The Alan Turing Institute

The Turing Way

Reproducible, inclusive, collaborative data science

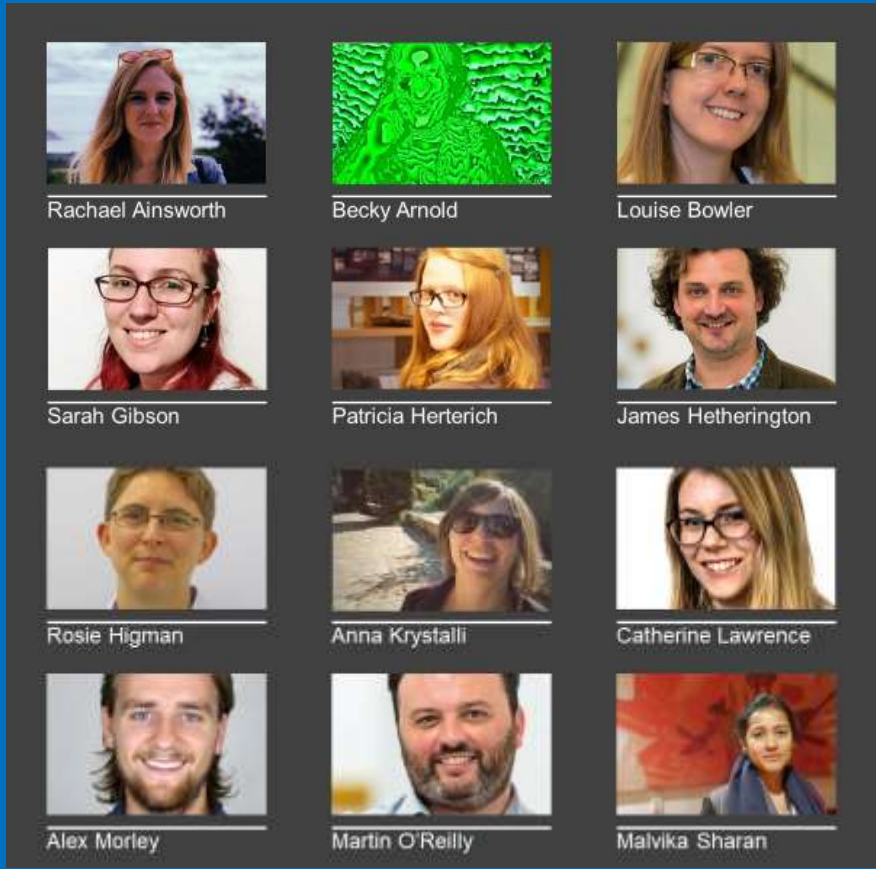
Kirstie Whitaker
Pronouns: she/her

#LoveData20 #TuringWay @kirstie_j

https://doi.org/10.5281/zenodo.3667204

The Turing Way is:

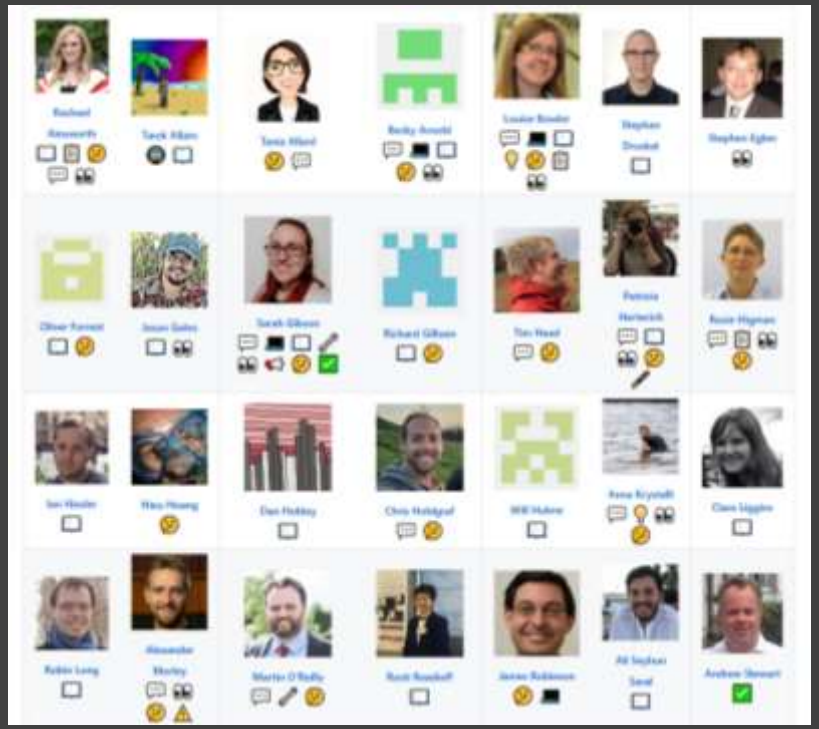
- a book
- a community
- a global collaboration
- a whole tonne of work

