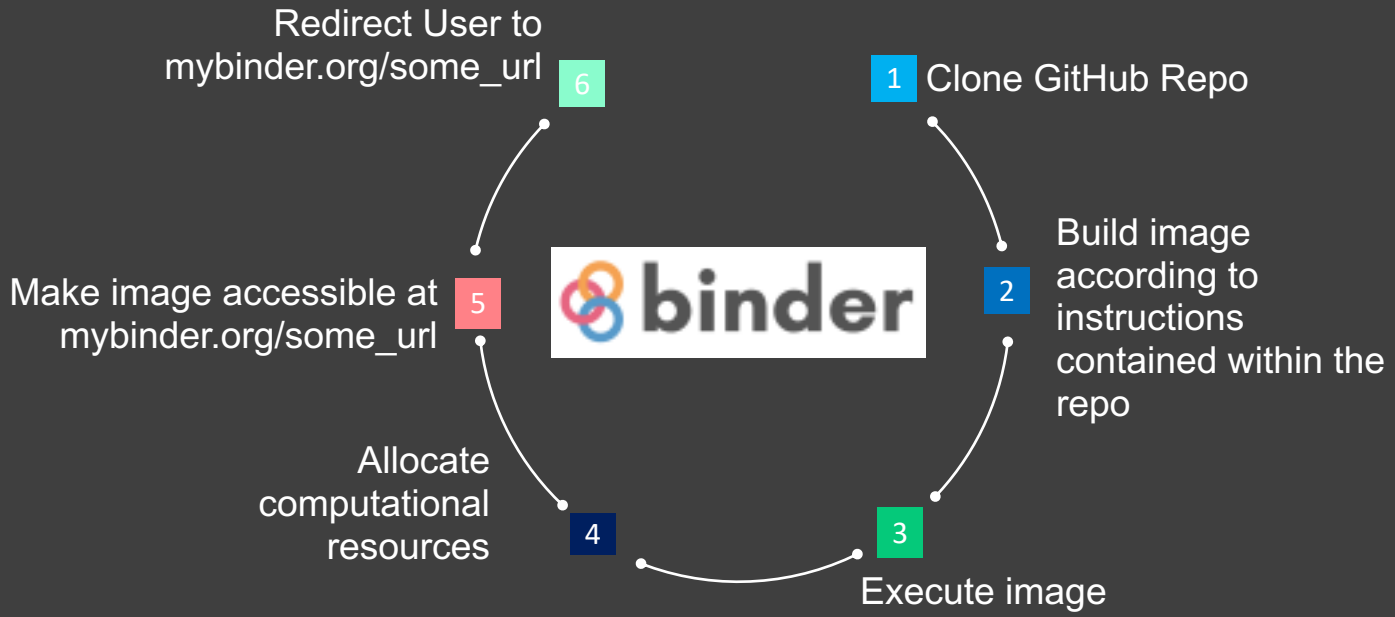




https://doi.org/10.5281/zenodo.3628296

BinderHub



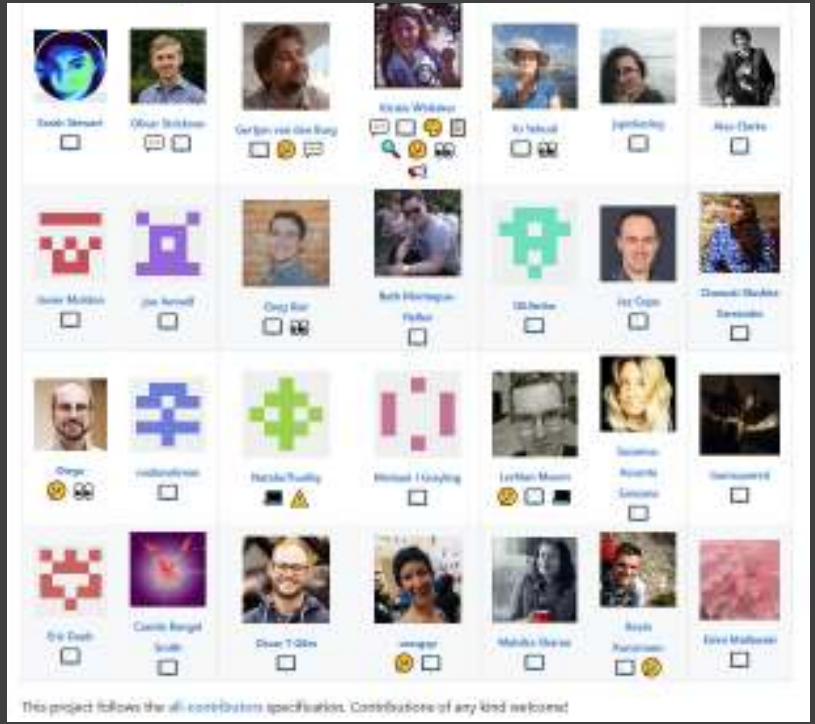


https://doi.org/10.5281/zenodo.3628296

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

#TuringWay

Thank you to our current (and future!) contributors



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

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