

Open and reproducible science

WHY AND HOW

Sara Villa

*Postdoctoral researcher @Denk's group
The Turing Way Contributor
OLS Open Seeds programme mentor*

sara.villa@kcl.ac.uk



@VillaScience

ALL IMAGES TAKEN FROM *The Turing Way* project illustration
by Scriberia. Used under a CC-BY 4.0 licence.
<http://doi.org/10.5281/zenodo.3332807>

OPEN
RESEARCH



OPPORTUNITY?

SENSITIVITY?

INCLUSIVITY?

FUNDING?

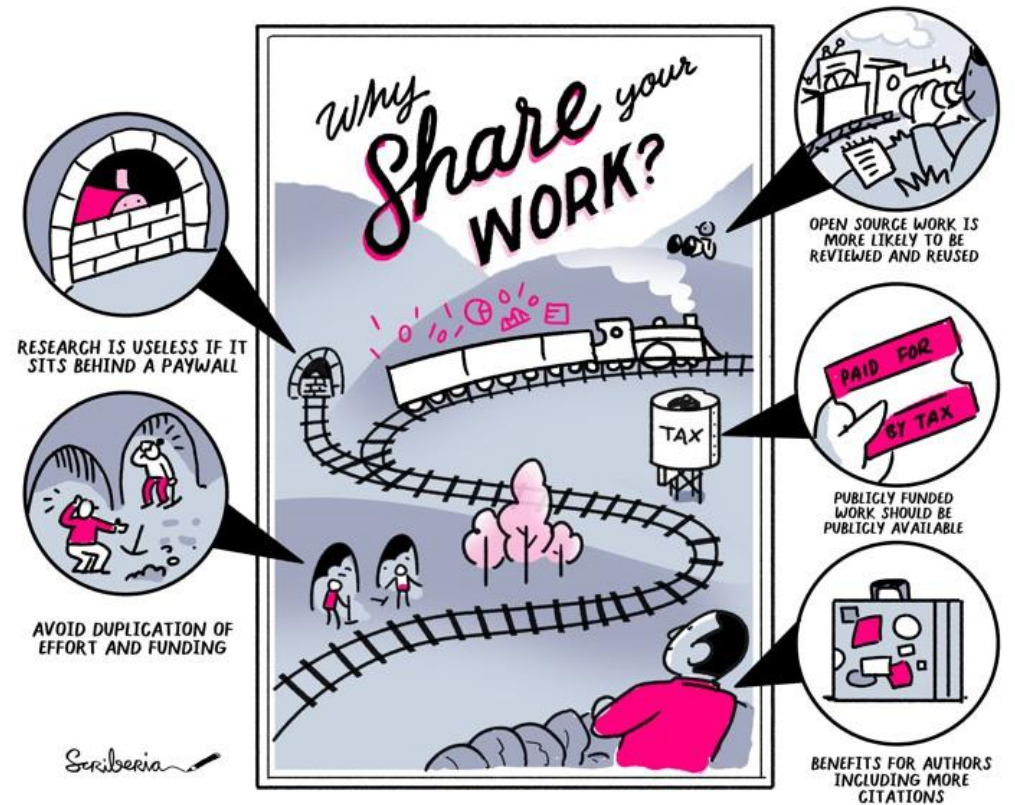
CLOSED
RESEARCH



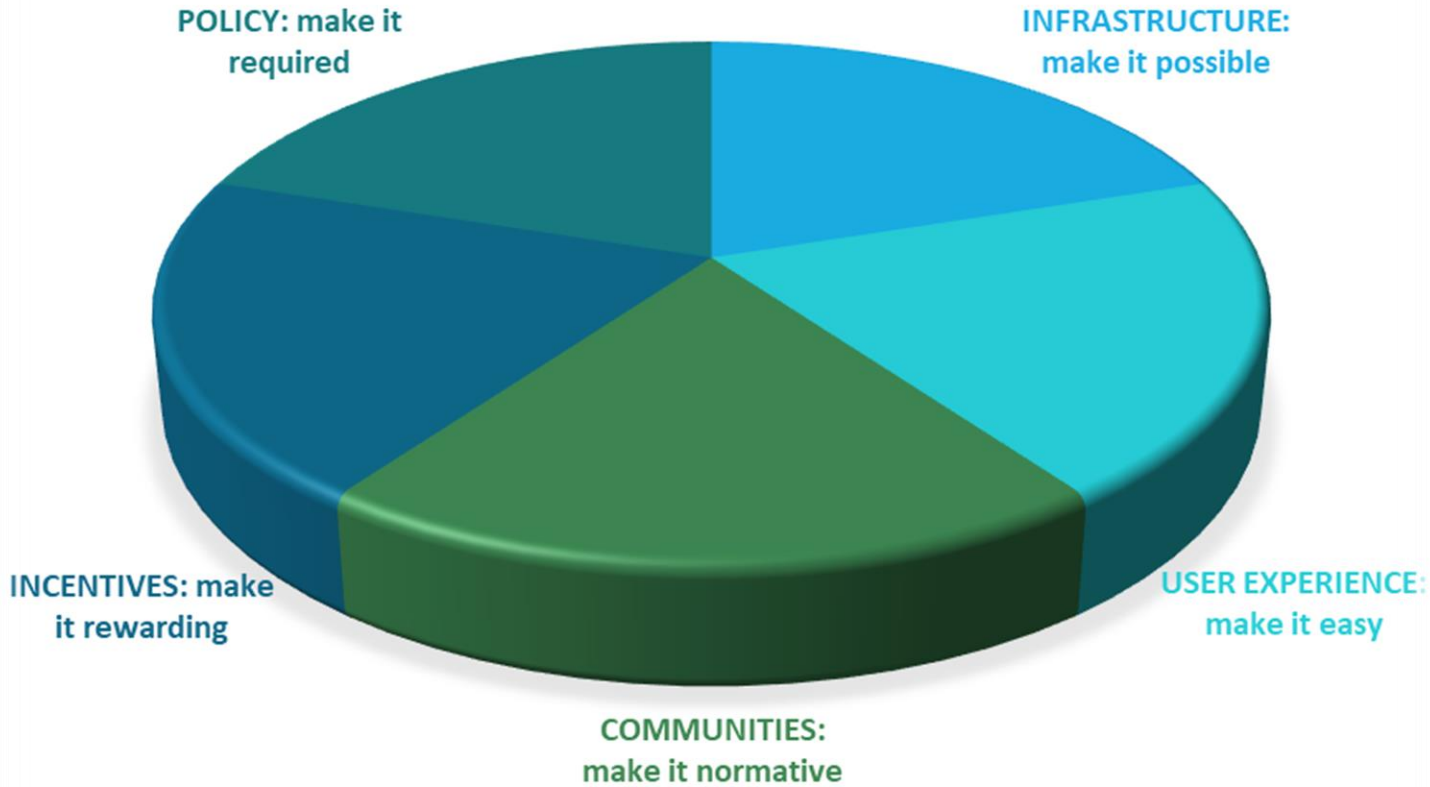
Scriberia 

WHY Open science

School	Belief	Aim
Infrastructure	Efficient research depends on the available tools and applications.	Creating openly available platforms, tools, and services for researchers.
Pragmatic	Knowledge-creation could be more efficient if researchers worked together.	Opening up the process of knowledge creation.
Measurement	Academic contributions today need alternative impact measurements.	Developing an alternative metric system for research impact.
Democratic	The access to knowledge is unequally distributed.	Making knowledge freely available for everyone.
Public	Research needs to be made accessible to the public.	Making research accessible for citizens.

