Research data management (RDM) guidance for i-CONN early stage researchers

15/09/2020

Instructor: Nicholas Syrotiuk

https://doi.org/d8wz

Session plan

F.A.I.R. data principles

15:00 [1]

Working reproducibly

10:00 [2]

Preserving research data

10:00 [3]

Summary

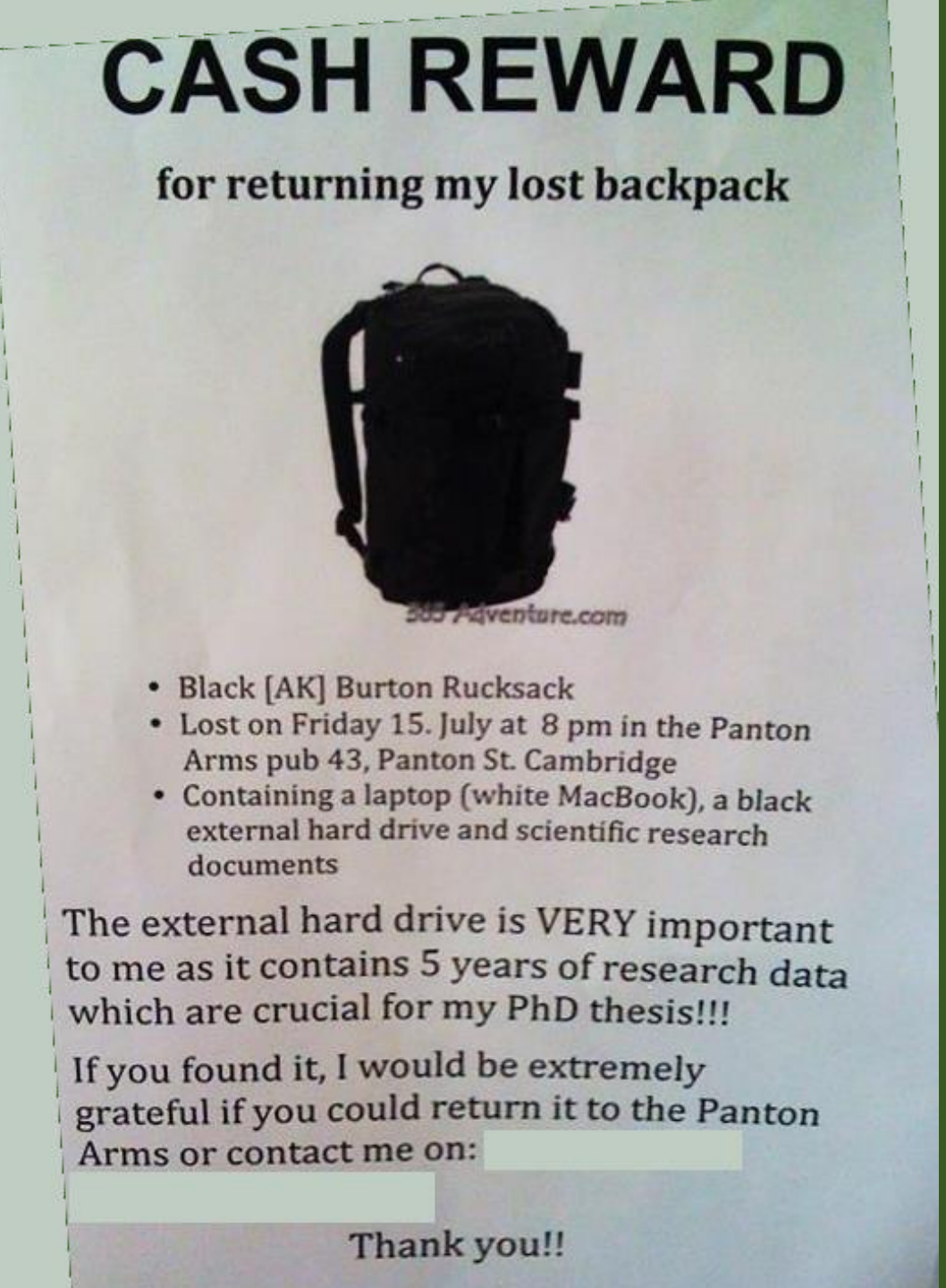
10:00 [4]

What is research data?

One definition: Anything which can be used to validate or replicate a research conclusion, or enrich understanding of the research process.

Why bother with data management?

2005/10/30:
Fire destroys Southampton research centre



F I V E years of research data lost

"Availability of research data declines rapidly with article age"--Vines et al.

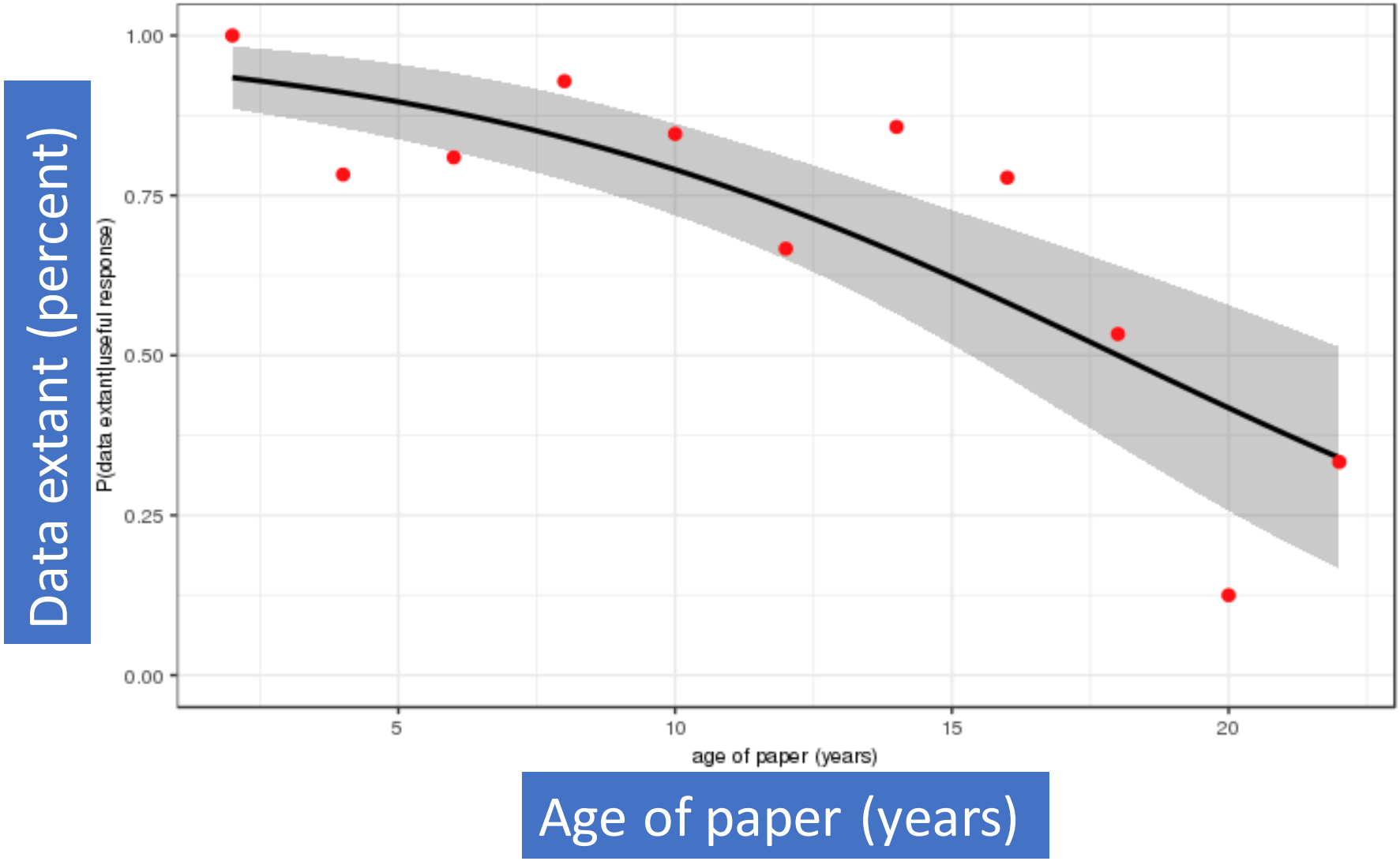


Image created by N Syrotiuk from the orignal research data and R script.