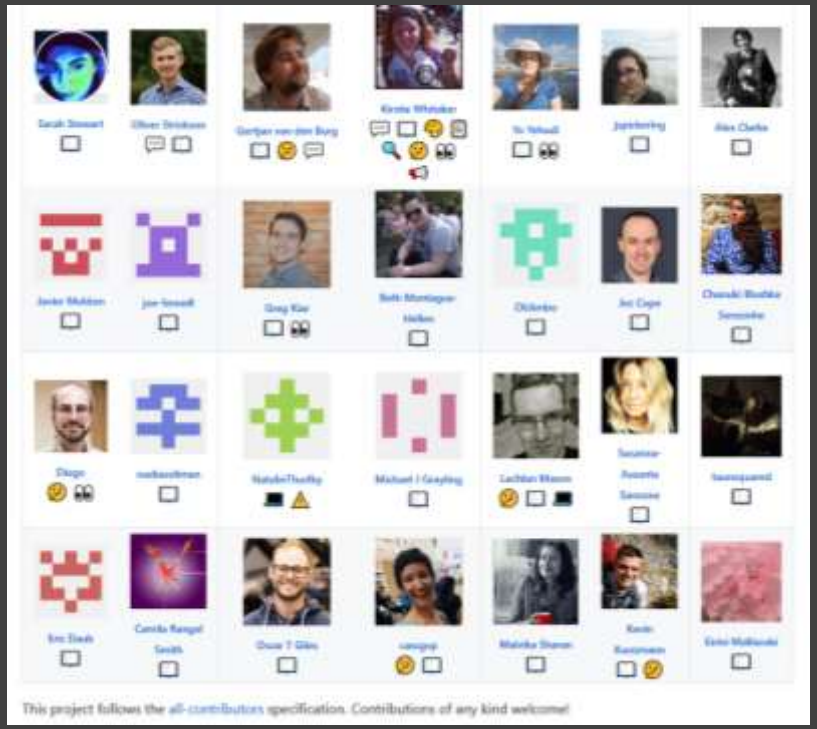


#LoveData20 #TuringWay @kirstie_j

https://doi.org/10.5281/zenodo.3667204

Thank you to all our contributors





https://github.com/alan-turing-institute/the-turing-way#contributors
https://allcontributors.org/docs/en/emoji-key

#LoveData20 #TuringWay @kirstie_j
https://doi.org/10.5281/zenodo.3667204

Picture credit: Chris Gorgolewski
#LoveData20 #TuringWay @kirstie_j
https://doi.org/10.5281/zenodo.3667204

#LoveData20 #TuringWay @kirstie_j
https://doi.org/10.5281/zenodo.3667204

https://www.bbc.co.uk/programmes/p0704h04
#LoveData20 #TuringWay @kirstie_j
https://doi.org/10.5281/zenodo.3667204

https://bletchleypark.org.uk
#LoveData20 #TuringWay @kirstie_j
https://doi.org/10.5281/zenodo.3667204

University network

#LoveData20 #TuringWay @kirstie_j

https://doi.org/10.5281/zenodo.3667204

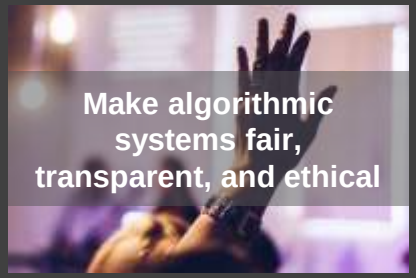
The Institute's partners and collaborators

