



# FAIR principles in life science research practice

**SBDI Days**

**2024-01-24**

**Wolmar Nyberg Åkerström**  
**National Bioinformatics**  
**Infrastructure Sweden**  
**(NBIS)**

**Angela Fuentes Pardo**  
**SciLifeLab Data Centre**

**Arnold Kochari**  
**SciLifeLab Data Centre**

[data-management@scilifelab.se](mailto:data-management@scilifelab.se)

<https://doi.org/10.17044/scilifelab.25091471>

# What is SciLifeLab Data Management?



- Collaborative activity between **SciLifeLab Data Centre** and the **Data Management team** at **NBIS**.
- About 17 members (project leaders and data stewards) working in different aspects of **Research Data Management (RDM)**
- Promote **Open Science**, **FAIR**, and **good RDM** practices
- Provide services and resources for **data management**, **IT** and **data sharing**
- Make **RDM** and **bioinformatics** support and training **easily accessible**



# Agenda

---

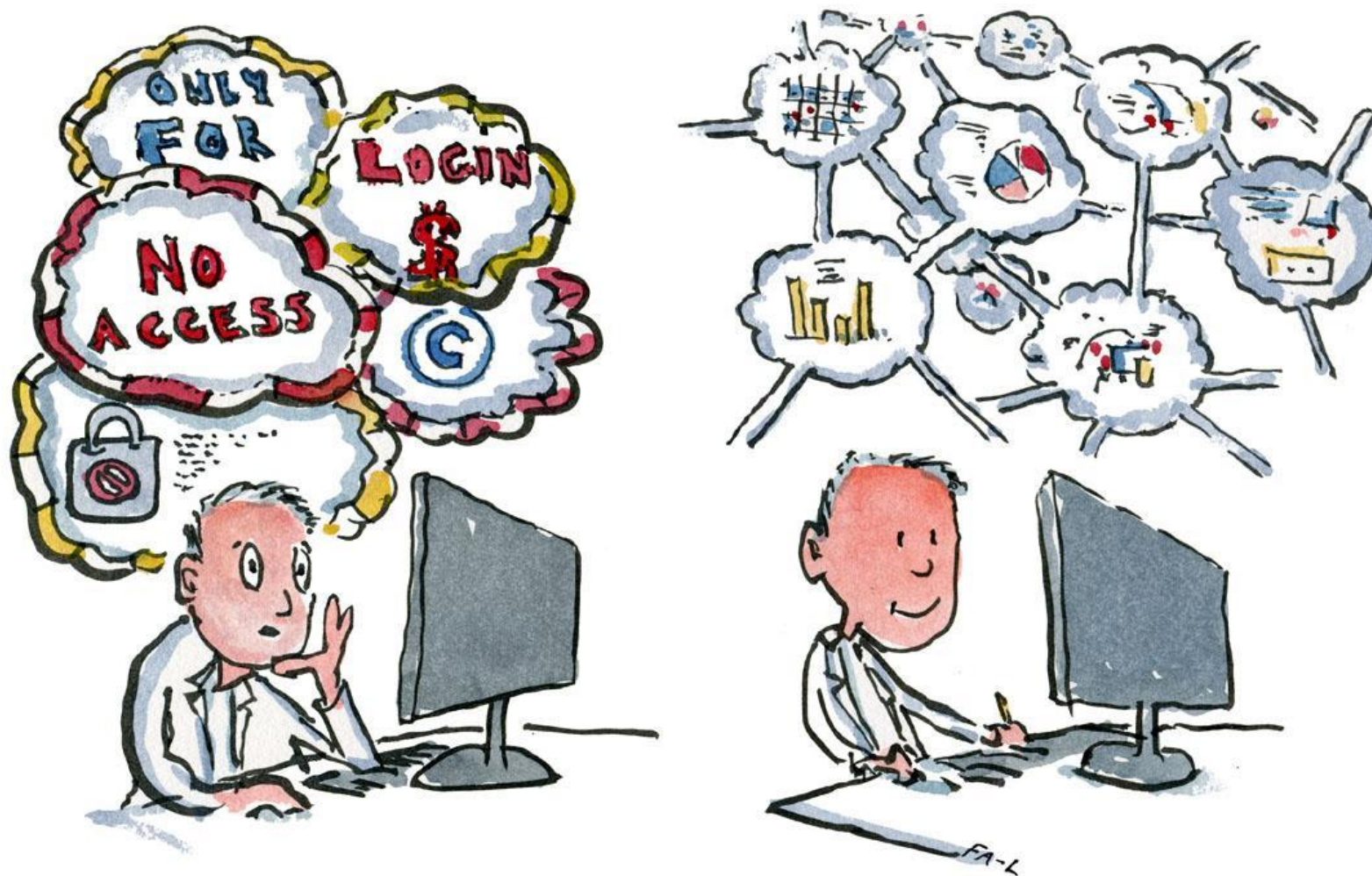


- Open Science and FAIR principles
- Effective data management practices
- Sharing code, tools, machine learning models
- Q&A



# Open Science and FAIR principles

By Angela



Frits Ahlefeldt

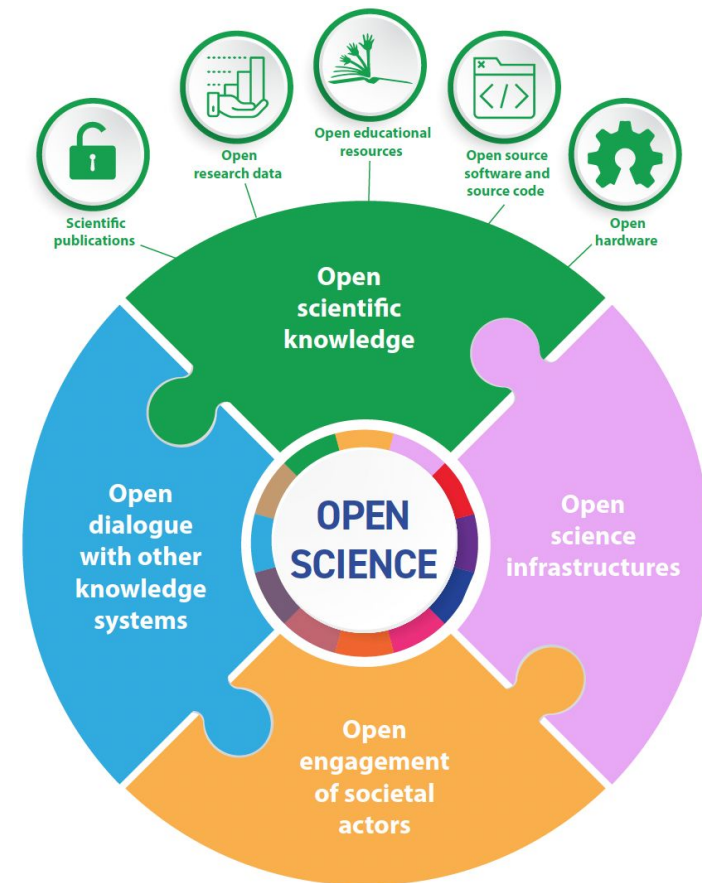
# Open Science



*“Make **scientific research** and its dissemination **accessible to all** levels of society”*