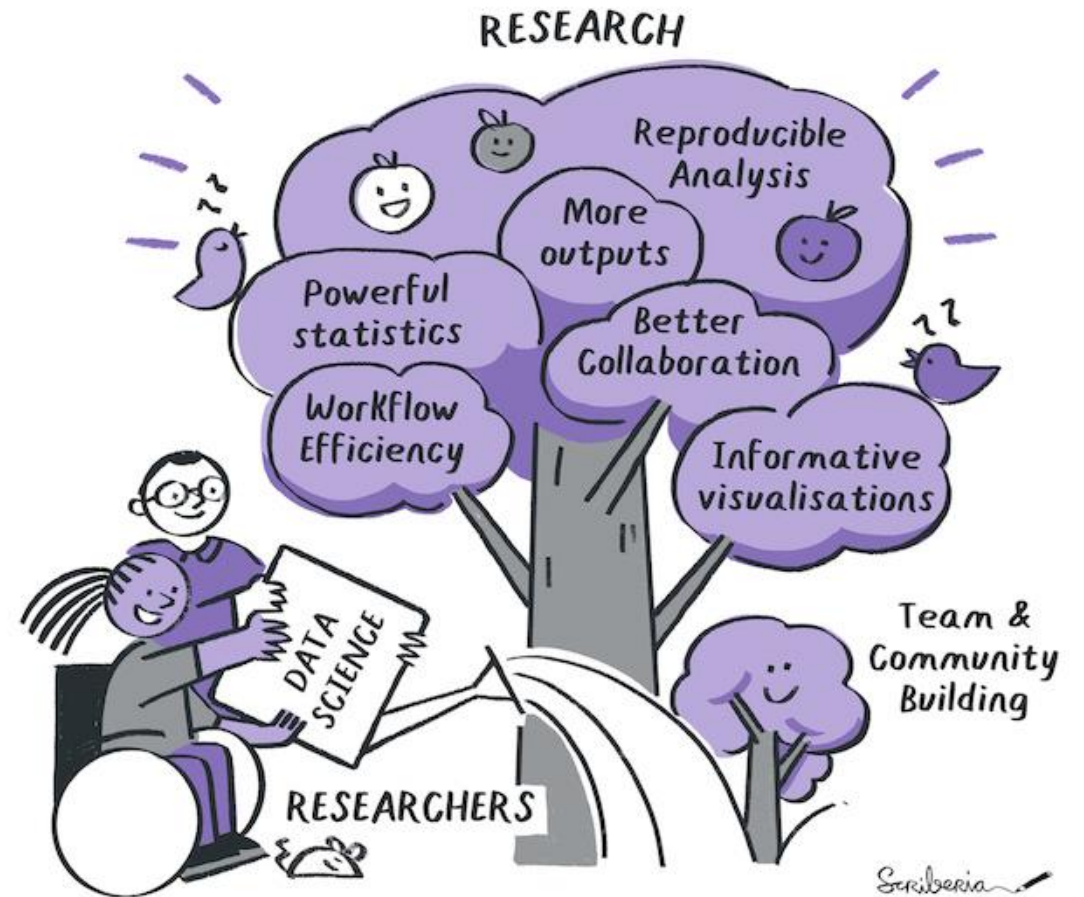


CHANGING RESEARCH CULTURE

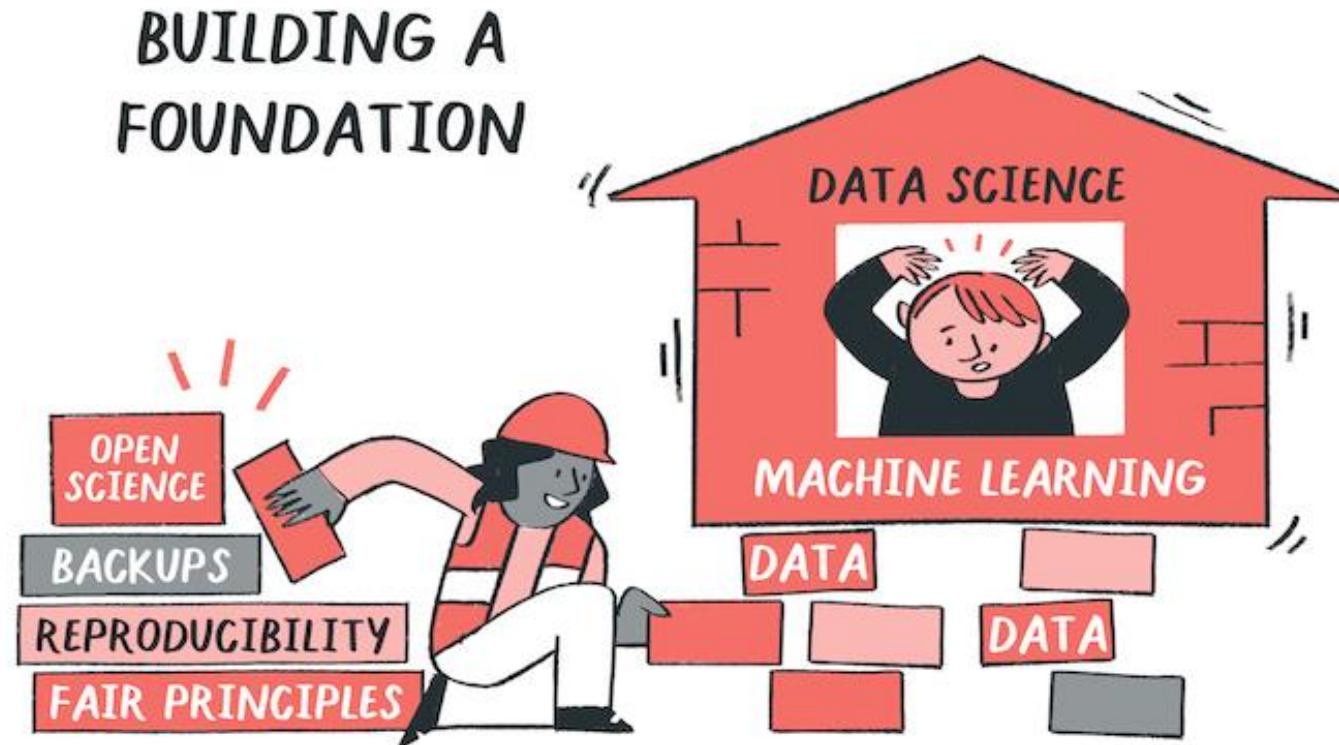


Reinterpreted from <https://www.cos.io/blog/cos-celebrates-10-years>

Every research project is a data science project



OS principles can be applied to ANY research field



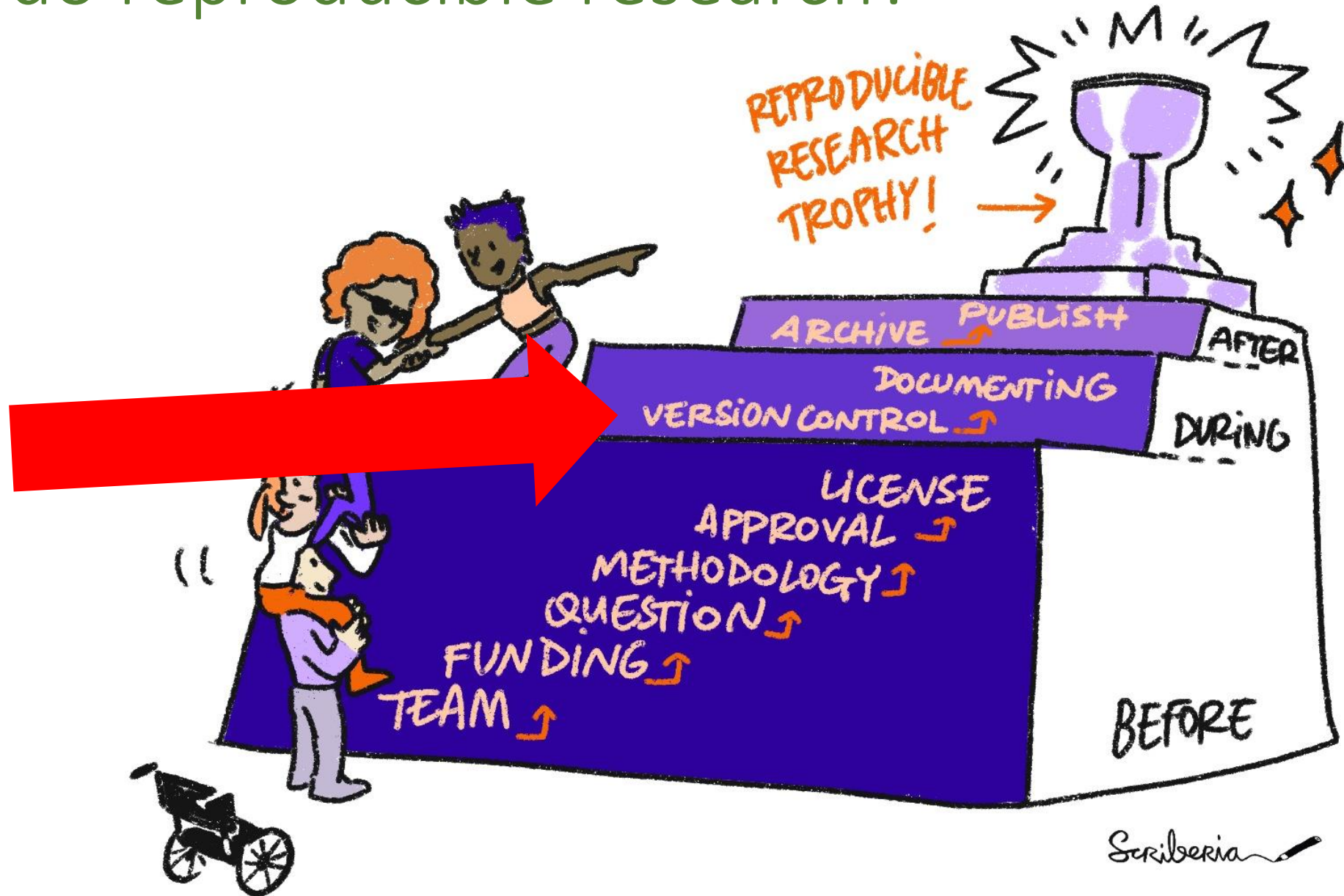
<https://rdmkit.elixir-europe.org/>

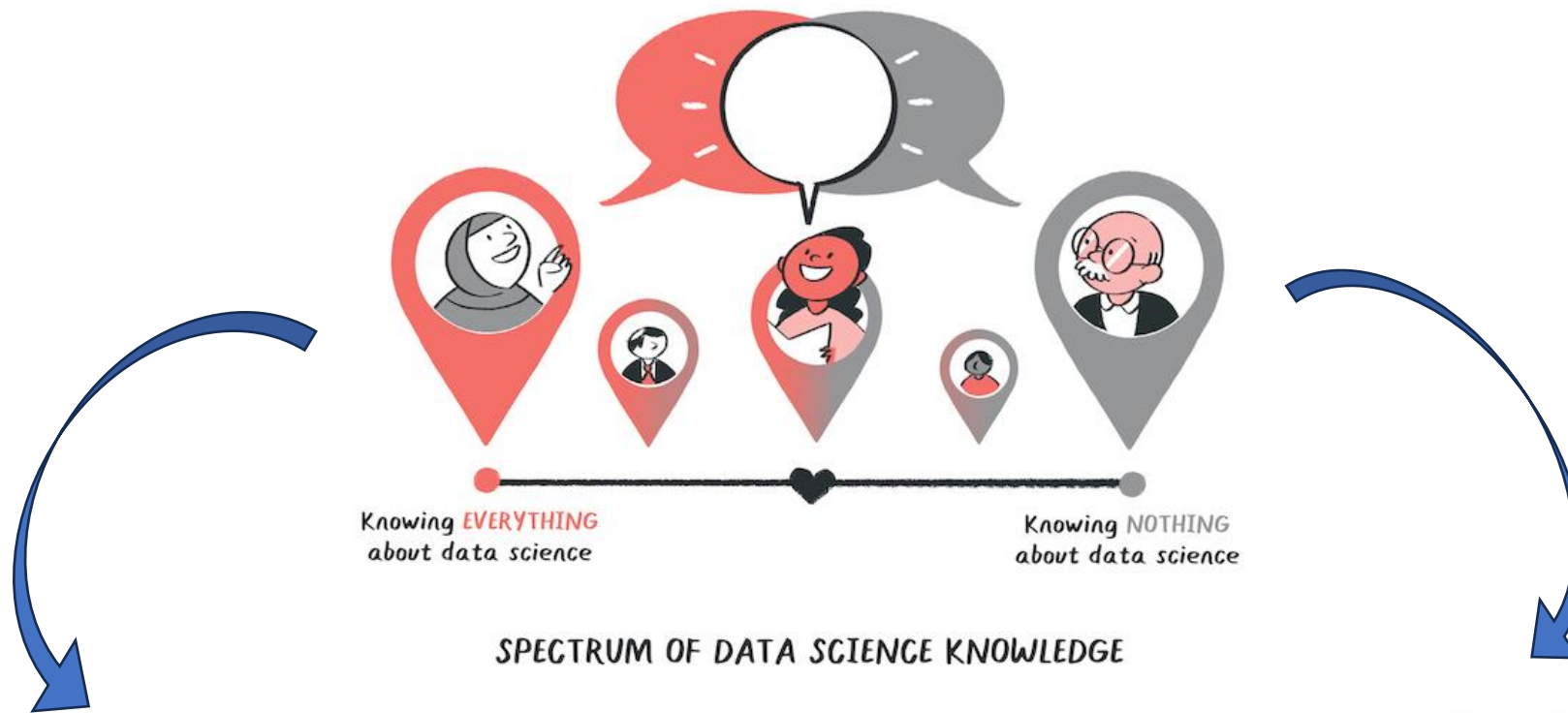
<https://osf.io/>







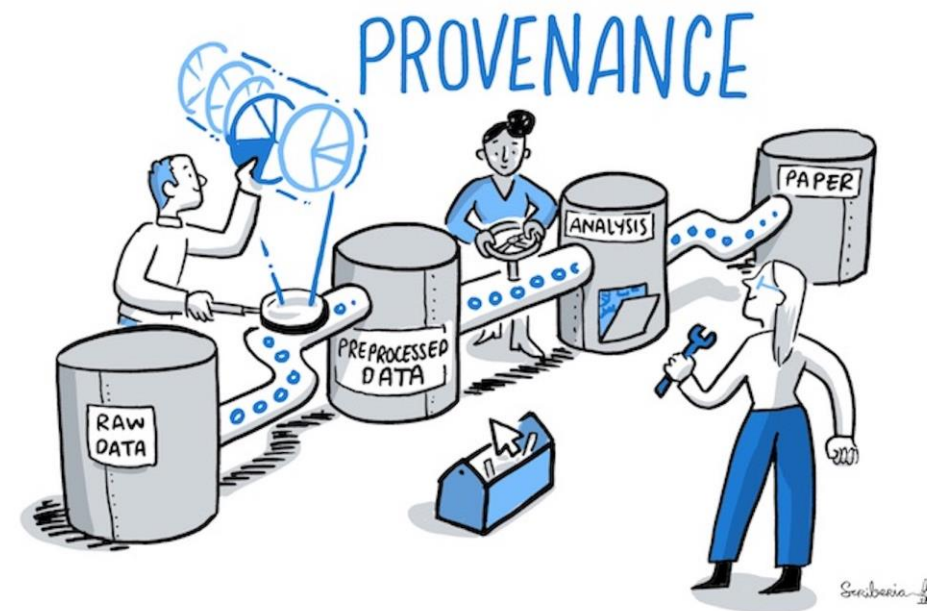
<https://www.go-fair.org/fair-principles/i1-metadata-use-formal-accessible-shared-broadly-applicable-language-knowledge-representation/>

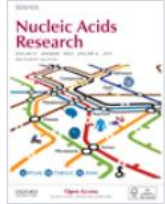
How to do reproducible research?





Interaction style	Graphical	Command line
What is reproduced?		