#LoveData20 #TuringWay @kirstie_j

https://doi.org/10.5281/zenodo.3667204

Challenges

Advance data science and artificial intelligence to…



#LoveData20 #TuringWay @kirstie_j

https://doi.org/10.5281/zenodo.3667204

The Turing Way

#LoveData20 #TuringWay @kirstie_j

https://doi.org/10.5281/zenodo.3667204

The Turing Way

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.

https://the-turing-way.netlify.com/introduction/introduction
#LoveData20 #TuringWay @kirstie_j
https://doi.org/10.5281/zenodo.3667204

The Turing Way

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.

https://the-turing-way.netlify.com/introduction/introduction
#LoveData20 #TuringWay @kirstie_j
https://doi.org/10.5281/zenodo.3667204

The Turing Way

Welcome to the Turing Way

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

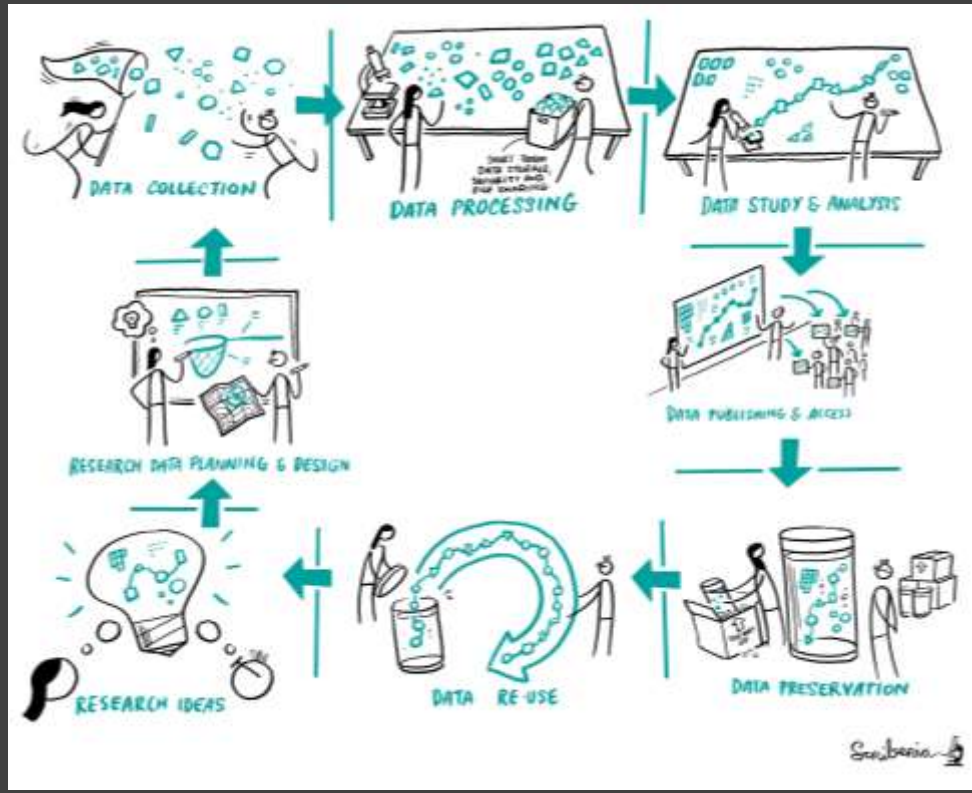
Our goal is to provide all the information that researchers need at the start of their projects to ensure that they are easy to reproduce at the end.

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.

https://the-turing-way.netlify.com/introduction/introduction
#LoveData20 #TuringWay @kirstie_j
https://doi.org/10.5281/zenodo.3667204

To be fully reproducible we have to cover all the steps of the research cycle

And that is super overwhelming…but we're here to help



#LoveData20 #TuringWay @kirstie_j
https://doi.org/10.5281/zenodo.3667204

		Data	
		Same	Different
Analysis	Same	Reproducible	Replicable
	Different	Robust	Generalisable

https://the-turing-way.netlify.com/reproducibility/03/definitions.html
#LoveData20 #TuringWay @kirstie_j
https://doi.org/10.5281/zenodo.3667204

https://the-turing-way.netlify.com/rdm/rdm.html
Wilkinson et al., Sci. Data, 2016. doi: 10.1038/sdata.2016.18