- Open scientific publications
- Open research data
- Open methods
- Open software and code
- Open educational resources
- Open hardware

*“As open as possible, but  
as closed as necessary”*



# Why does Open Science matter?



Compliant  
with grant  
rules



Open  
Science

- Required by funders
  - EU Open Science policy



# Why does Open Science matter?



Compliant  
with grant  
rules

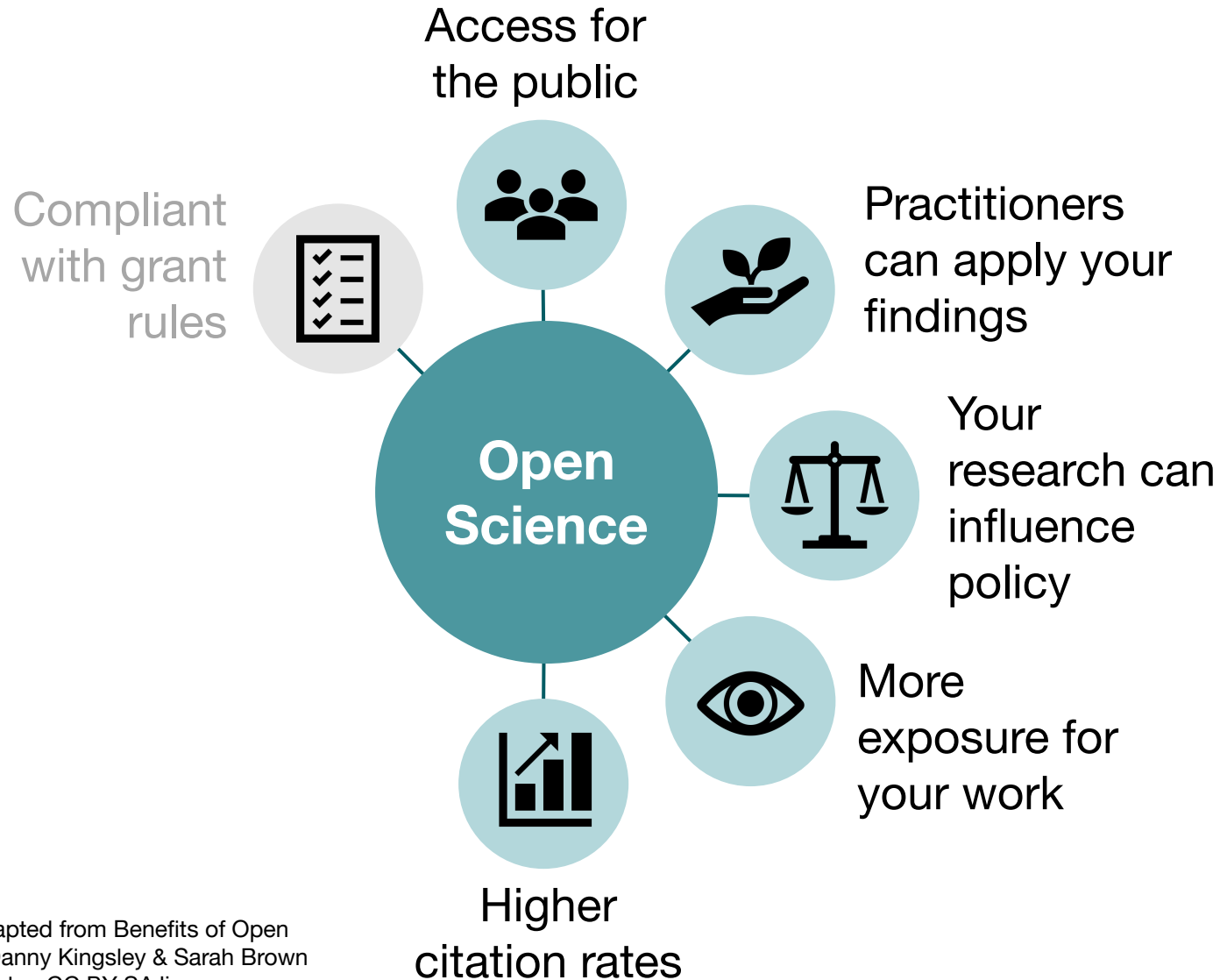


Open  
Science

- Required by funders
  - EU Open Science policy
  - National policies

The screenshot shows the top of a webpage from the Kungliga biblioteket (KB). The header includes the KB logo, navigation links for 'Nytt från KB', 'Kontakt', 'Tjänster', 'Sök', and 'Meny'. Below the header, there is a breadcrumb trail: 'Startsida • Nytt från KB • National guidelines for promoting open science in Sweden'. The main heading is 'National guidelines for promoting open science in Sweden'. Below this, the date '15 januari 2024' and a link 'Open Science (In English)' are visible. The main text states: 'On behalf of the Swedish government, the National Library of Sweden (Kungliga biblioteket, KB) has developed national guidelines for open science. The guidelines are intended to provide support and guidance to actors in Sweden who have an important role to play in the transition to open science.'