Software & versions	 binder	 CONDA
Entire system		





Volume 47, Issue D1
08 January 2019

Article Contents

Abstract

- INTRODUCTION
- INTERACTIVE ANALYSIS REPORTS
- DATA MINING
- INTEGRATIVE DATA ANALYSIS TOOLS
- CONCLUSION
- DATA AVAILABILITY
- COMPUTATIONAL RESOURCES
- SUPPLEMENTARY DATA
- ACKNOWLEDGEMENTS

JOURNAL ARTICLE

Stemformatics: visualize and download curated stem cell data

Jarny Choi, Chris M Pacheco, Rowland Mosbergen, Othmar Korn, Tyrone Chen, Isha Nagpal, Steve Englart, Paul W Angel, Christine A Wells

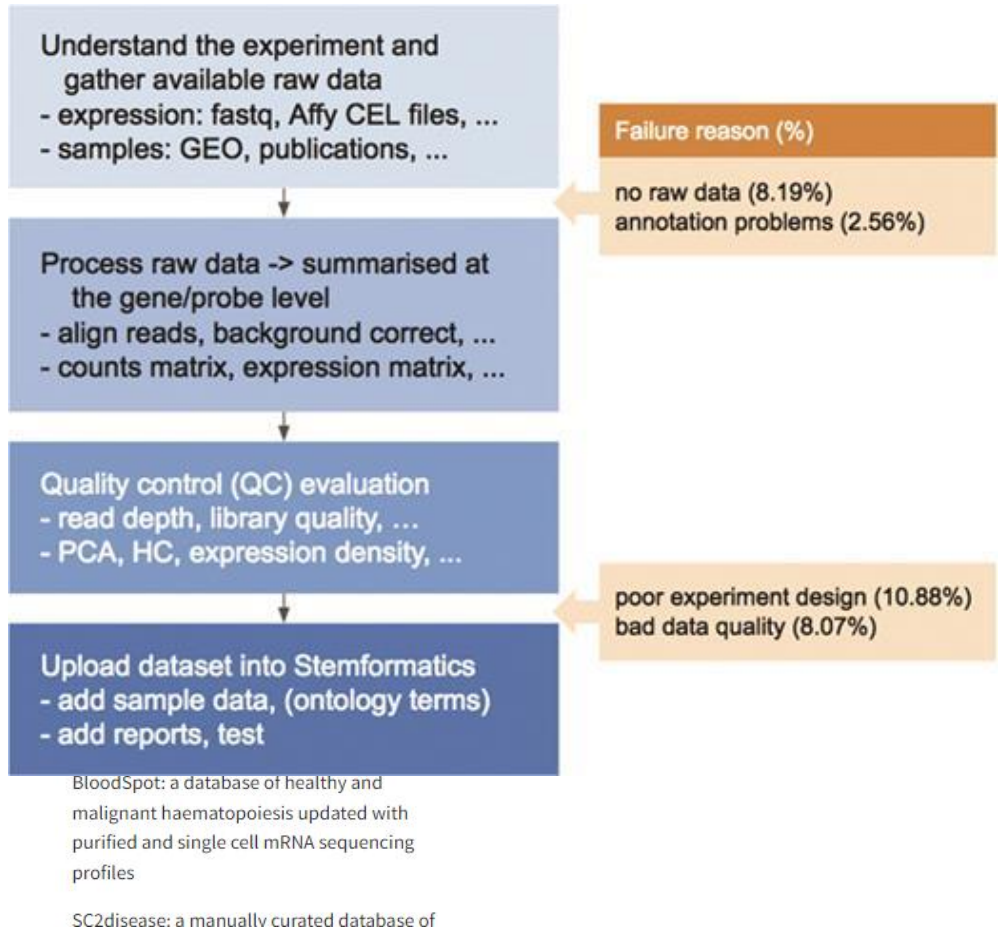
Nucleic Acids Research, Volume 47, Issue D1, 08 January 2019, Pages D841–D846,
<https://doi.org/10.1093/nar/gky1064>

Published: 08 November 2018 [Article history ▾](#)