Part one:

WHAT ARE THE F.A.I.R. DATA PRINCIPLES?

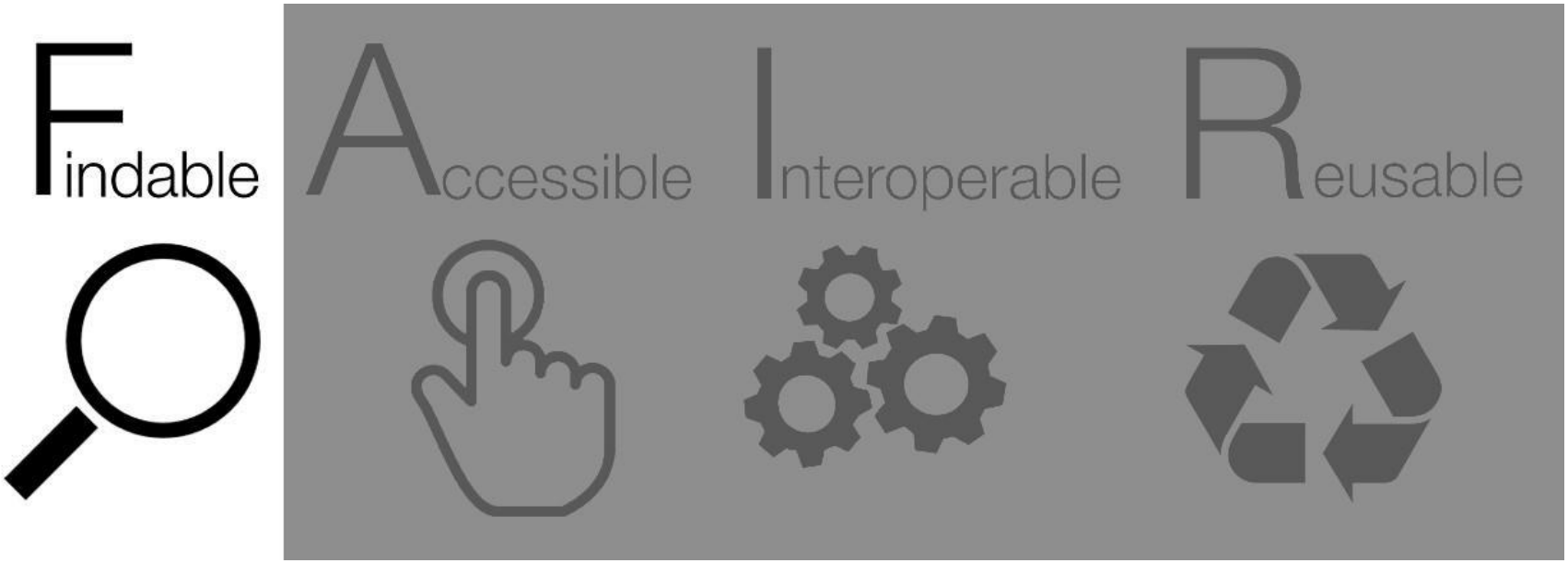
Brief history

2014: Term coined

2016: Fifteen principles published in Nature:

DOI: http://doi.org/10.1038/sdata.2016.18

2020: Guidance on web site: https://www.go-fair.org/fair-principles/


Findable
Accessible
Interoperable
Reusable

Creator(s)

Title

Publisher

Publication year

Subject keywords

Assign a persistent identifier

DOI: Digital object identifer

ARK: Archival resource key

URN: Uniform resource name

PURL: Persistent uniform resource locator

Minting process:

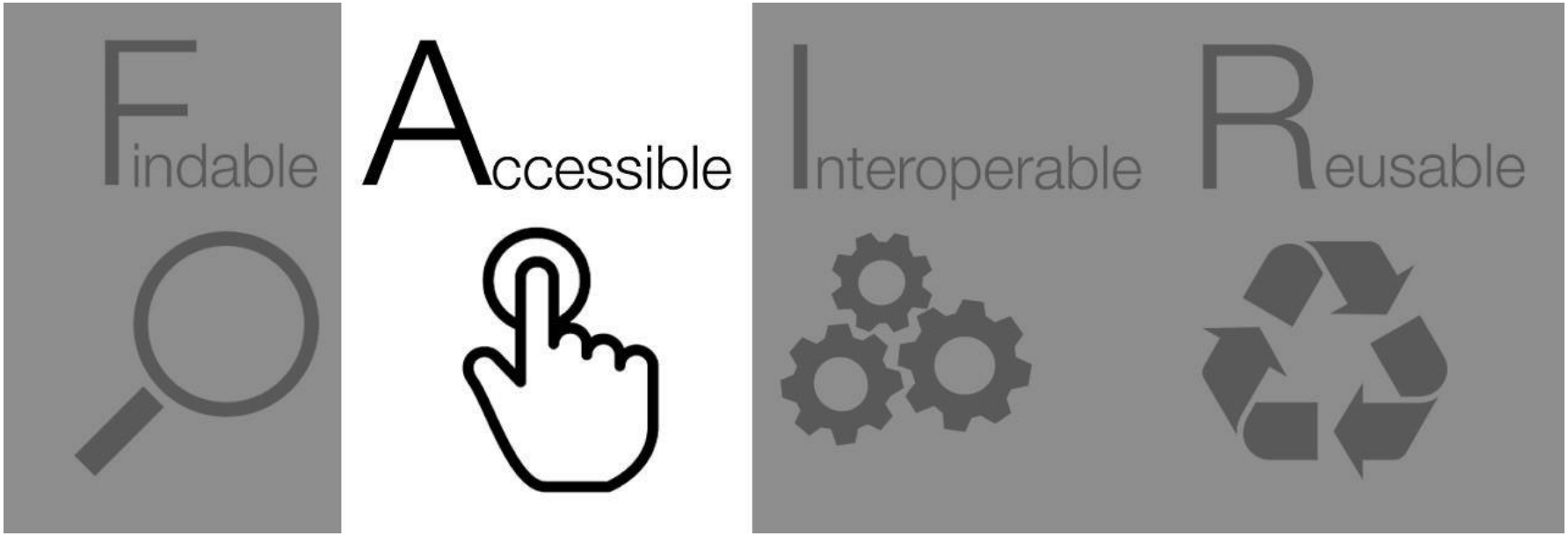
1. Deposit data

2. Add bibliographic metadata

3. Mint/publish DOI

[4. Cite the data]

https://search.datacite.org/


Findable
Accessible
Interoperable
Reusable

As open as possible,
as closed as necessary

Funders expect research data to be published openly in order to facilitate reproducible and transparent research