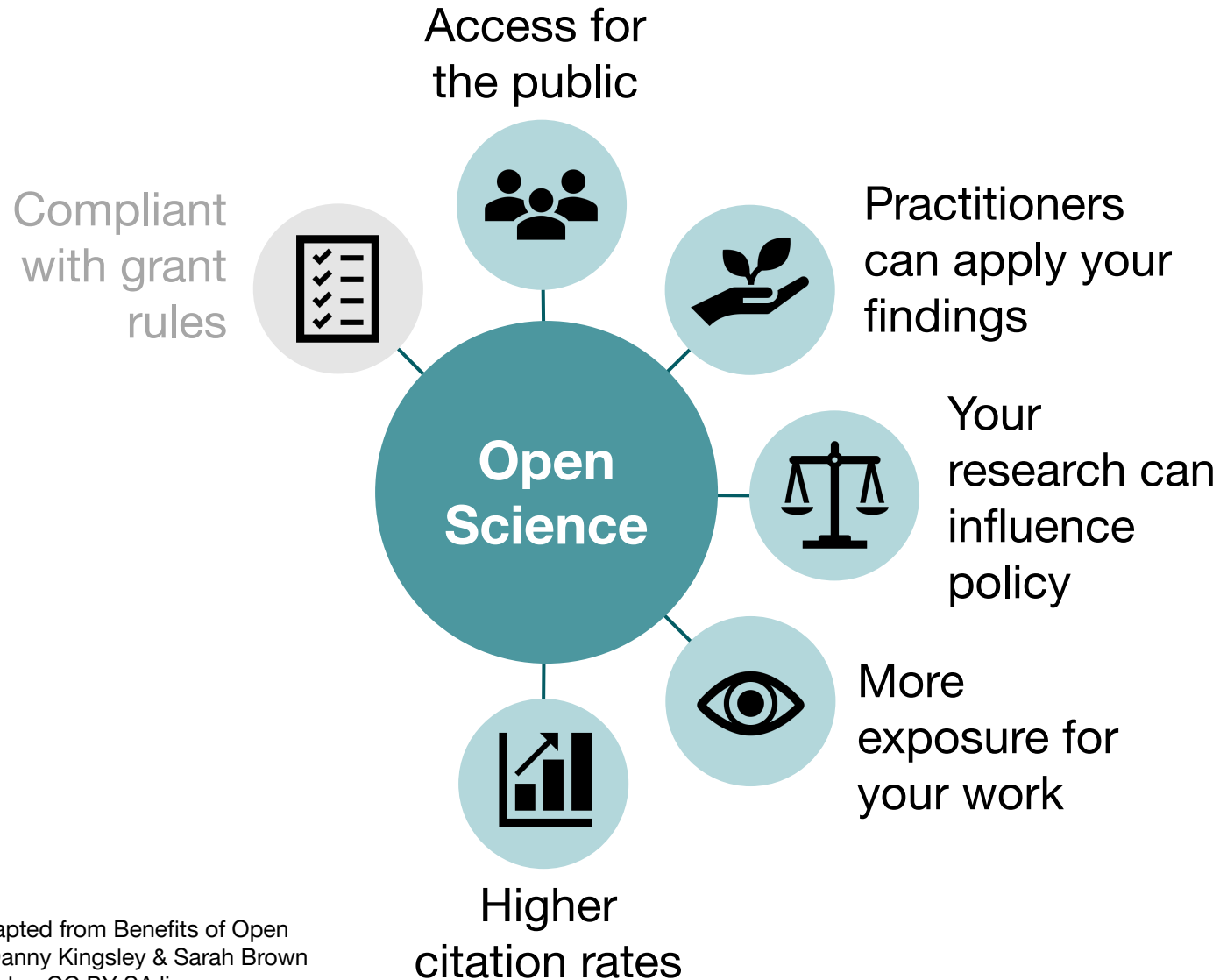
# Why does Open Science matter?



- **Benefits to society**

- Innovation
- Sustainability
- Better utilisation of public funding
- Increased public awareness and interest
- Improve transparency, visibility, and trust

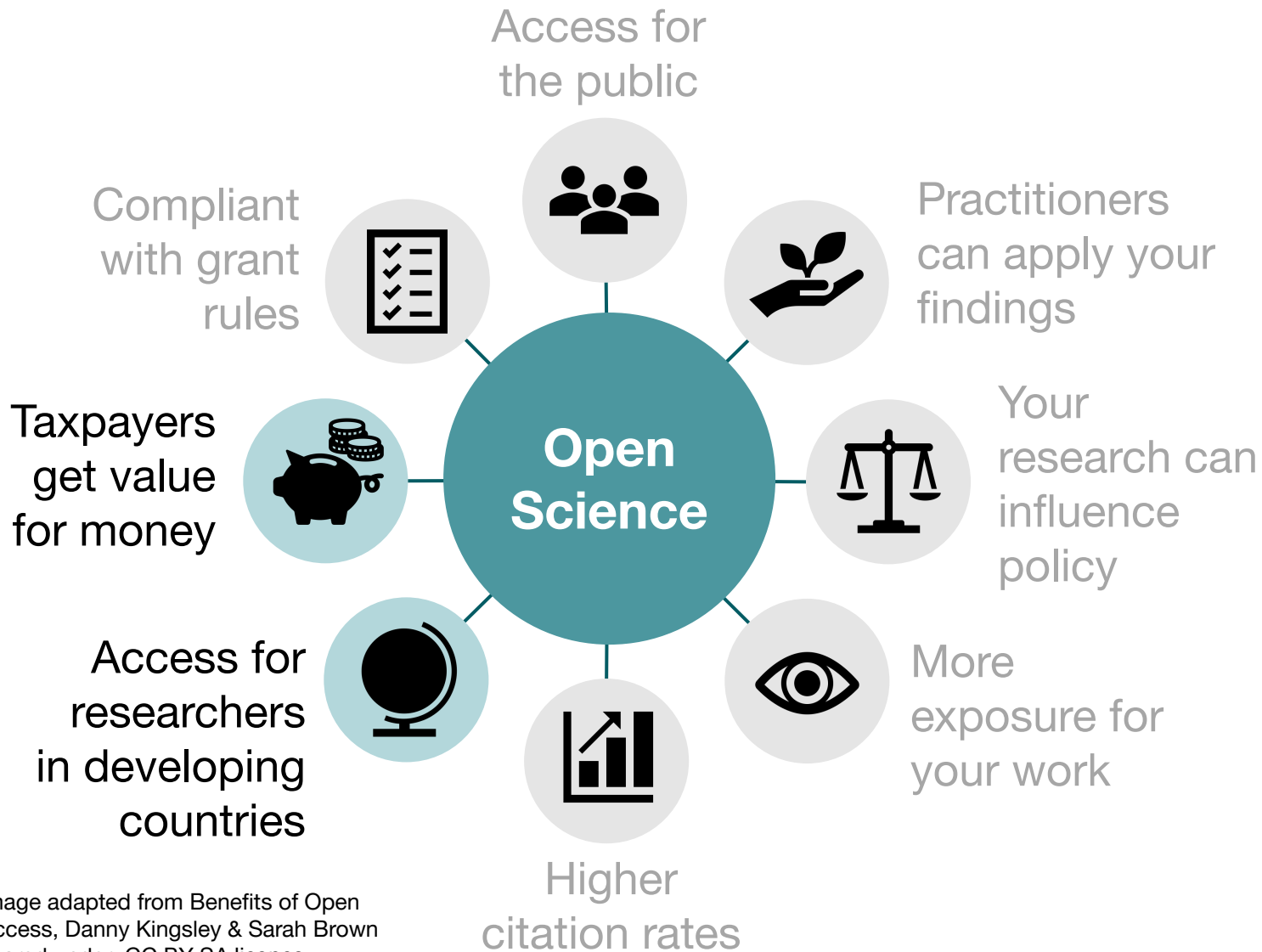
# Why does Open Science matter?