PDF Split View Cite Permissions Share ▾

Abstract

Stemformatics is an established gene expression data portal containing over 420 public gene expression datasets derived from microarray, RNA sequencing and single cell profiling technologies. Developed for the stem cell community, it has a major focus on pluripotency, tissue stem cells, and staged differentiation. Stemformatics includes curated ‘collections’ of data relevant to cell reprogramming, as well as hematopoiesis and leukaemia. Rather than simply rehosting datasets as they appear in public repositories, Stemformatics uses a stringent set of quality control metrics and its own pipelines to process handpicked datasets from raw files. This means that about 30% of datasets processed by Stemformatics fail the quality control metrics and never make it to the portal, ensuring that Stemformatics data are of high quality and have been processed in a consistent manner. Stemformatics provides easy-to-use and intuitive tools for biologists to visually explore the data, including



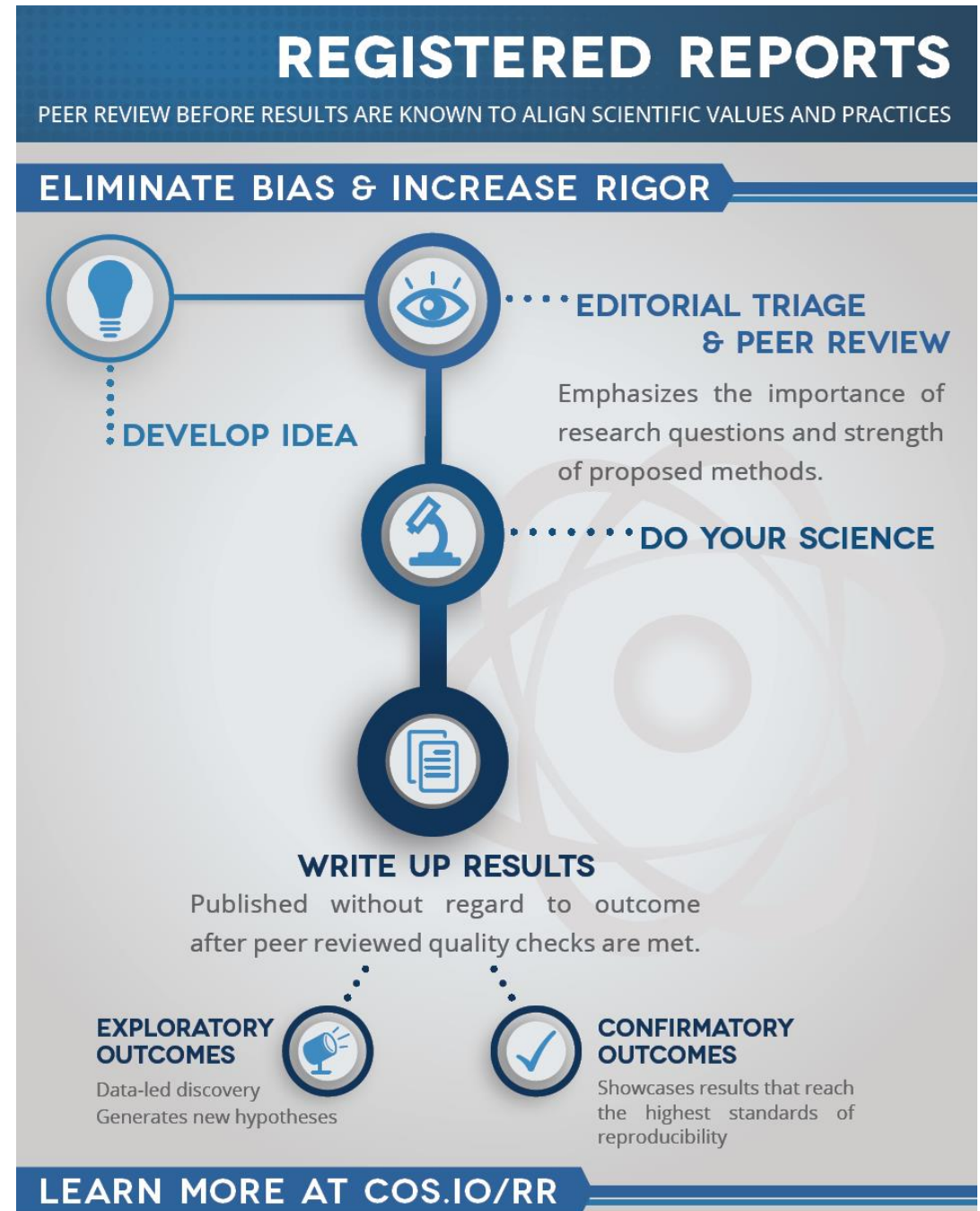
Insufficient sample size,
power calculations...
are basics often driving
to a poor quality study

https://lakens.github.io/statistical_inferences/

Feedback from
project design
and
publish
negative results
too!

<https://www.bnacredibility.org.uk/academia>

<https://www.cos.io/initiatives/registered-reports>



Version control



TRACK PROJECT HISTORY



Scriberia 

Learning new tools is worthy for your research most of the times...

HOW LONG CAN YOU WORK ON MAKING A ROUTINE TASK MORE EFFICIENT BEFORE YOU'RE SPENDING MORE TIME THAN YOU SAVE?
(ACROSS FIVE YEARS)

		HOW OFTEN YOU DO THE TASK					
		50/DAY	5/DAY	DAILY	WEEKLY	MONTHLY	YEARLY
HOW MUCH TIME YOU SHAVE OFF	1 SECOND	1 DAY	2 HOURS	30 MINUTES	4 MINUTES	1 MINUTE	5 SECONDS
	5 SECONDS	5 DAYS	12 HOURS	2 HOURS	21 MINUTES	5 MINUTES	25 SECONDS
	30 SECONDS	4 WEEKS	3 DAYS	12 HOURS	2 HOURS	30 MINUTES	2 MINUTES
	1 MINUTE	8 WEEKS	6 DAYS	1 DAY	4 HOURS	1 HOUR	5 MINUTES
	5 MINUTES	9 MONTHS	4 WEEKS	6 DAYS	21 HOURS	5 HOURS	25 MINUTES
	30 MINUTES		6 MONTHS	5 WEEKS	5 DAYS	1 DAY	2 HOURS
	1 HOUR		10 MONTHS	2 MONTHS	10 DAYS	2 DAYS	5 HOURS
	6 HOURS				2 MONTHS	2 WEEKS	1 DAY
1 DAY					8 WEEKS	5 DAYS	

What does a quantitative researcher do?

Read literature

Write papers

Bring in funding

Collect / access data

Design experiments

Set strategy

Mentor colleagues

Run statistical analyses

Promote work to decision makers

Manage a team

@TuringWay @kirstie_j

Interpret results

<https://doi.org/10.5281/zenodo.7749650>

Peer review

Read
literature

Write papers

Bring in funding

Collect / access data

Design
experiments

What does a quantitative researcher do?

Set strategy

Testing &
quality control

Mentor
colleagues

Manage budgets

Run statistical analyses

Promote work to
decision makers

Report on
progress

Manage a team

@TuringWay @kirstie_j

Interpret results

<https://doi.org/10.5281/zenodo.7749650>

What does a quantitative researcher do?

Peer review

Manage (cloud)

compute

resources

Deliver trainings

Write papers

Bring in funding

Collect / access data

Set strategy

Run statistical analyses

Scope user requirements

@TuringWay @kirstie_j

@TuringWay @kirstie_j

<https://doi.org/10.5281/zenodo.7749650>

Report on progress

Manage a team

Visualisation

Promote work to decision makers

Manage budgets

Mentor colleagues

Generalise code to software

Interpret results

Testing & quality control

Design experiments

Read literature

What does a quantitative researcher do?

Peer review

Manage (cloud)

compute

resources Archive data &

code
Write papers

Bring in funding

Collect / access data

Set strategy

Run statistical analyses

Scope user requirements

@TuringWay @kirstie_j

<https://doi.org/10.5281/zenodo.7749650>

Manage a team

Visualisation

Promote work to decision makers

Report on progress

Manage budgets

Mentor colleagues

Generalise code to software

Interpret results

Testing & quality control

Design experiments

Deliver trainings

Read literature

What does a quantitative researcher do?

Read literature
Design experiments
Testing & quality control
Mentor colleagues
Generalise code to software
Interpret results

Deliver trainings
Write papers
Track impact
Promote work to decision makers
Report on progress

Manage (cloud) compute resources
Archive data & code
Write papers
Manage budgets
Promote work to decision makers
Report on progress

Commercialise a product
Bring in funding
Collect / access data
Set strategy
Run statistical analyses
Visualisation
Manage a team

Peer review
Bring in funding
Collect / access data
Set strategy
Run statistical analyses
Visualisation
Manage a team
Scope user requirements

@TuringWay @kirstie_j
<https://doi.org/10.5281/zenodo.7749650>

Attend / chair meetings
Read literature
Design experiments
Testing & quality control
Mentor colleagues
Generalise code to software
Interpret results