Exceptions to sharing

Ethical reasons

Public safety reasons

Commercial reasons

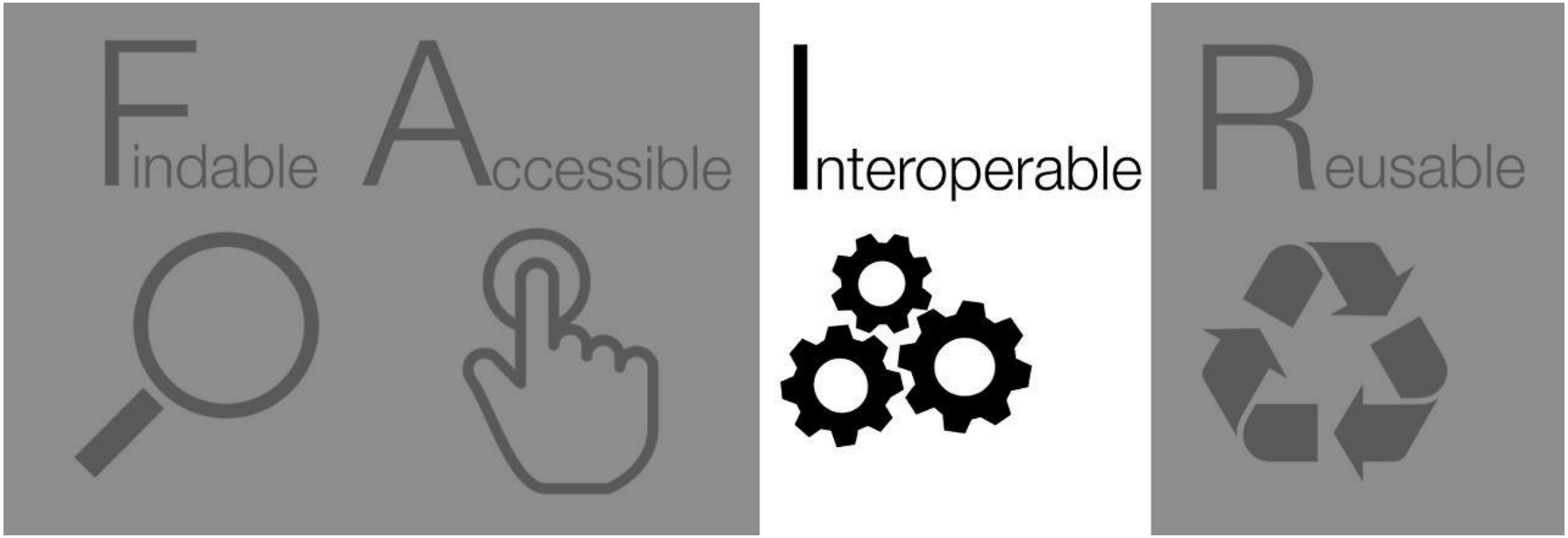
Data access statement: Is your research data published openly?

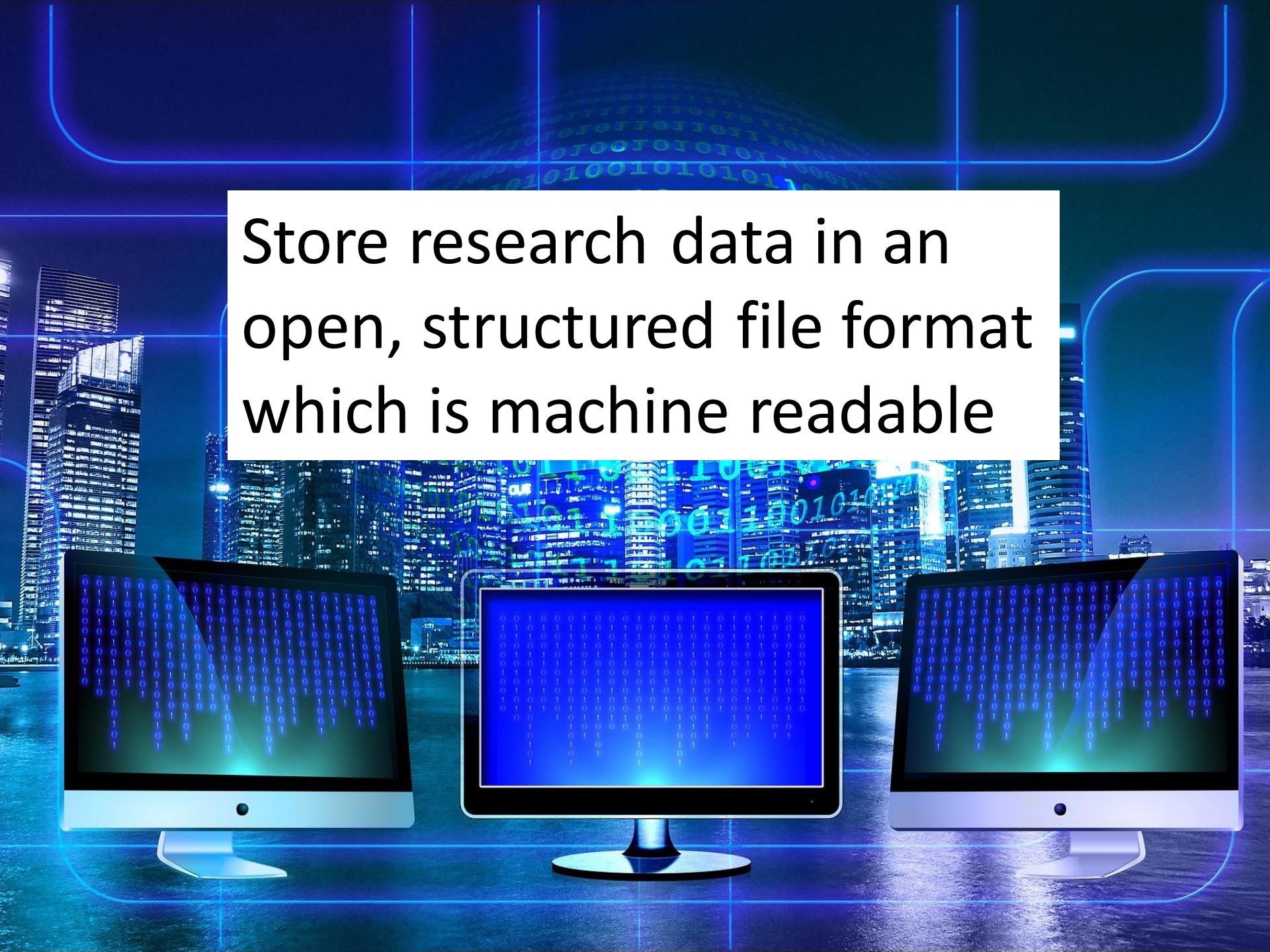
1. Yes. Here's the DOI.