- Benefits to researchers

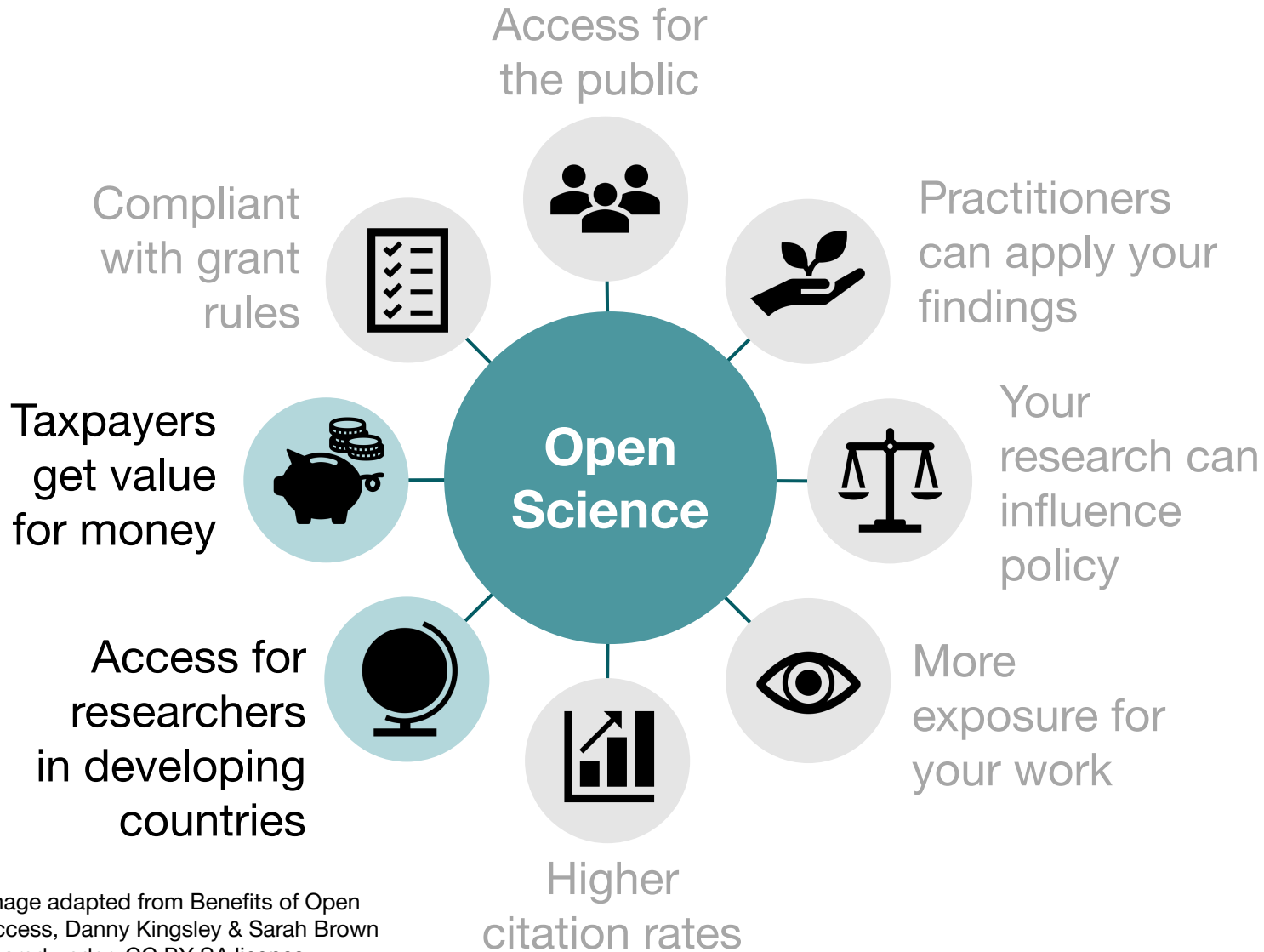
- Combining data
- Collaboration
- Addressing new questions
- Faster development of methodology
- Faster dissemination
- Increased citation
- Reproducibility

# Why does Open Science matter?



- It is the right thing to do
  - Publicly funded research data should be accessible to all
  - Published results and conclusions should be possible to be checked by others
  - Applying equity benefits all

# Why does Open Science matter?



- It is the right thing to do  
*Doing "sloppy" science & not being open and transparent, could result in:*
  - Waste of resources
  - Contributing to the current research credibility crisis
  - Contributing to the current reproducibility crisis