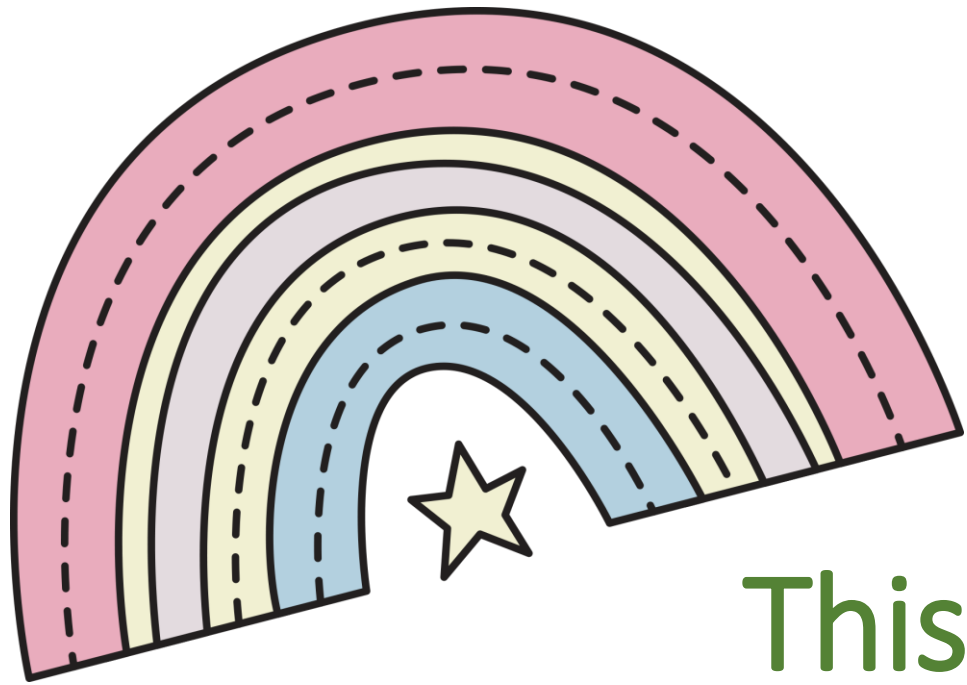
Manage (cloud) compute resources
Deliver trainings
Share work with the public
Track impact
Manage budgets
Promote work to decision makers
Report on progress

Commercialise a product
Archive data & code
Give talks
Collect / access data
Set strategy
Maintain ethical & legal principles
Run statistical analyses
Host events
Visualisation
Manage a team

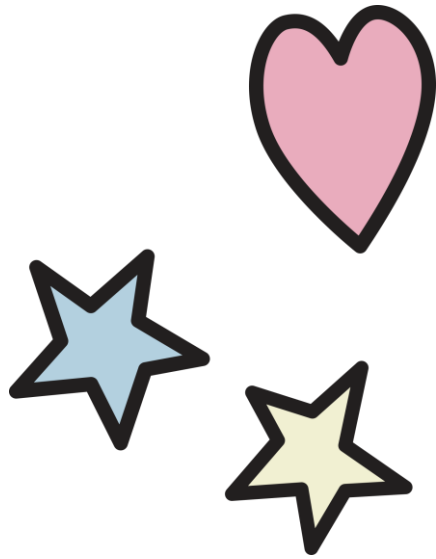
Peer review
Promote EDI and work-life balance
Bring in funding
Maintain ethical & legal principles
Scope user requirements

What does a quantitative researcher do?

@TuringWay @kirstie_j
<https://doi.org/10.5281/zenodo.7749650>

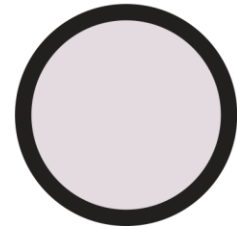
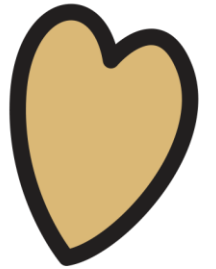
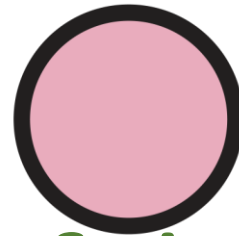
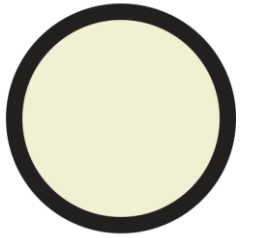
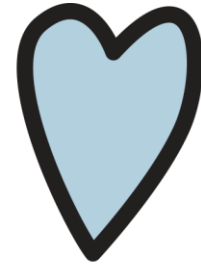
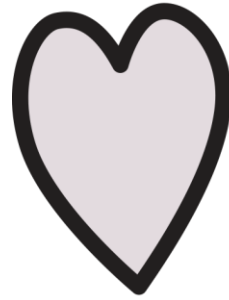


This person is a unicorn



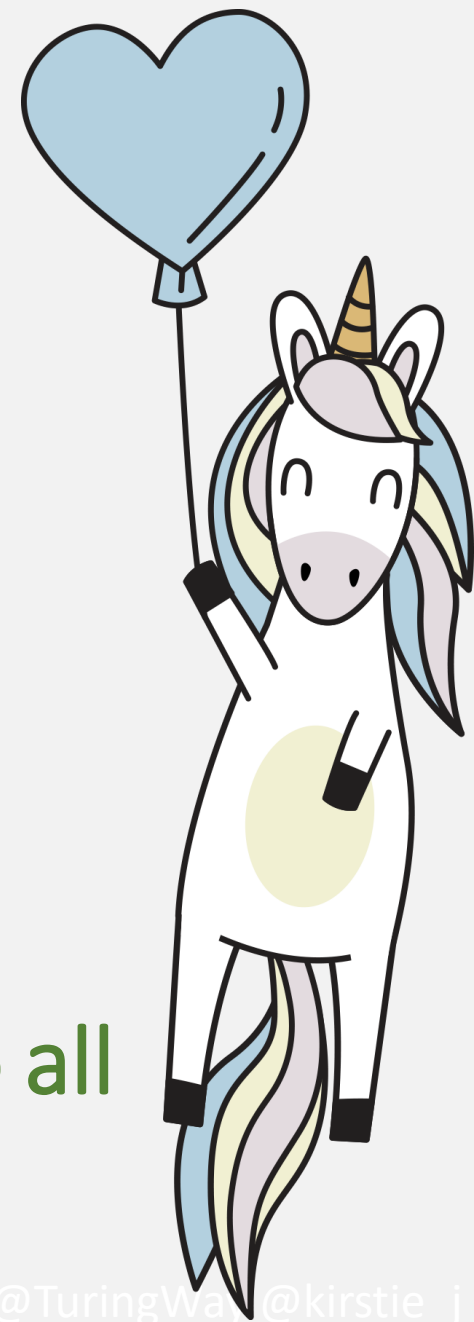


They are delightful



But they don't
exist

We can't expect individuals to be able to do all
these tasks to the highest standard



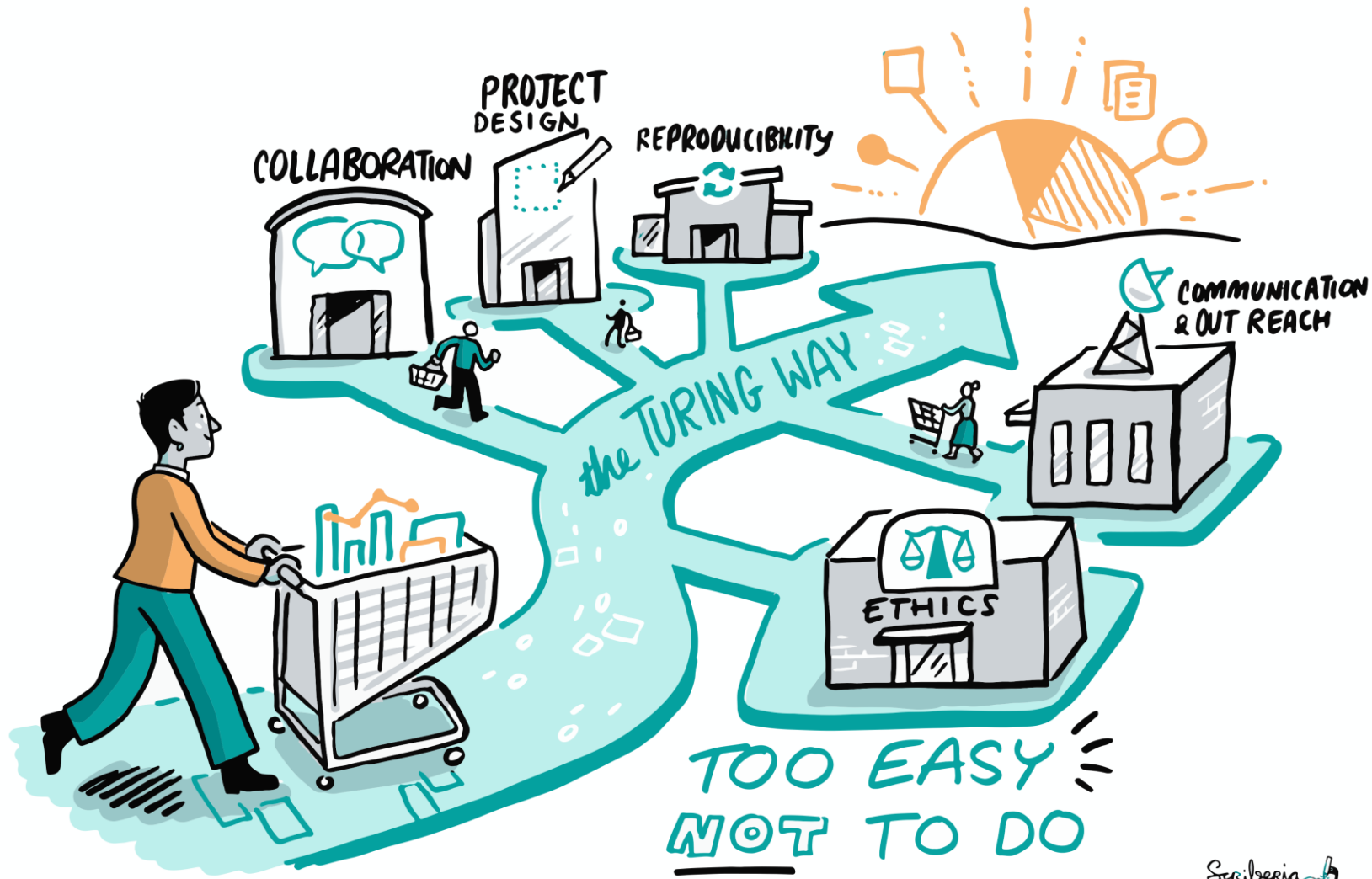
There is one way...