#LoveData20 #TuringWay @kirstie_j
https://doi.org/10.5281/zenodo.3667204

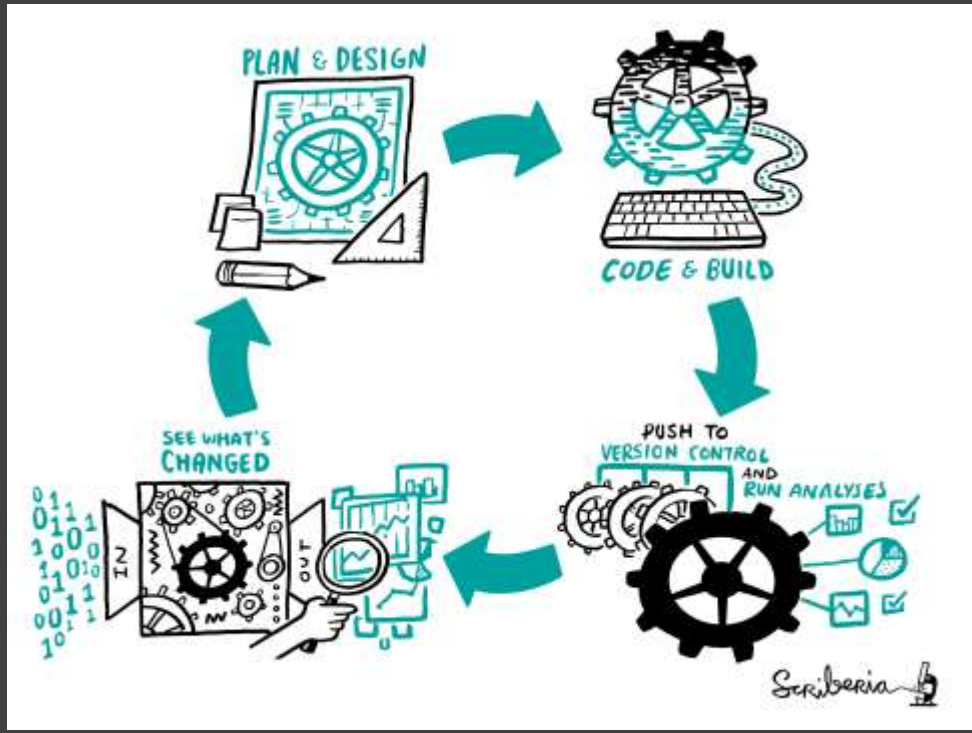
Is your code doing what you think its doing?

```
Assert.AreEqual(
    GetTimeOfDay(),
    “Morning” )
```

https://www.toptal.com/qa/how-to-write-testable-code-and-why-it-matters
https://the-turing-way.netlify.com/testing/testing.html

#LoveData20 #TuringWay @kirstie_j
https://doi.org/10.5281/zenodo.3667204

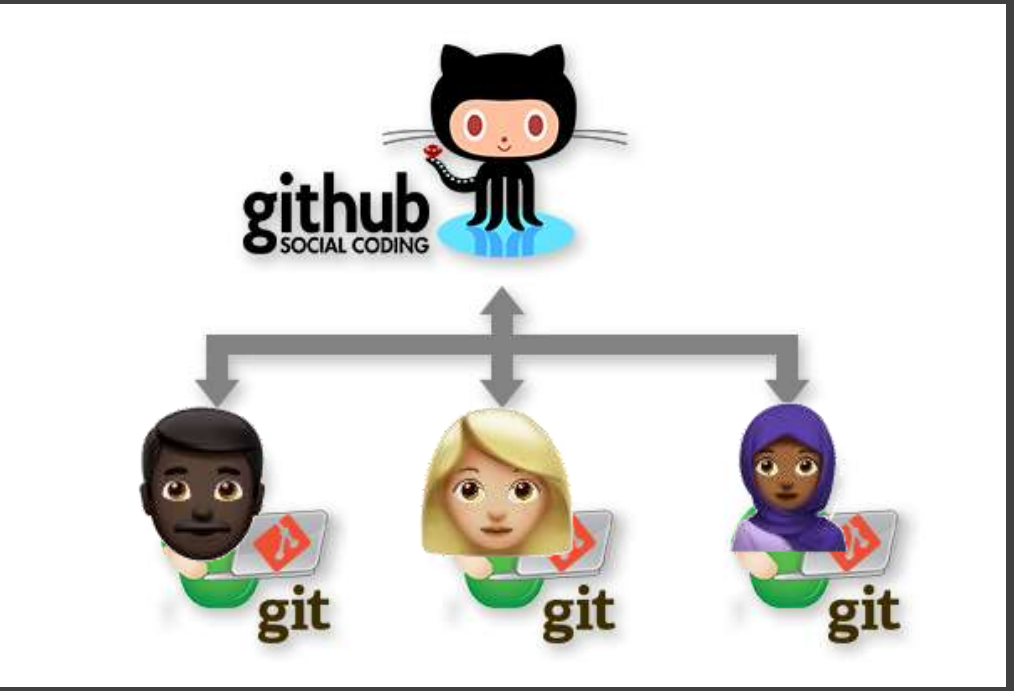
- Plan and design your experiment
- Write down those steps in code
- Push to version control and run the analyses
  - Traditionally done on the cloud, but the important part is that **all steps** are run **every time**
- Test to see what's changed