# Why does Open Science matter?

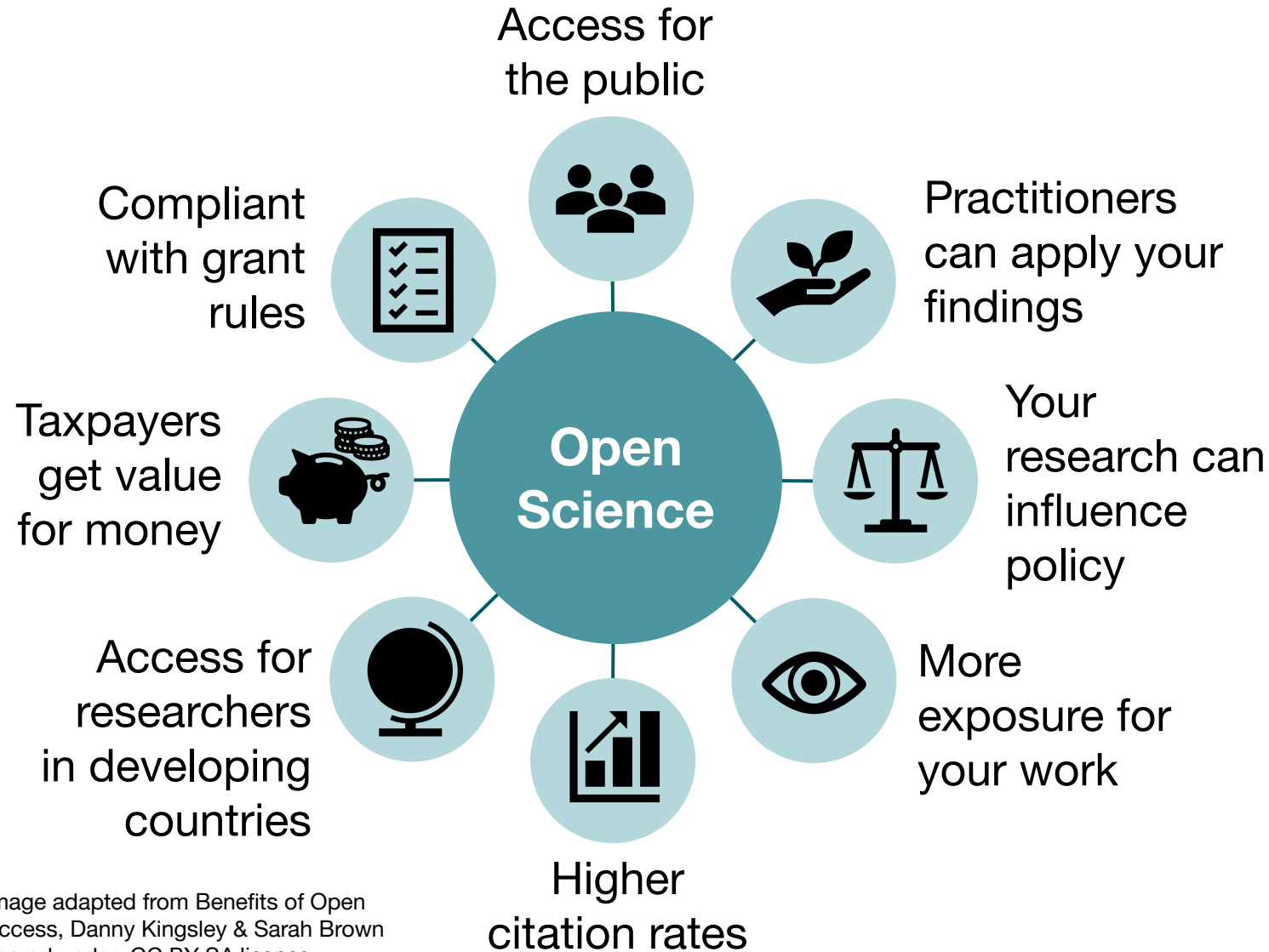


Image adapted from Benefits of Open Access, Danny Kingsley & Sarah Brown shared under [CC BY-SA licence](#)