Learning new tools might be worthy... **or it might not sometimes**

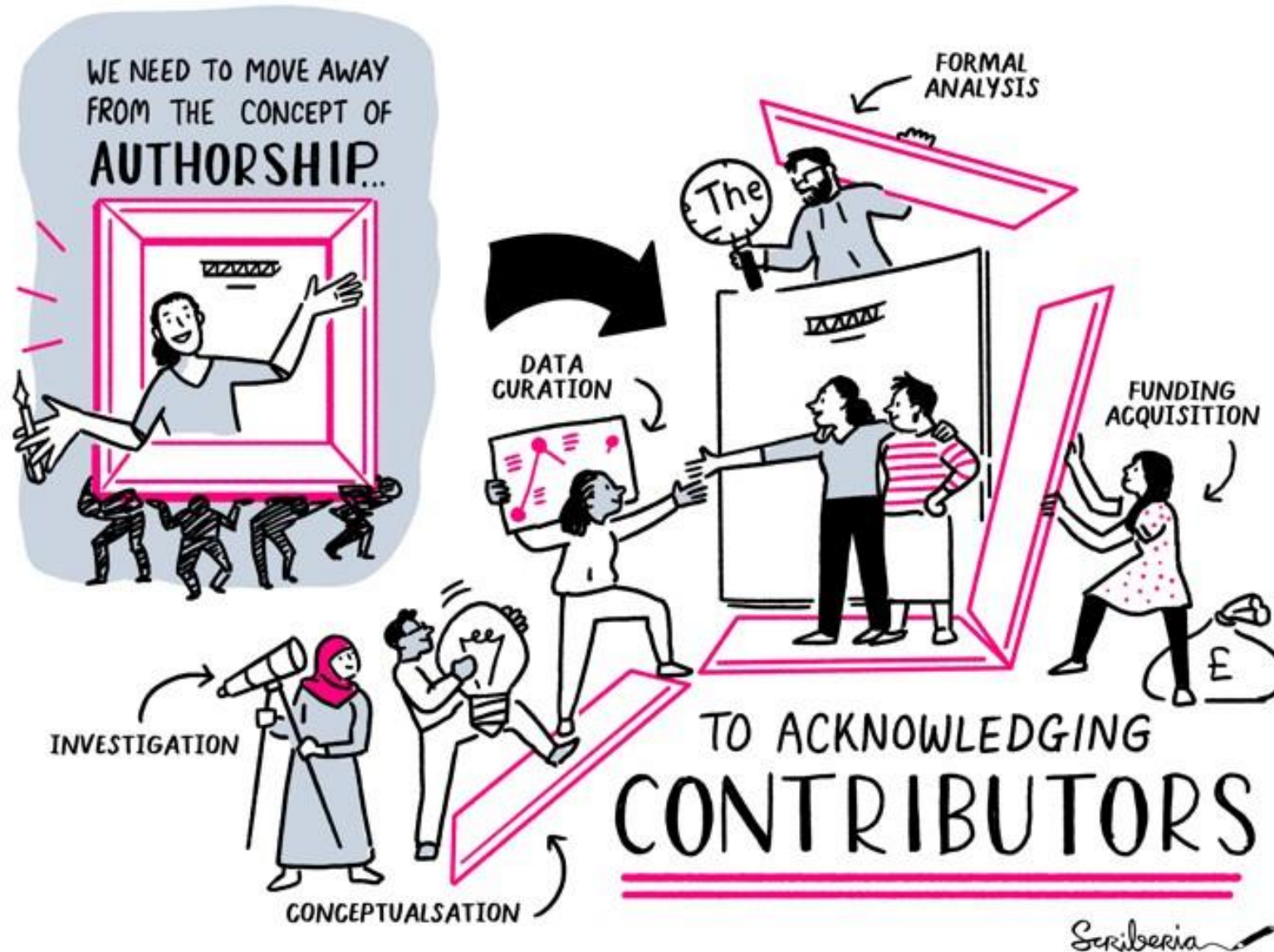
HOW LONG CAN YOU WORK ON MAKING A ROUTINE TASK MORE EFFICIENT BEFORE YOU'RE SPENDING MORE TIME THAN YOU SAVE?
(ACROSS FIVE YEARS)

		HOW OFTEN YOU DO THE TASK					
		50/DAY	5/DAY	DAILY	WEEKLY	MONTHLY	YEARLY
HOW MUCH TIME YOU SHAVE OFF	1 SECOND	1 DAY	2 HOURS	30 MINUTES	4 MINUTES	1 MINUTE	5 SECONDS
	5 SECONDS	5 DAYS	12 HOURS	2 HOURS	21 MINUTES	5 MINUTES	25 SECONDS
	30 SECONDS	4 WEEKS	3 DAYS	12 HOURS	2 HOURS	30 MINUTES	2 MINUTES
	1 MINUTE	8 WEEKS	6 DAYS	1 DAY	4 HOURS	1 HOUR	5 MINUTES
	5 MINUTES	9 MONTHS	4 WEEKS	6 DAYS	21 HOURS	5 HOURS	25 MINUTES
	30 MINUTES		6 MONTHS	5 WEEKS	5 DAYS	1 DAY	2 HOURS
	1 HOUR		10 MONTHS	2 MONTHS	10 DAYS	2 DAYS	5 HOURS
	6 HOURS				2 MONTHS	2 WEEKS	1 DAY
	1 DAY					8 WEEKS	5 DAYS

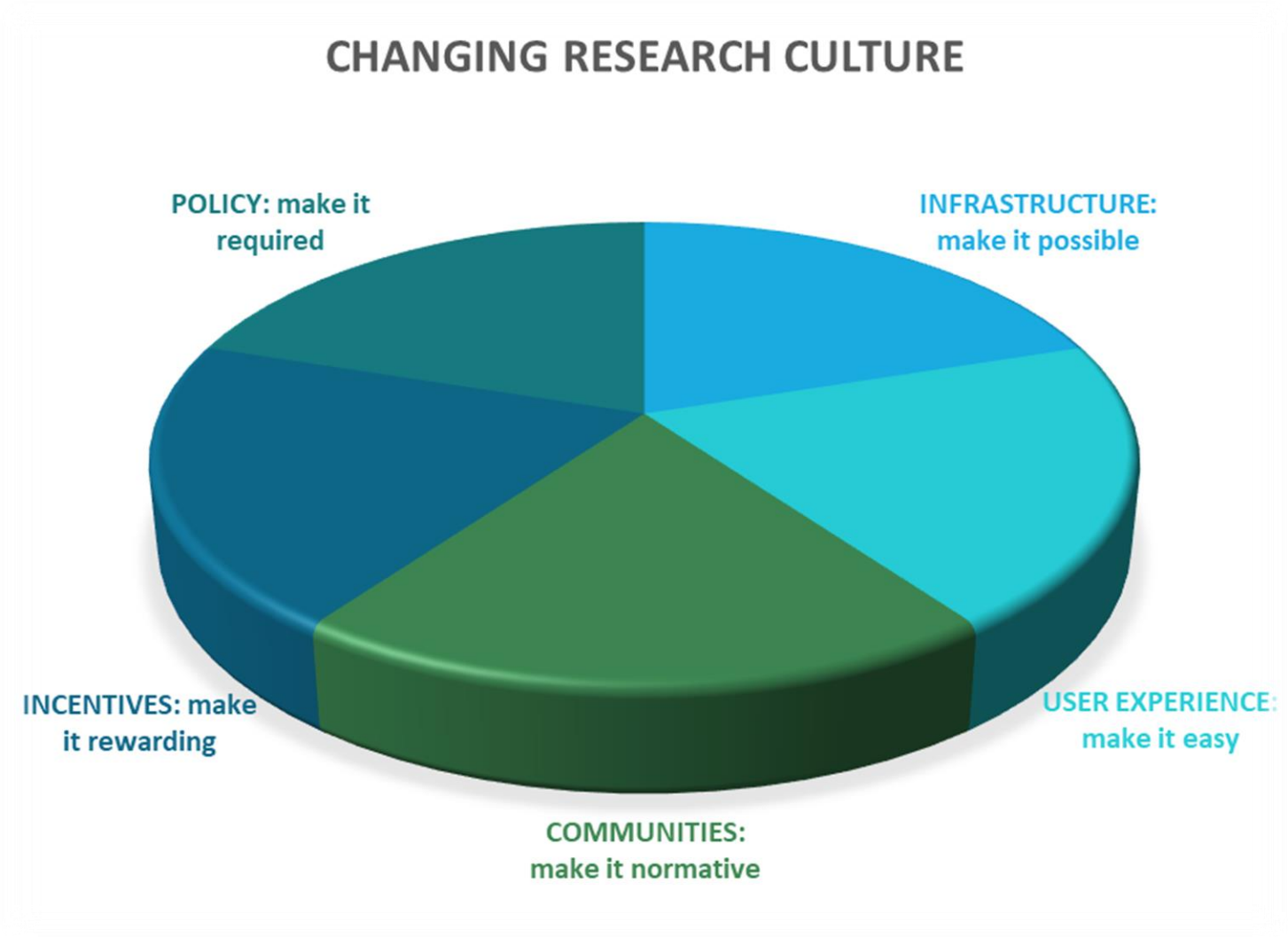
There is another way... The Turing way!