2. Yes, with restrictions as described in the Non-disclosure Agreement.

3. No, the data is too sensitive.

4. No new data was generated.

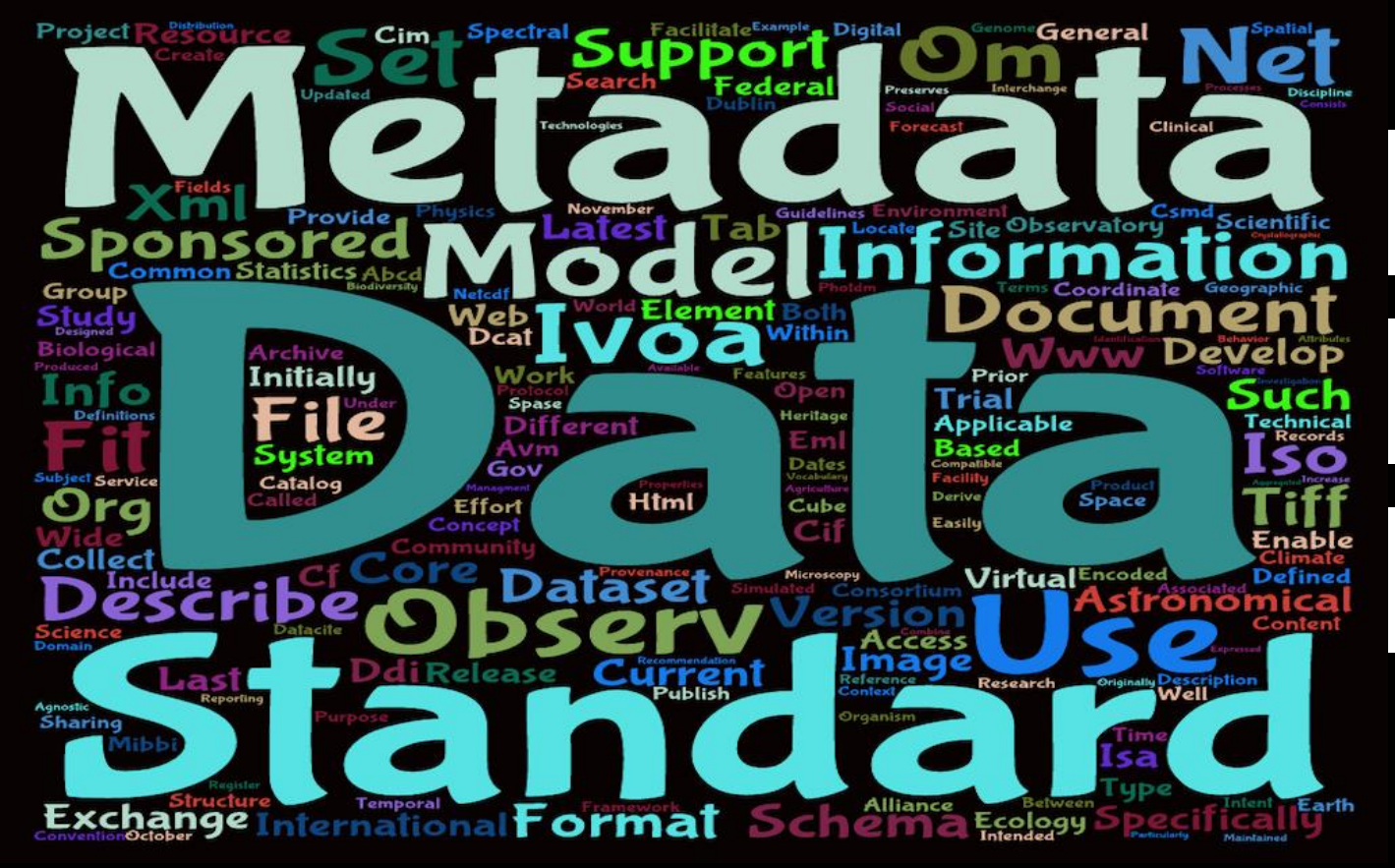

Findable
Accessible
Interoperable
Reusable


Store research data in an open, structured file format which is machine readable


Avoid
proprietary
file formats

Describe data elements of structured research data using a standard vocabulary which includes resolvable, global identifiers linking to explanations

Attach metadata as soon as possible after data capture



Standards

Tools

Use cases

http://rd-alliance.github.io/metadata-directory

Example of discipline-specific metadata

Standard

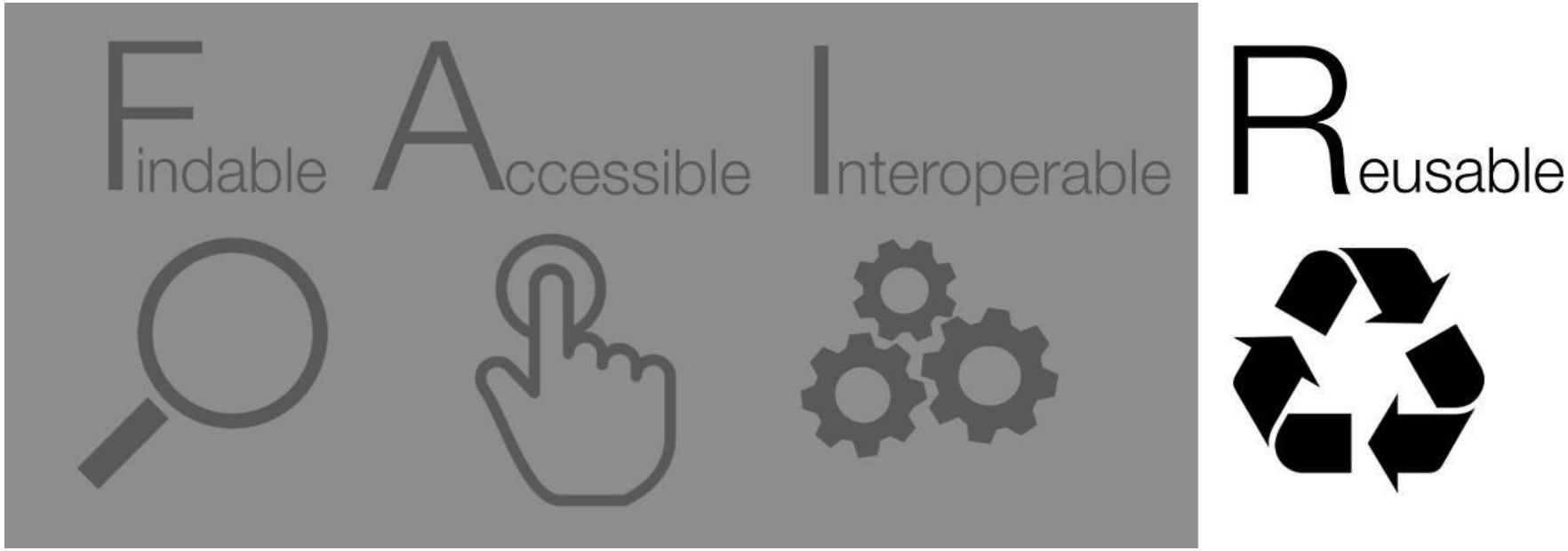
Ecological Metadata Language (EML) for earth, environmental and ecological sciences (in XML format)