# How?

From this



To this



Frits Ahlefeldt

# Reuse by people and machines

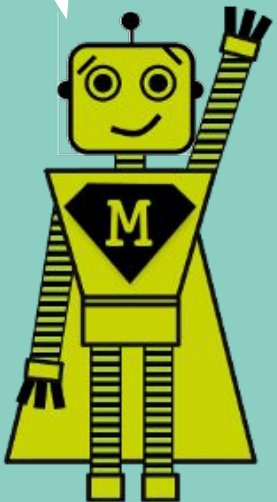
More value to publicly funded research

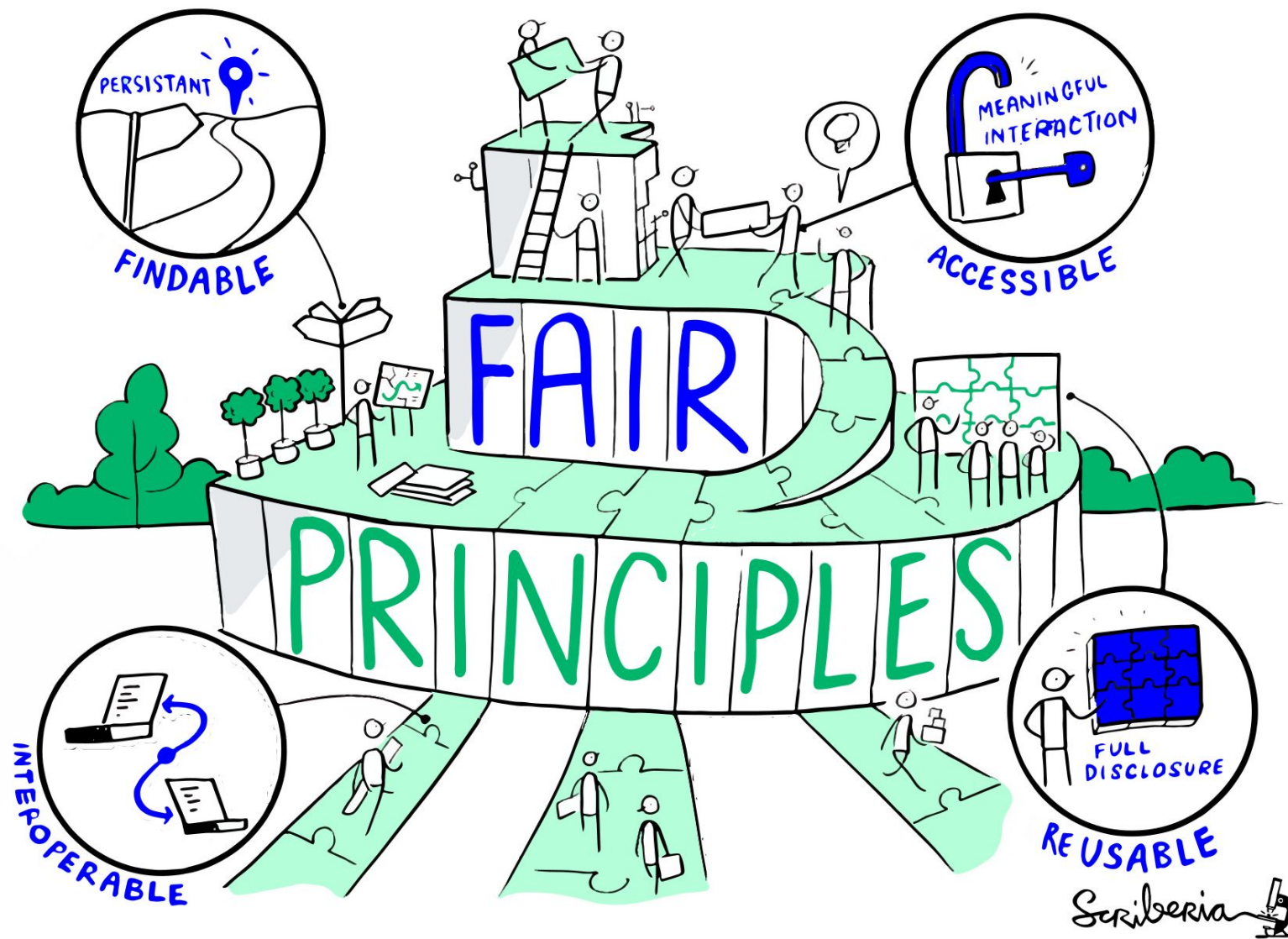
Improve peer-review process

Works with my software

High research integrity

Better research output



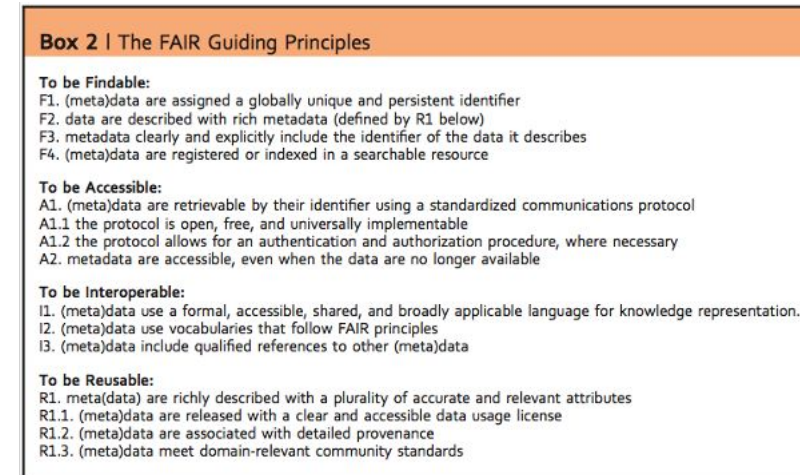


# FAIR principles



- Guidelines to make your data available to (you) others for reuse

Wilkinson, M. D., et al. (2016) The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, 160018 <http://dx.doi.org/10.1038/sdata.2016.18>



DOI: 10.1038/sdata.2016.18

Where is the metadata?

Where are the files that were generated 5 years ago?

What does this variable name mean?

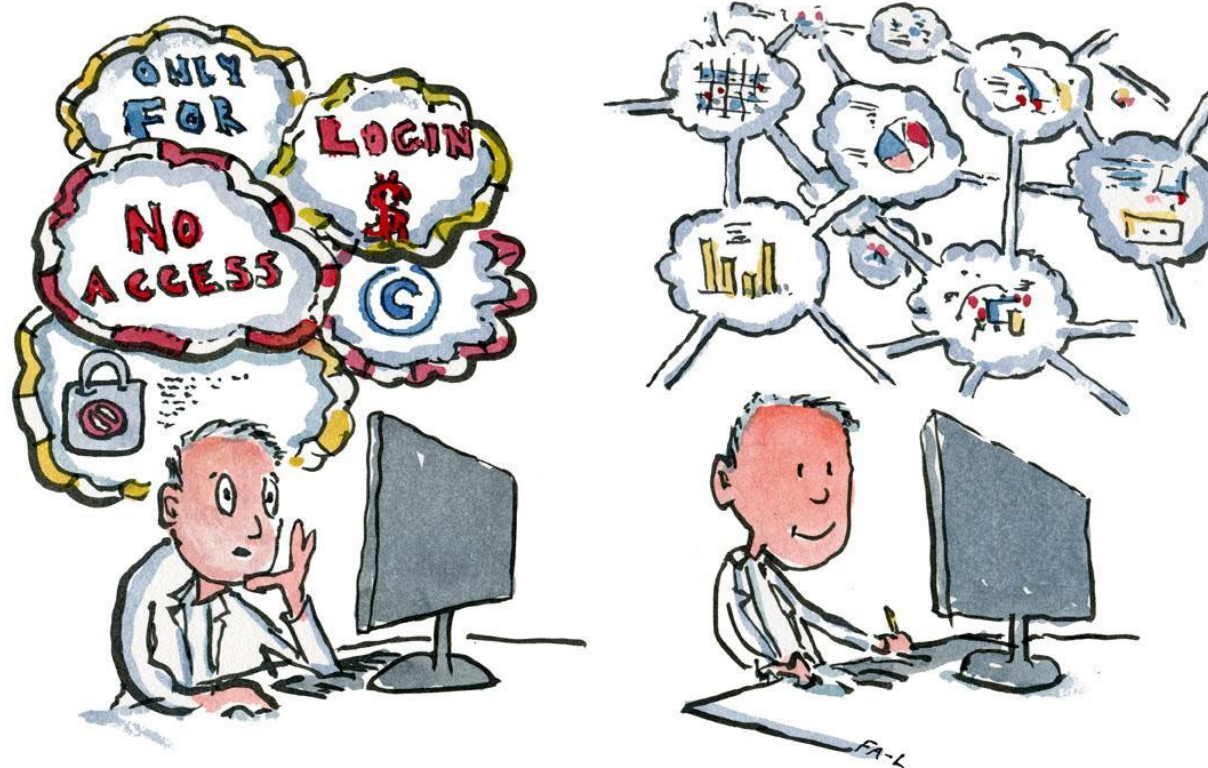
What equipment was used to acquire the data?

What does this ID mean?

How does this code work?

How do I open a .xhfg file?

Am I allowed to reuse this?