Key aspect for collaboration in neuroscience



It's never THAT easy



Reinterpreted from <https://www.cos.io/blog/cos-celebrates-10-years>



<https://the-turing-way.netlify.app/index.html>

<https://software.ac.uk/>



The Turing Way

Search this book...

Home

- Home for Reproducible Research
- Home for Project Design
- Home for Communication
- Home for Collaboration
- Home for Ethical Research
- Community Handbook
- Word

Welcome

Welcome to *The Turing Way* handbook to reproducible, ethical and collaborative data science.

The Turing Way project is open source, open collaboration, and community-driven. We involve and support a diverse community of contributors to make data science accessible, comprehensible and effective for everyone. Our goal is to provide all the information that researchers and data scientists in academia, industry and the public sector need to ensure that the projects they work on are easy to reproduce and reuse.

Top Tip

The Turing Way is not meant to be read from start to finish. Start with a concept, tool or method that you need now, in your current work. Browse the different guides that make up the book, or use the search box to search for whatever you would like to learn about first.

All stakeholders, including researchers, software engineers, project leaders and funding teams, are encouraged to use *The Turing Way* to understand their roles and responsibility of reproducibility in data science. You can inspect our resources on [GitHub](#), contribute to the project as described in our [contribution guidelines](#) and re-use all materials ([see the License](#)).

Please [join our Slack Workspace](#) to connect and discuss your ideas or suggestions with *The Turing Way* members.



Visit our [GitHub Repository](#)
this book is powered by [Jupyter Book](#)

FOR YOU TO START... plenty of resources/tools out there... HAPPY TO HELP!

<https://openlifesci.org/>



OLS program | OLS-8 | OLS-7 | Policies | About | Services



The OLS program is for people interested in **applying open principles** in their work and **becoming Open Science ambassadors** in their communities.

About

This is a **16-week long personal mentorship and cohort-based training**, where participants (organisers, hosts, mentors and project leads/mentees) of this program will:

- **share** their expertise and gain knowledge essential to create, lead, and sustain an Open Science project
- **connect** with members across different projects, communities, backgrounds, and identities
- **empower** each other to become effective Open Science ambassadors in their communities

Participants join this program with a **project** that they either are already working on or want to develop during this program **individually or in teams**.



Don't be overwhelmed... just choose from the Open science buffet!



Don't try to stuff yourself on everything, select what works for this study and let's steadily improve our fields..

Take your pick from the 'buffet' of open science practices from transparency, statistics, preregistration, multi-lab collaborations...

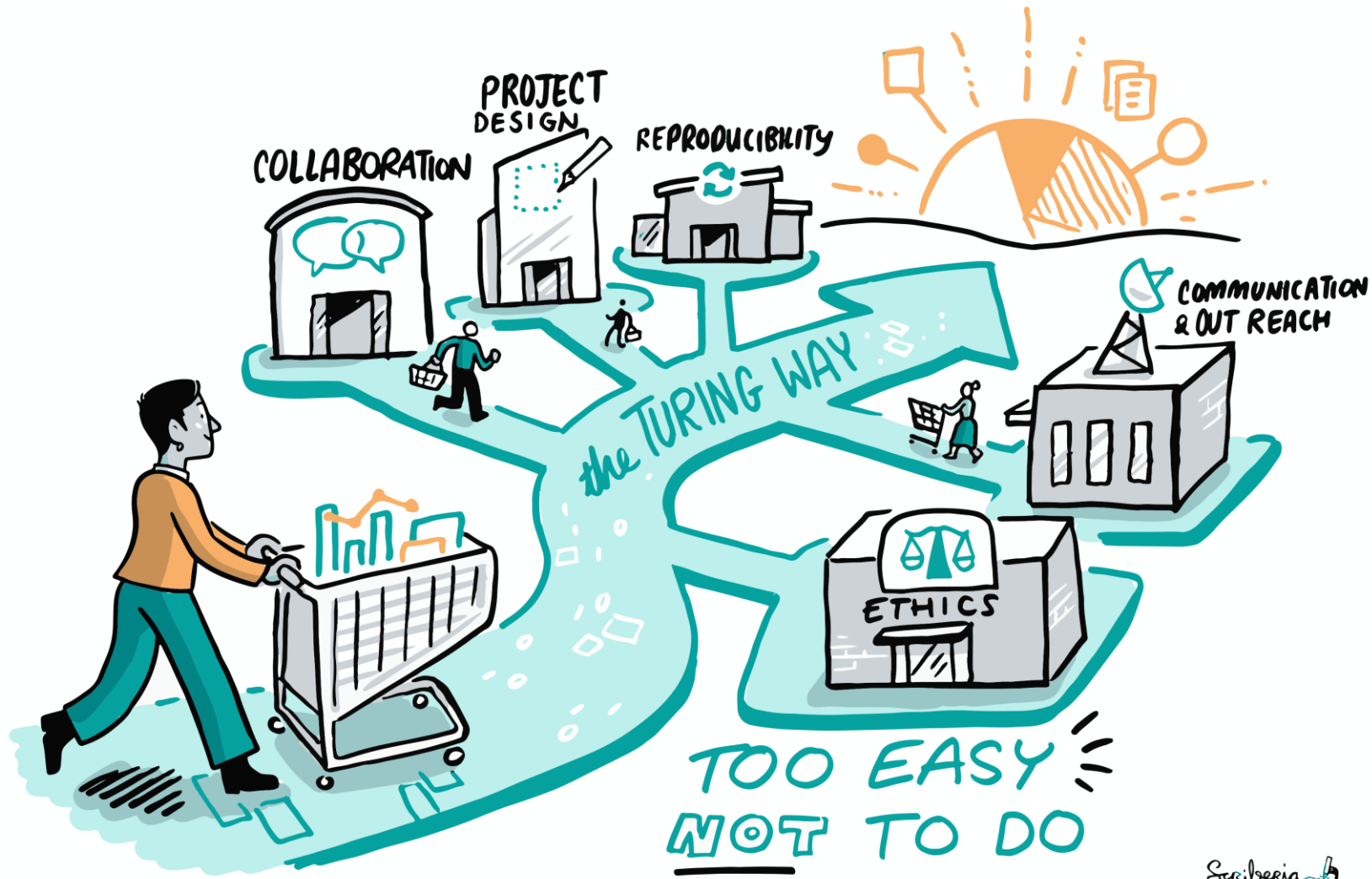
open science buffet" metaphor (h/t Christina Bergmann)



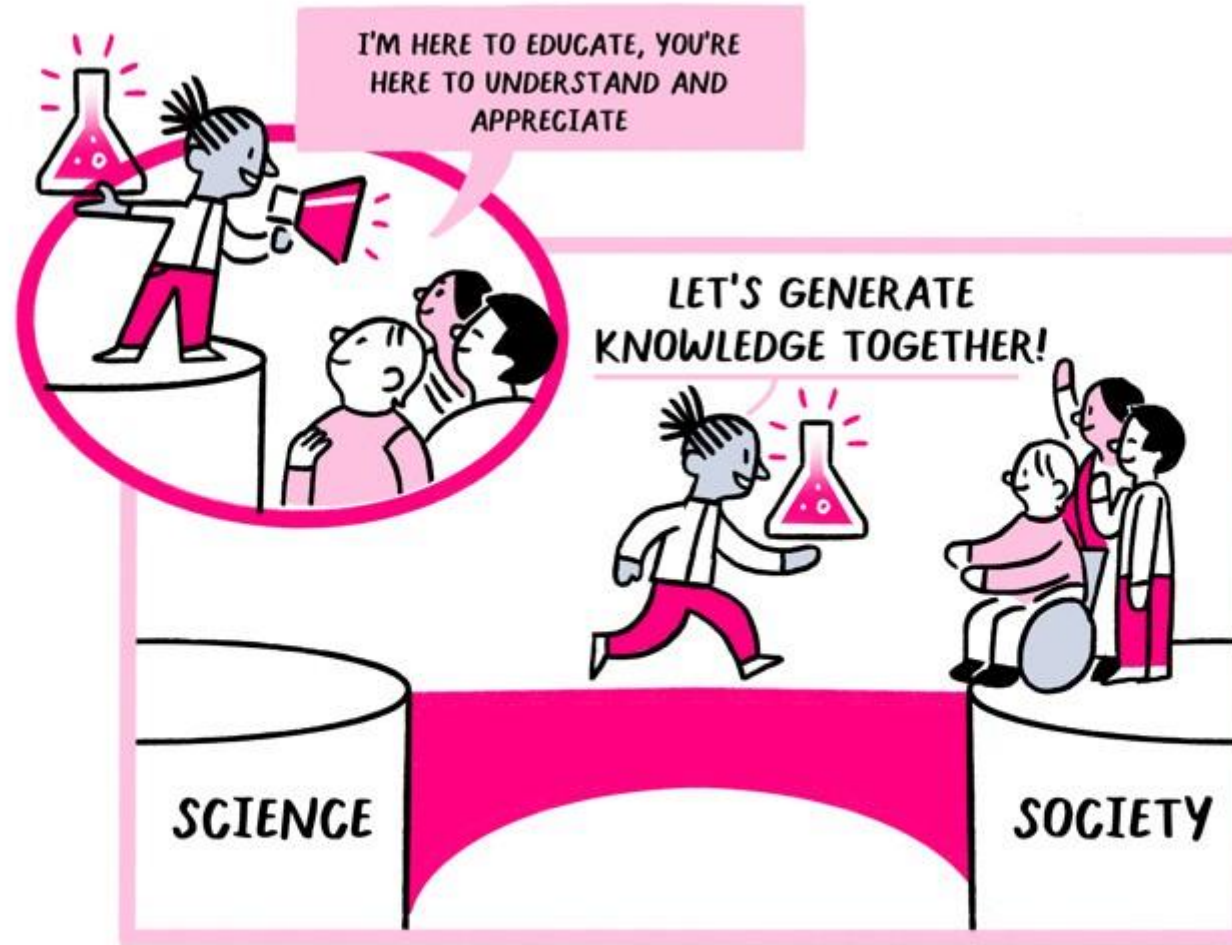
ALL IMAGES TAKEN FROM *The Turing Way* project illustration
by Scriberia. Used under a CC-BY 4.0 licence.