https://the-turing-way.netlify.com/continuous_integration/continuous_integration.html

#LoveData20 #TuringWay @kirstie_j

https://doi.org/10.5281/zenodo.3667204



https://the-turing-way.netlify.com/collaborating_github/collaborating_github.html
https://the-turing-way.netlify.com/version_control/version_control.html
https://neurohackademy.org

#LoveData20 #TuringWay @kirstie_j
https://doi.org/10.5281/zenodo.3667204

Open Leadership Principles

Understanding
You make the work accessible and clear

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

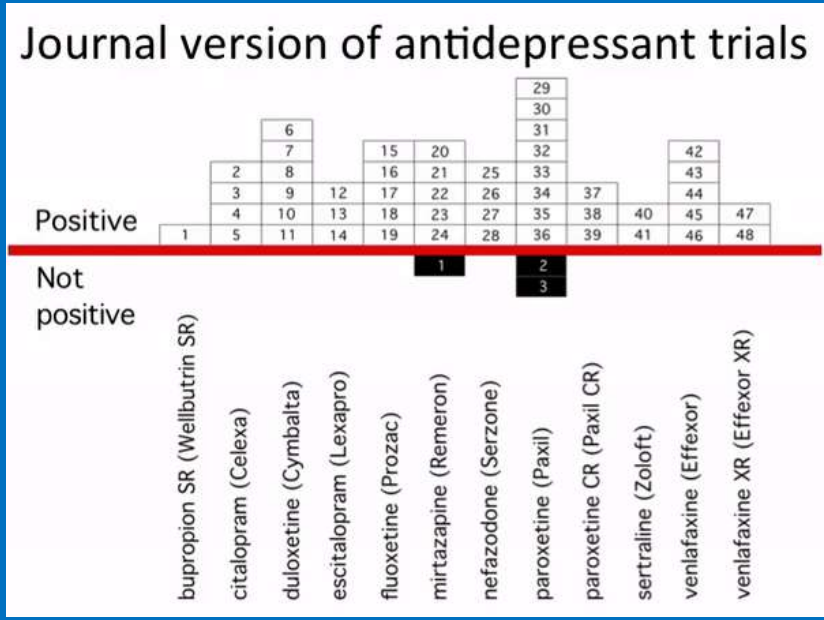
Participation & Inclusion
You build shared ownership and agency to make the work inviting and sustainable for all.

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

moz://a

#LoveData20 #TuringWay @kirstie_j
https://doi.org/10.5281/zenodo.3667204

Ethical and transparent research goes beyond reproducibility



https://twitter.com/eturnermd1/status/737436322344927232
https://doi.org/10.6084/m9.figshare.3381379

#LoveData20 #TuringWay @kirstie_j
https://doi.org/10.5281/zenodo.3667204

Understand factors that contribute to poor research reproducibility
Provide training and disseminate best practice
Support and test interventions to improve reproducibility
Ensure coordination with stakeholders

- Launched March 2019
- Local network leads at >40 UK institutions
- Supported by a range of stakeholders

Marcus Munafò | Laura Fortunato | Malcolm MacLeod | Alex Collins | Chris Chambers