Frits Ahlefeldt



# Findable



- Data should be easily found by both, humans and machines
- Data have a globally **unique persistent identifier** (e.g. a *DOI*)
- Data are described by **metadata** (*data about data*)

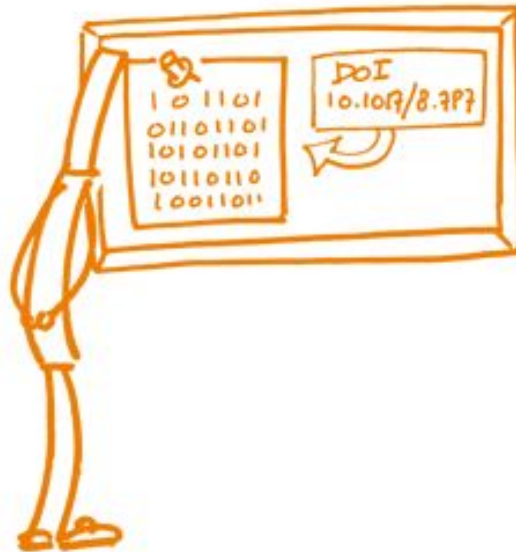


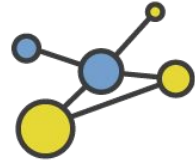


# Accessible



- Data are retrievable through a **standardised communication protocol** (open, free, allowing authentication & authorisation when necessary) (e.g. *http, sftp, etc.*)
- **Metadata are accessible, even if data is no longer available**





# I nteroperable



- Data can be **easily integrated with other data**, so it can be utilised by other applications (analysis, storage, processing)
- Metadata use **standardised vocabularies** that follow the FAIR principles (*standardised ways of capturing information about the data*)

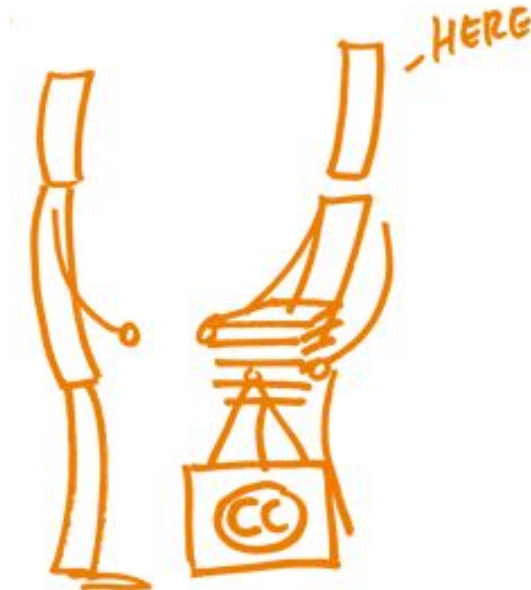




# Reusable



- Data and related metadata should be well described, so the data can be reused, replicated and/or combined in different settings
- Data have a clear data usage **license** (*under what conditions it can be reused*)