<http://doi.org/10.5281/zenodo.3332807>

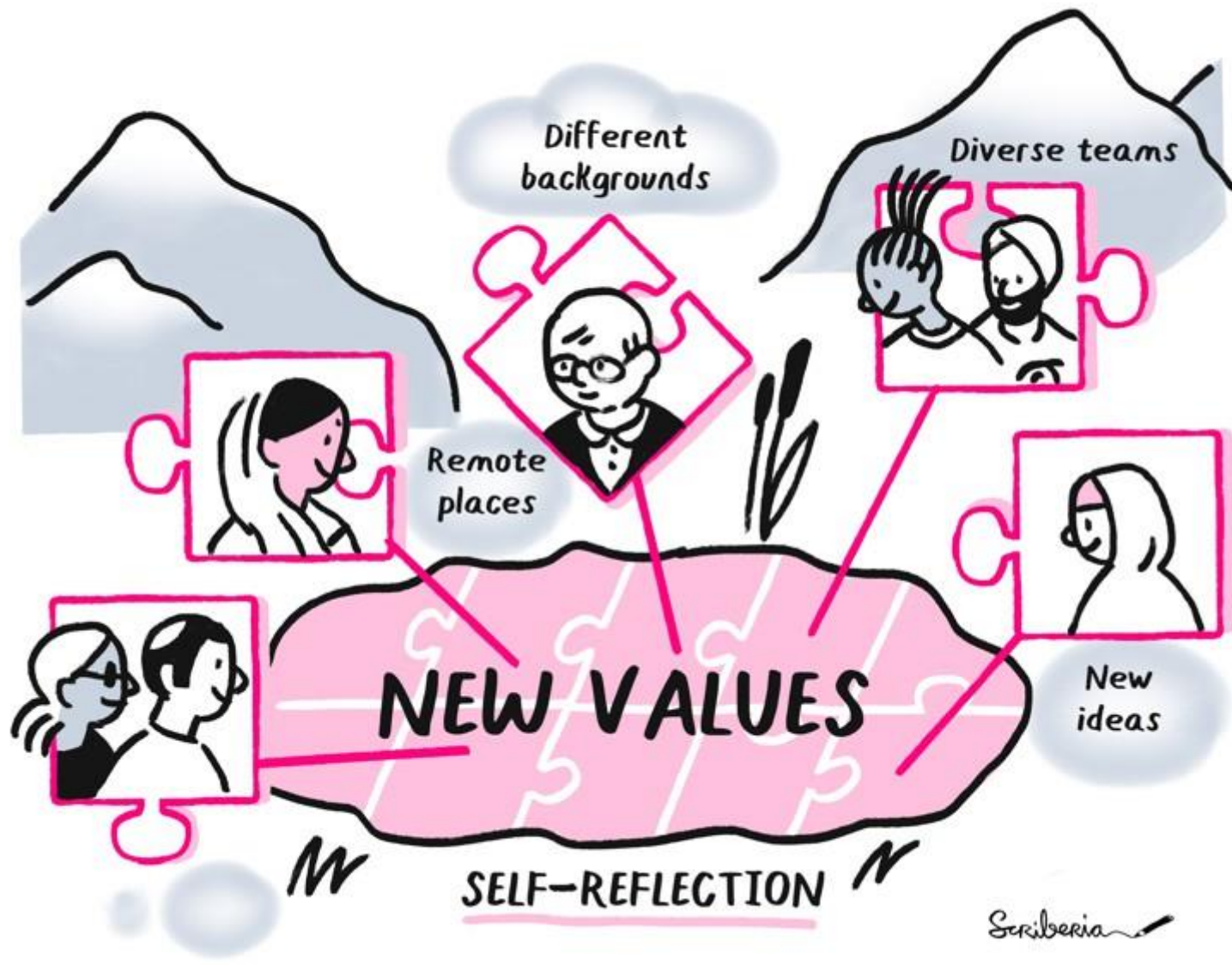
There is another way... The Turing way!



Self-reflection and change of mindset

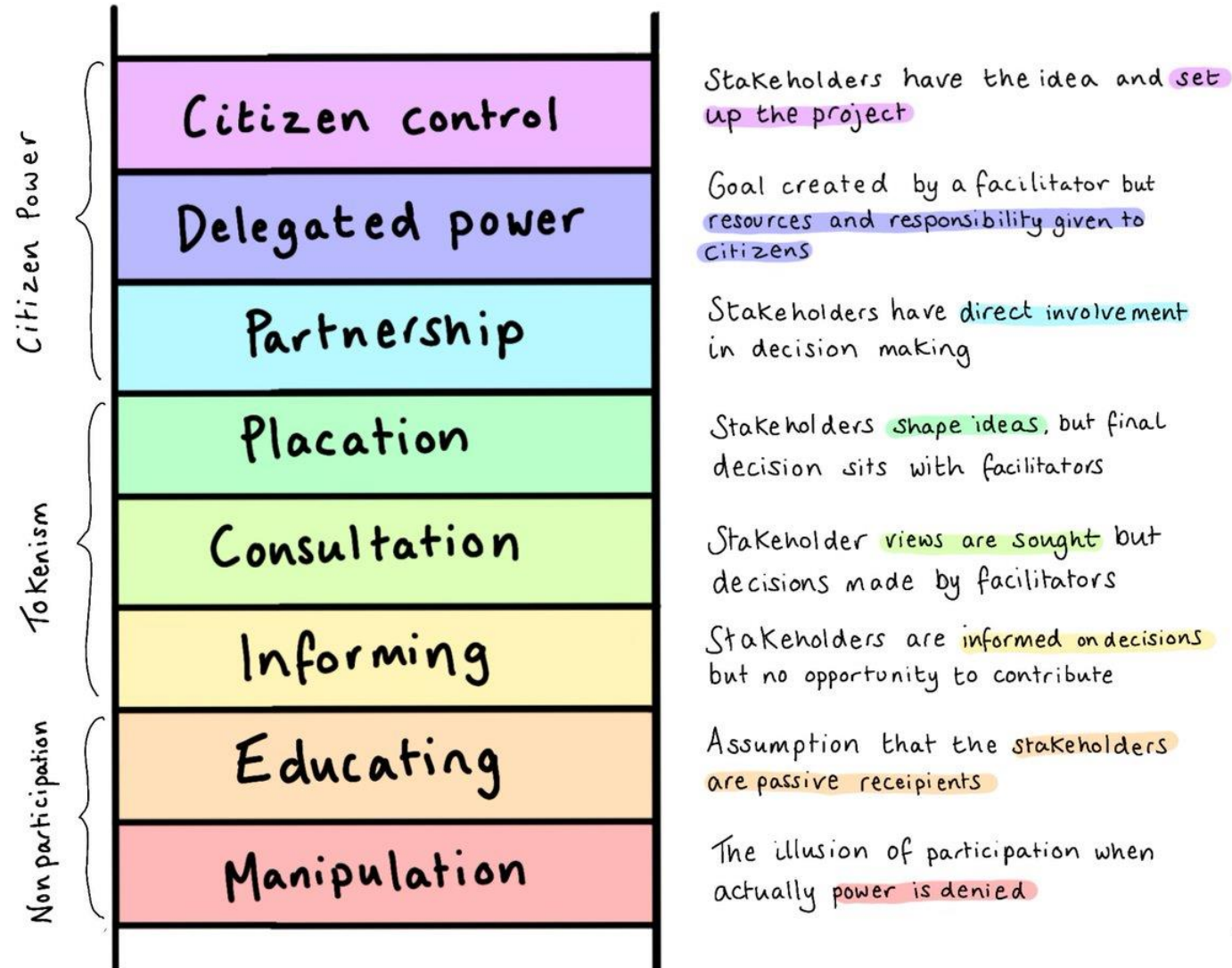


COMMUNICATION - NOT TO SOCIETY SOCIETY BUT WITH IT



Ladder of Participation

(Arnstein, 1969)



@creative.clinical. Psychologist

Drawn by Juliet Young