Tool

Morpho
https://knb.ecoinformatics.org/tools/morpho

Use case

Knowledge Network for Biocomplexity (KNB) data repository

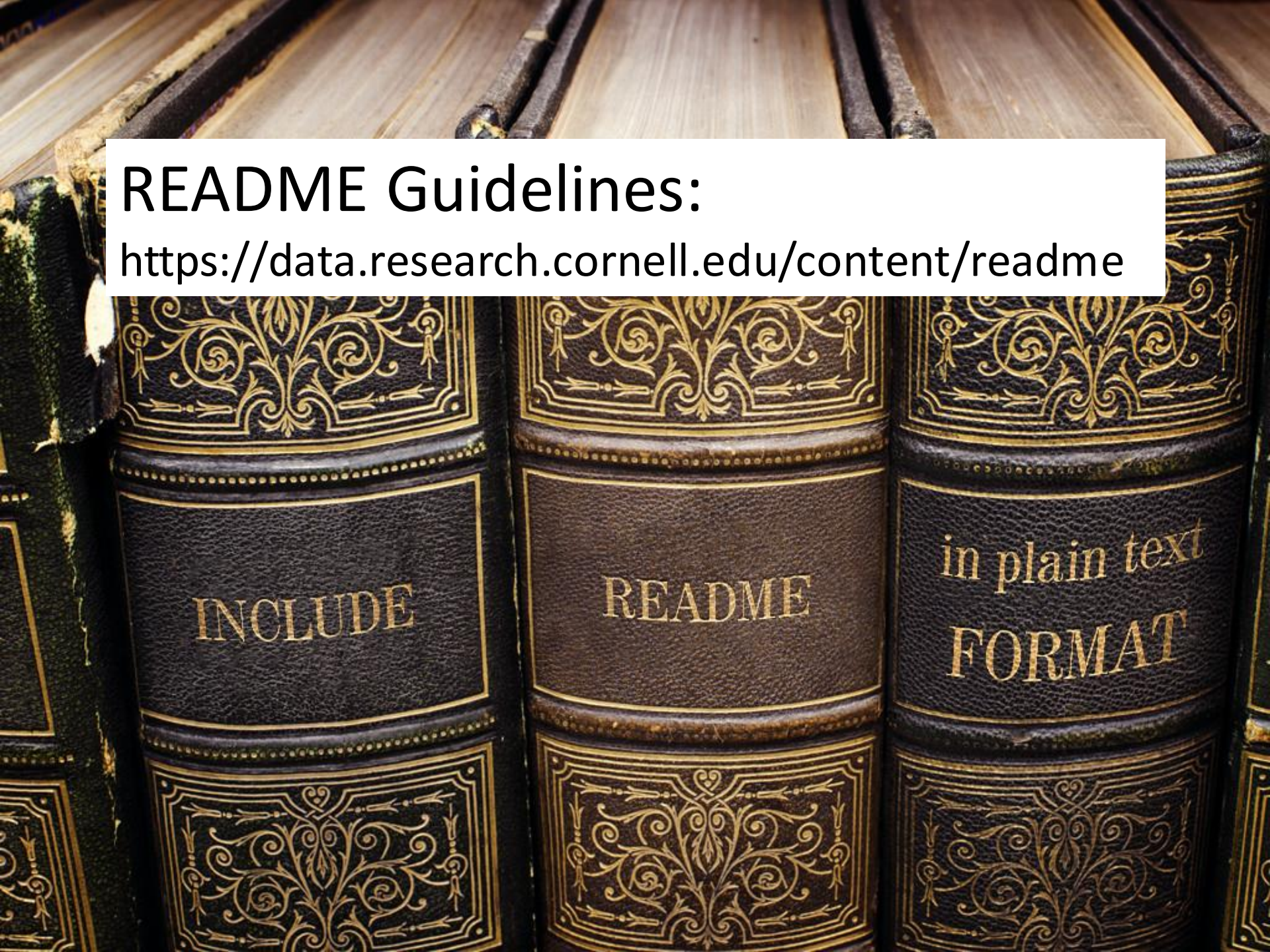

Findable
Accessible
Interoperable
Reusable

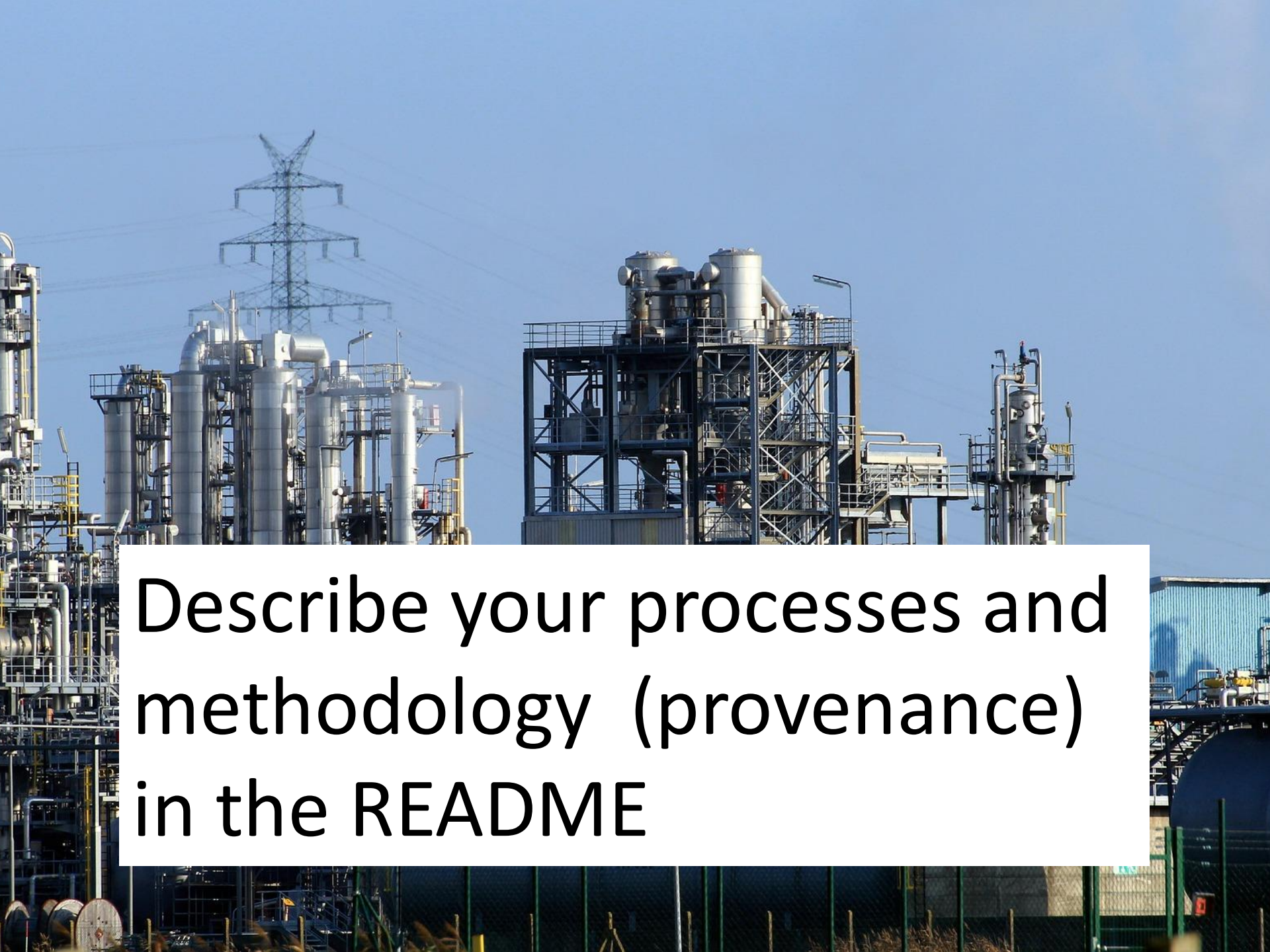
Documentation

Good data comes with good documentation

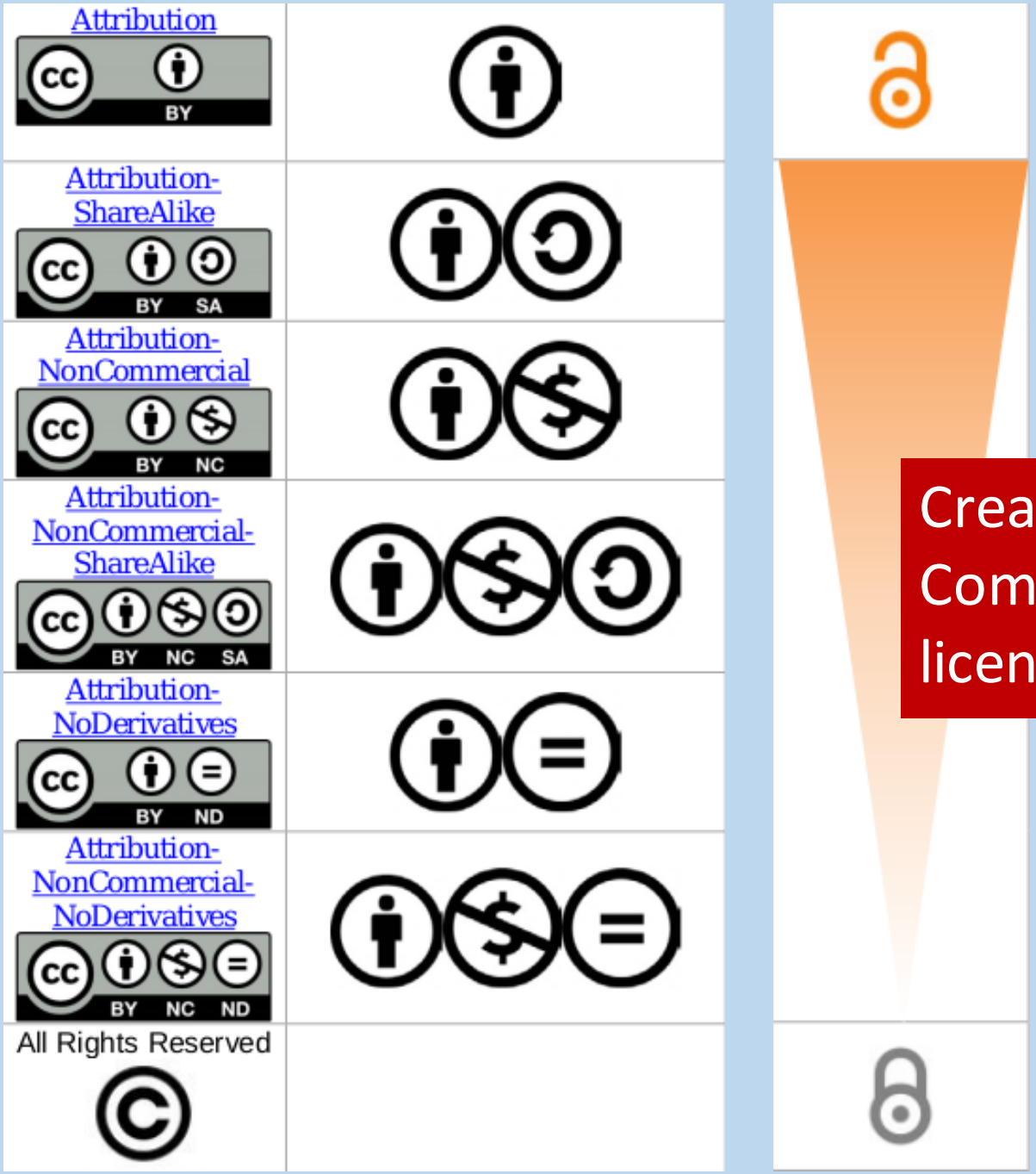
README Guidelines:

https://data.research.cornell.edu/content/readme





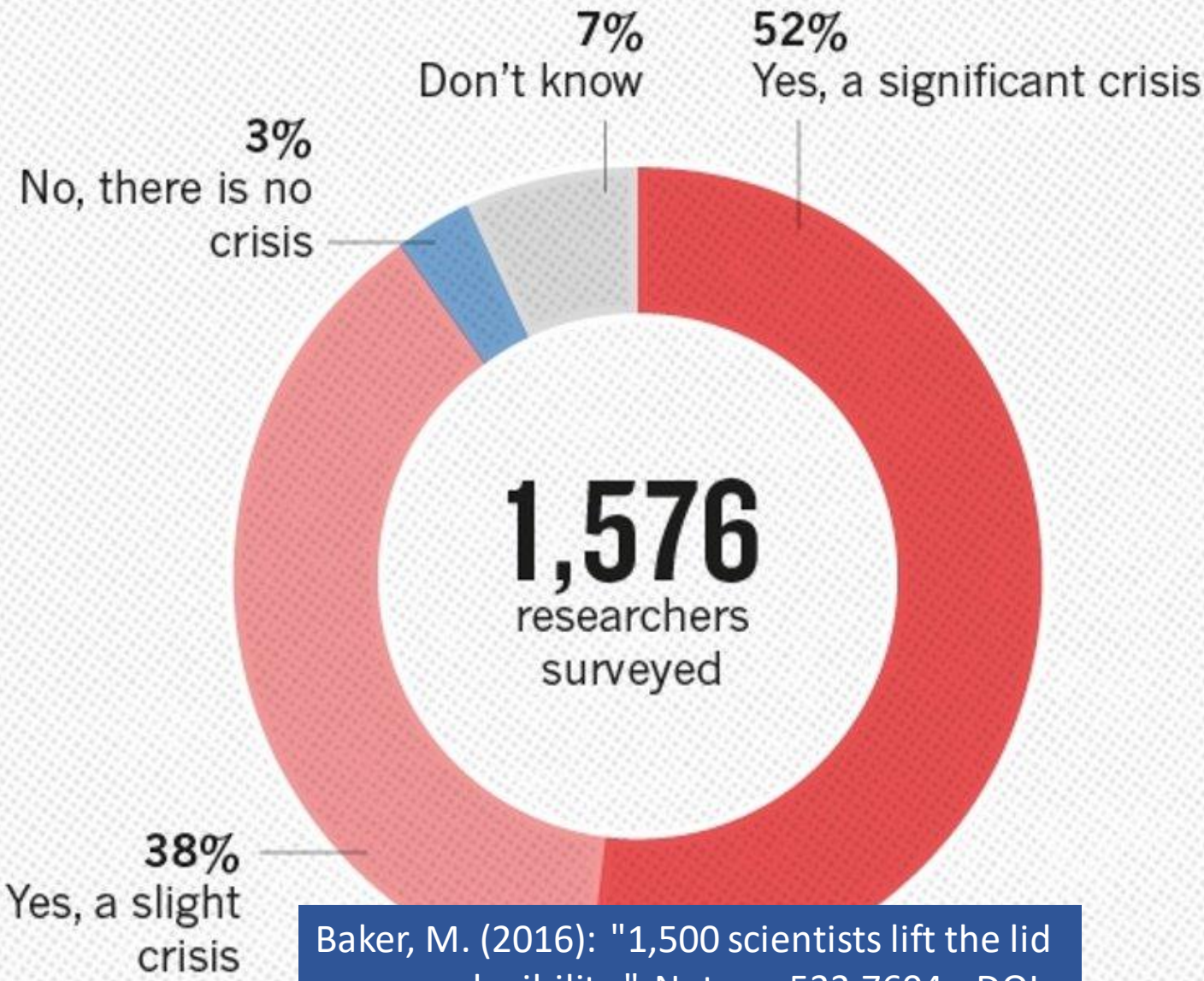
Describe your processes and methodology (provenance) in the README


Attribution
BY
Attribution-ShareAlike
BY SA
Attribution-NonCommercial
BY NC
Attribution-NonCommercial-ShareAlike
BY NC SA
Attribution-NoDerivatives
BY ND
Attribution-NonCommercial-NoDerivatives
BY NC ND
All Rights Reserved
Open
Creative Commons licences
Closed

Part two:

WORKING REPRODUCIBLY

IS THERE A REPRODUCIBILITY CRISIS?



Baker, M. (2016): "1,500 scientists lift the lid on reproducibility." Nature, 533:7604. DOI: http://doi.org/10.1038/533452a

Levels of reproducibility

According to Florian Markowetz (2015)

Avoid beginners' mistakes

Keep files organised

Name files in a meaningful way

Avoid scattering files

Lowest level of reproducibility

Computational reproducibility

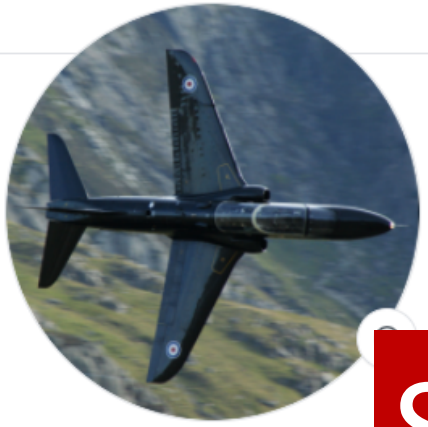
Scripting tools:
Python, R, Perl, ...

Notebook tools:
IPython, Jupyter, ...