UKRN

@ukrepro #LoveData20 #TuringWay @kirstie_j

https://doi.org/10.5281/zenodo.3667204

UKRN Initiatives

- Registered reports: https://cos.io/rr
- Accountable Replications Policy: https://royalsocietypublishing.org/rsos/replication-studies
- Open Research Working Groups: https://osf.io/vgt3x
- ReproducibiliTea: https://osf.io/3ed8x
- Octopus: https://octopus-hypothesis.netlify.com
- Framework for Open and Reproducible Research Training: https://forrt.netlify.com
- Consortium-Based Student Projects: https://forrt.netlify.com
- Laboratory Efficiency Assessment Framework: https://www.ucl.ac.uk/greenucl/resources/labs/leaf-laboratory-efficiency-assessment-framework
- Open Research Primers: https://www.bristol.ac.uk/psychology/research/ukrn/about/resources

https://doi.org/10.1016/j.tics.2019.12.002
@ukrepro #LoveData20 #TuringWay @kirstie_j
https://doi.org/10.5281/zenodo.3667204

Thank you

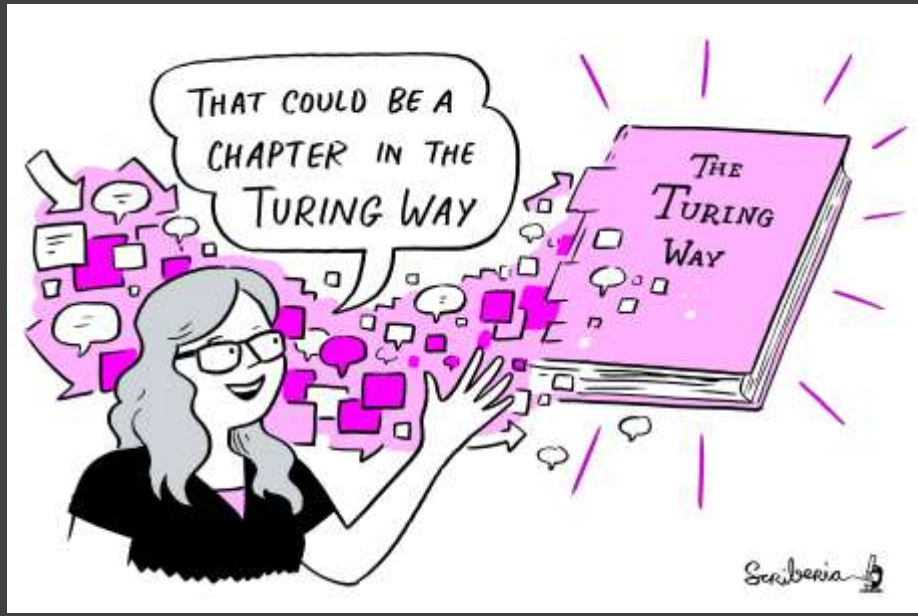
- Book: https://the-turing-way.netlify.com
- Newsletter: https://tinyletter.com/TuringWay
- GitHub: https://github.com/alan-turing-institute/the-turing-way
- Chat: https://gitter.im/alan-turing-institute/the-turing-way
- Next Collaboration Café: 19 February at 7pm UK, 8pm CEST
- This work was supported 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. Unsplash photos by Adolfo Felix,Daniil Silantev, James Pond, Kinson Leung, Mateo Vrbnjak, Toa Heftiba, Thomas Q, Waldemar Brandt. Noun Project icons by Aybige, Luis Prado, Edward Boatman, Becris, Rose Alice Design, Hyemm.work. Original artwork by Scriberia: https://doi.org/10.5281/zenodo.3332807

#LoveData20 #TuringWay @kirstie_j

https://doi.org/10.5281/zenodo.3667204

Extension in 2020

- Expand scope to all data science practices
  - Reproducibility
  - Scoping and designing a data science project
  - Ethics
  - Communication and visualisation
  - Collaborative working