Next level of reproducibility

Search or jump to... / Pull requests Issues Marketplace Explore

Overview Repositories 22 Projects Packages



Nicholas Syrotiuk
sefnyn

Edit profile

4 followers · 6 following · 24

Highlights

Arctic Code Vault Contributor

Organizations

Pinned

Customize your pins

scholix

Scripts to find links between research papers and corresponding research data

Python 6 2

JavaScript

carousel

RDM slides published as a simple image carousel

JavaScript

rom IMF International Financial

nd Postgres backend. Persistent

Ruby

Software version control systems

e.g., Git, Mercurial, Subversion

144 contributions in the last year

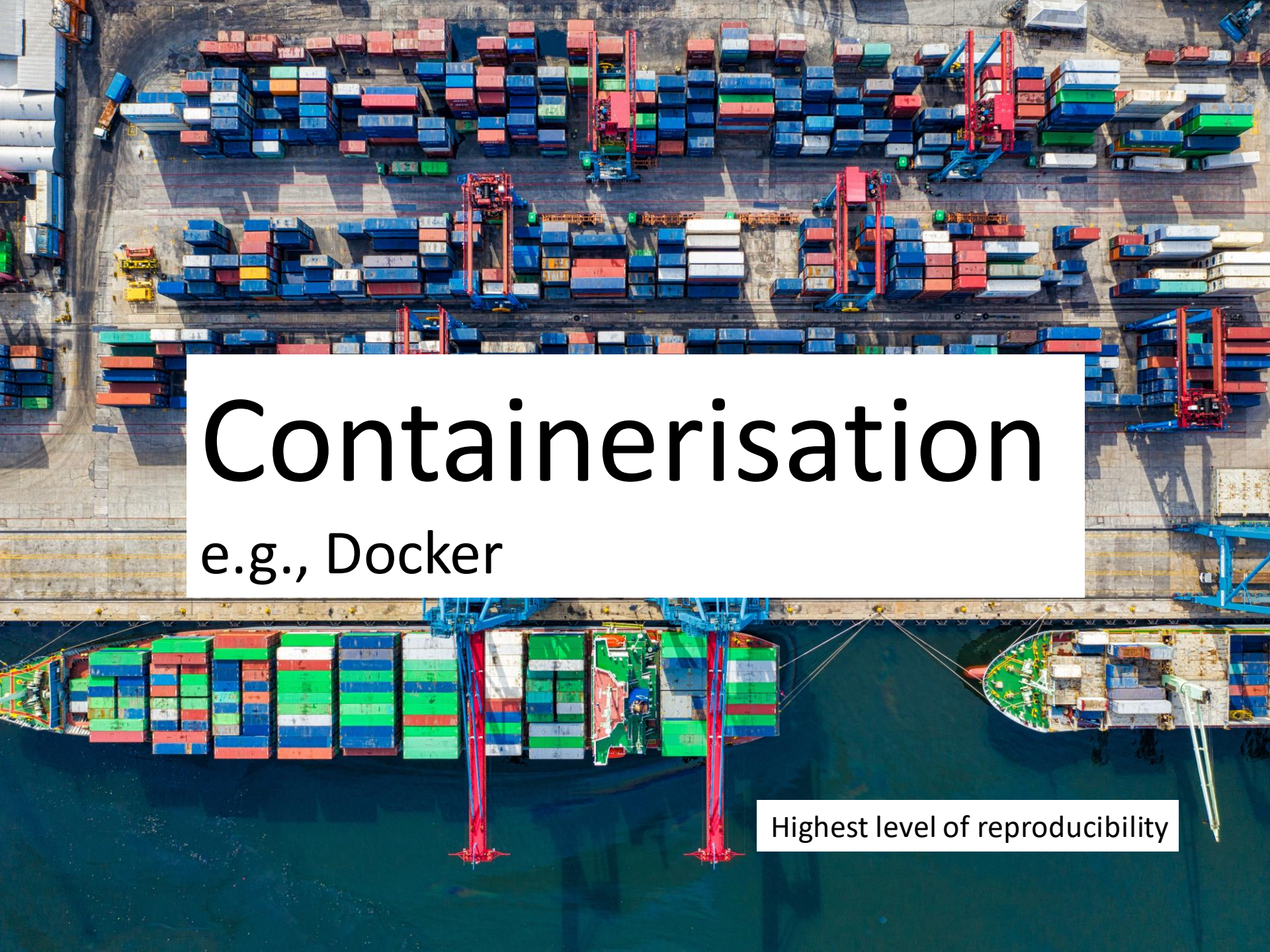
Contribution settings

Sep Oct Nov Dec Jan Feb Mar Apr May Jun Jul Aug

Learn how we count contributions.

Less More

Next level of reproducibility



Containerisation

e.g., Docker

Highest level of reproducibility

Part three:

PRESERVING RESEARCH DATA FOR THE LONG TERM

Preparation and organisation

- Provide all necessary files
- Ensure accessibility of files
- Descriptive file names
- Logical file structure
- Include a README

Anonymisation

https://amnesia.openaire.eu/

Choosing a research data repository

Three types of repository

1 Subject-specific

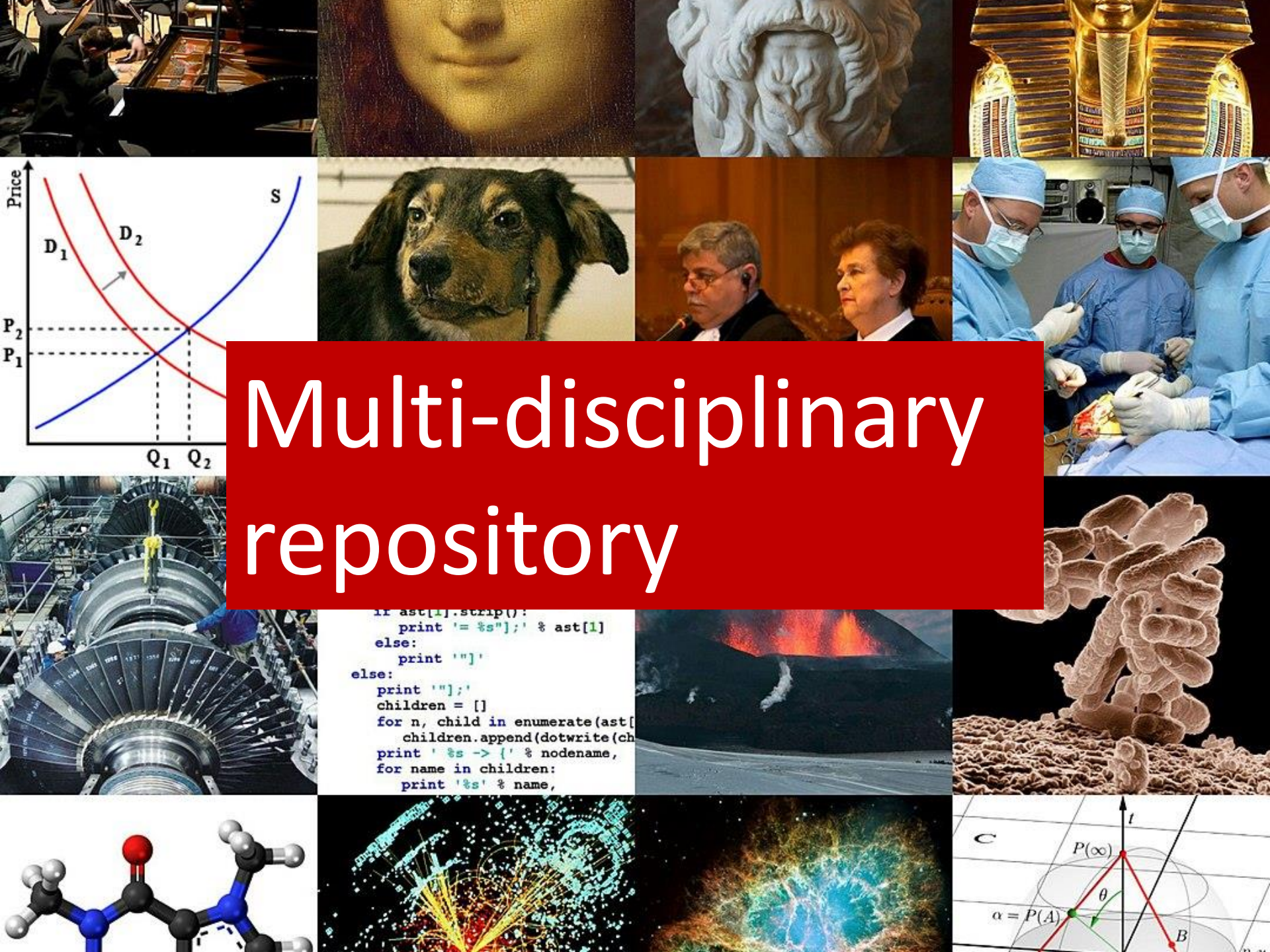
2 Multi-disciplinary

3 Institutional

Find a subject-specific repository

re3data.org

REGISTRY OF RESEARCH DATA REPOSITORIES



Multi-disciplinary repository

Mendeley

zenodo

DRYAD

OSF

Open Science Framework

figshare

Examples of multi-disciplinary repositories

Collections

https://collections.durham.ac.uk/

Durham University

Enter search terms | Go | All

About | Help | Contact | Login

Institutional repo: Durham research data repository

Share Your Work

Terms of Use

Recently Uploaded

Depositor	File Details
N.G. Chancellor	Finding spin-glass ground... qwspinglass_data.tar.gz quantum computing, quantum walk, quantum optimisation, quantum algorithms, spin glass

Tweets by @DurhamRdm

Durham RDM Retweeted

DurhamResearchOnline
@DROdurham

Accepted manuscript now available on DRO: Kilby, Karen E. (2018) 'Julian of Norwich, Hans Urs von Balthasar, and the status of suffering in Christian theology.', New blackfriars., 99 (1081). pp. 298-311. dro.dur.ac.uk/22207/

5h

Add Data Search About User Guide Support Log In

DataverseNL

Metrics 48,176 Downloads

Another institutional repository

Welcome to DataverseNL

Store, share and publish research data online. Use the slider below to access the dataverses of the DataverseNL partners. If you want to try out the DataverseNL features, please visit our demo-site.

Utrecht University

Maastricht University

Tilburg University

University of Groningen

Search this dataverse... Find Advanced Search

- [x] Dataverses (370)
- [x] Datasets (1,442)
- [ ] Files (9,522)

Dataverse Category

Organization or Institution (111)
Research Project (61)
Research Group (59)
Department (35)

1 to 10 of 1,812 Results Sort

The Attentional Blink is Related to the Microsaccade Rate Signature

Sep 1, 2020 - Cognitive Neuroscience

Roberts, Mark J.; Lange, Gesa; Van Der Veen, Tracey; Lowet, Eric; de Weerd, Peter, 2020, "The Attentional Blink is Related to the Microsaccade Rate Signature", https://doi.org/10.34894/5YNXRI, DataverseNL, V1

The reduced detectability of a target T2 following discrimination of a preceding target T1 in the attentional blink (AB) paradigm is classically interpreted as a consequence of reduced attention to T2 due to attentional allocation to T1. Here, we investigated whether AB was relat...

Data citation

Format:
Creator (PublicationYear): Title. Version. Publisher. (resourceTypeGeneral). Persistent identifer

Example:
Breckon, T; Tiancheng, G (2018): Pretrained neural network models for Guo 2018 study, TensorFlow format. Durham University. (dataset). DOI: http://doi.org/10.15128/r23j333226h

Source: Metadata Schema Documentation for the Publication and Citation of Research Data, version 4.3. DOI: https://doi.org/10.14454/7xq3-zf69

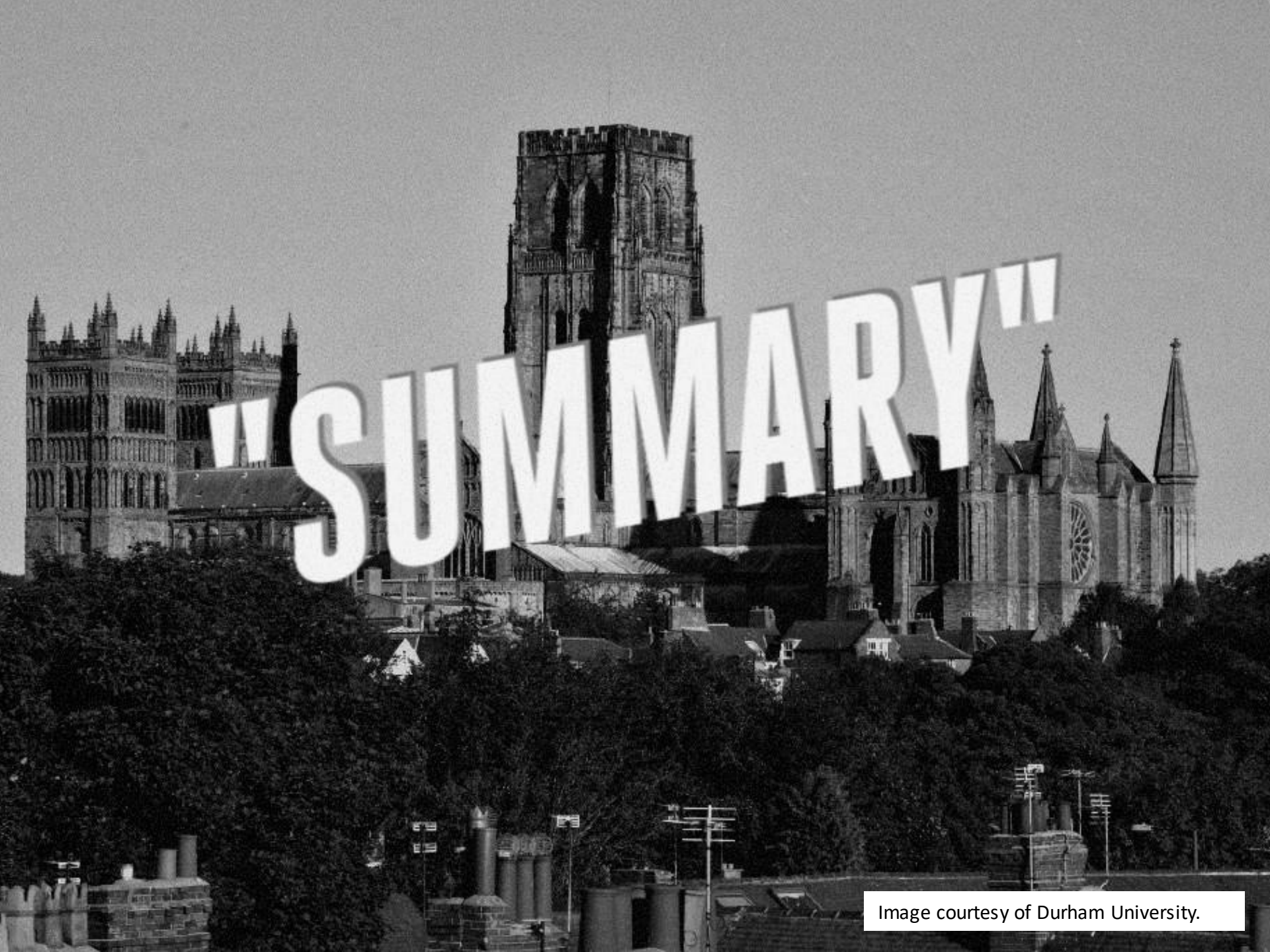


Image courtesy of Durham University.

Summary (1): Basic do's and don'ts of data management

DO	DON'T
Have a plan for managing research data	Make it up as you go along
Keep backups. Make this easy with automated syncing services like Dropbox, provided your data isn't too sensitive	Carry the only copy around on a memory card, your laptop, your phone, etc
Describe your data as you collect it. This makes it possible for others to interpret it, and for you to do the same a few years down the line	Leave this till the end. The quality of metadata decreases with time, and the best metadata is created at the moment of data capture
Save your work in open file formats, where possible, and use accepted metadata standards to enable like-with-like comparison	Invent new 'standards' where community norms already exist
Deposit your data in a data centre or repository, and link it to your publications	Be afraid to ask for help. This will exist both within institutions, and via national / European support organisations

Summary (2): Data management rules of thumb

- Without intervention, data + time = no data
    - *See:* Vines, T. H., et al. (2014): "The availability of research data declines rapidly with article age," Current Biology 24(1): 94-97. DOI: http://doi.org/10.1016/j.cub.2013.11.014
- Following F.A.I.R. data principles and sharing research data can lead to making more progress as a research community collectively
- Working reproducibly and writing good RDM documentation ultimately saves time
- Publish data in a repository in order to preserve it for the long term
    - Not all data should be published or shared
    - Publish in one place only

Thank you

Nicholas Syrotiuk

@DurhamRdm

http://doi.org/d8wz

References:

Vines, T. H., et al. (2014): "The availability of research data declines rapidly with article age," Current Biology 24(1): 94-97. DOI: https://doi.org/10.1016/j.cub.2013.11.014

Vines, T. H., et al. (2013): "The availability of research data declines rapidly with article age." Dryad Digital Repository. (dataset). DOI: https://doi.org/10.5061/dryad.q3g37

UK Research and Innovation (2018): "Guidance on best practice in the management of research data." UKRI web site.

Markowetz, F. (2015): "Five selfish reasons to work reproducibly." Genome Biol 16, 274. DOI: https://doi.org/10.1186/s13059-015-0850-7

Piccolo, S.R. and Frampton, M.B. (2016): "Tools and techniques for computational reproducibility." GigaScience, Volume 5, Issue 1, s13742–016–0135–4. DOI: https://doi.org/10.1186/s13742-016-0135-4

Credits:

F.A.I.R. data image:
Licenced by Sangya Pundir under CC-BY-SA 4.0.
https://commons.wikimedia.org/wiki/File:FAIR_data_principles.jpg

CC licence image:
Licenced by Creative Commons Australia under CC BY 4.0.
http://creativecommons.org.au/know-your-rights/

Image of academic disciplines:
Licenced by the Collective under CC BY-SA 3.0.
https://commons.wikimedia.org/w/index.php?curid=62858409