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

#LoveData20 #TuringWay @kirstie_j

https://doi.org/10.5281/zenodo.3667204

My definition of data science

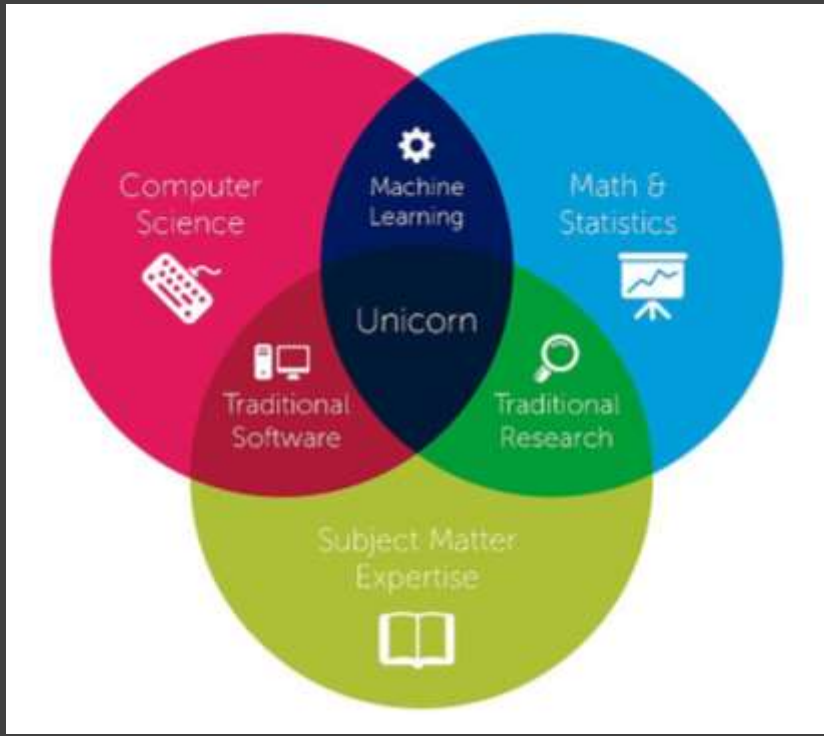
“Applying software, data management or data analysis practices from one domain to another, including the reuse of data, to answer a new research question.”

#LoveData20 #TuringWay @kirstie_j

https://doi.org/10.5281/zenodo.3667204

Data science is:

– innovative
– collaborative
– interdisciplinary research



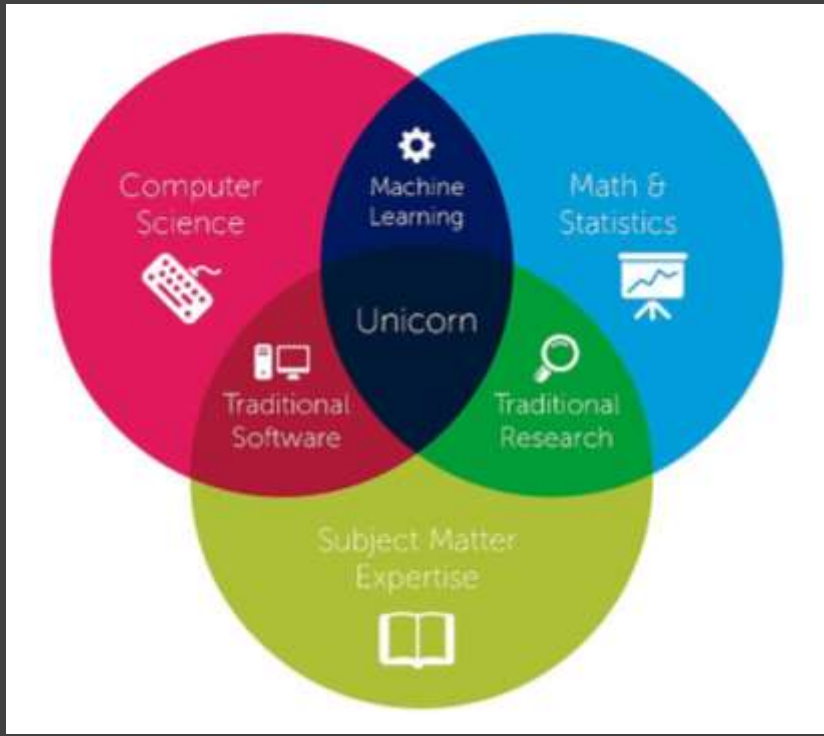
https://www.luther.edu/computer-science/data-science-major/why-study

https://doi.org/10.5281/zenodo.3667204

Data science is:

– innovative
– collaborative
– interdisciplinary

research delivered by teams of experts



https://www.luther.edu/computer-science/data-science-major/why-study

https://doi.org/10.5281/zenodo.3667204