**F**indable



**A**ccessible



**I**nteroperable



**R**eusable



# Effective data management practices

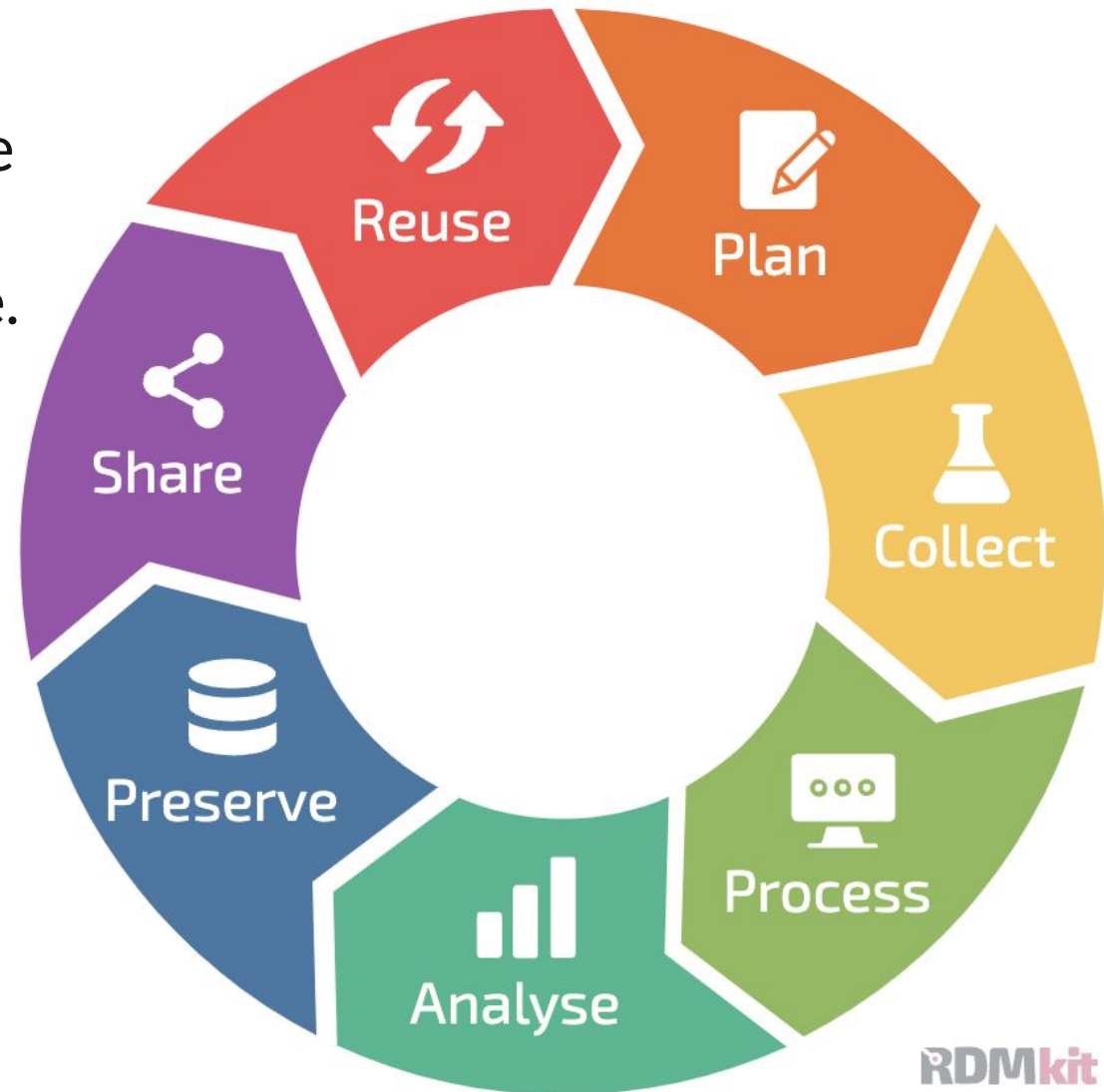
By Wolmar

# Adopt a life cycle approach to research outputs



<https://data-guidelines.scilifelab.se/>

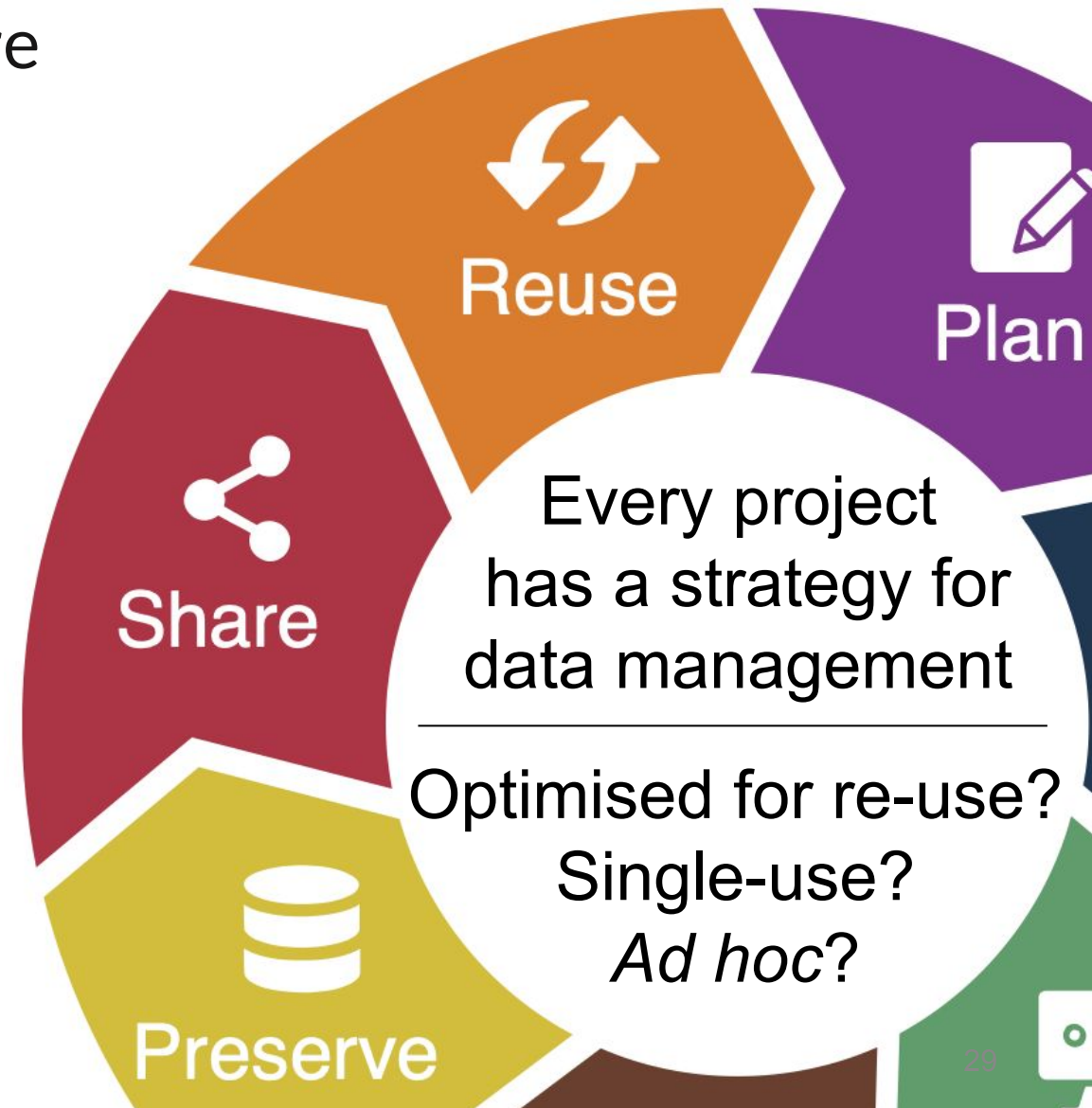
- **Get Support**—for anyone involved in life science research that is affiliated with a Swedish university or research institute.
- **Events & Training**—upcoming conferences, webinars, workshops, and training opportunities.
- **Topics**—covering the entire data life cycle with links to relevant resources.



# Life cycle perspective in context



- Good data management practices are **important in all phases** and covers aspects such as
  - Ethics and legislation
  - Information security
  - Research documentation
  - Project organisation
- Access to research outputs also needs to be secured **beyond the project's** time frame



# Life cycle perspective in context (II)



Study & data  
design

Sampling  
& specimen  
collection

Sample  
preparation

Sample analysis  
& data generation

Data processing  
to prepare inputs  
for analysis

Data  
analysis

Communicating  
results

**Procedures**

**Biosamples and instruments**

**Data and computational workflows**

**Outputs**

Information  
systems

Laboratory  
workspaces

Biobank or  
Collection

Data  
delivery

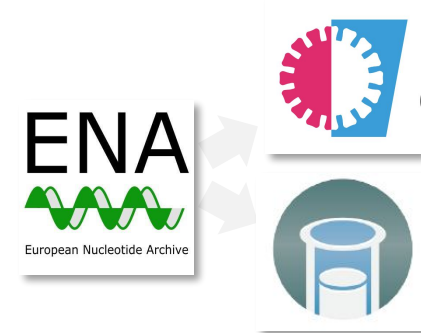
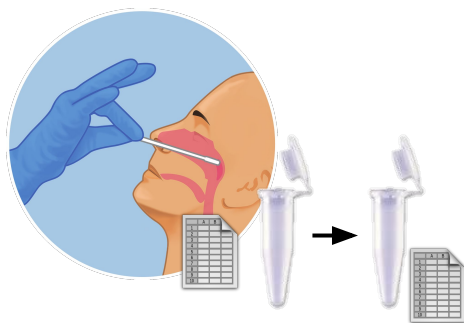
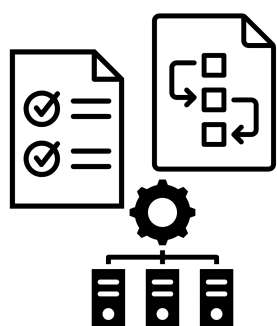
Digital  
workspaces

Local data  
archive etc

Research  
databases

“Protocol” & “project plan” icons by Justin Blake, and “infrastructure” icon by Eko Purnomo, from thenounproject.com

# Research outputs and metadata in context (I)



Study & data  
design

Sampling  
& specimen  
collection

Sample  
preparation

Sample analysis  
& data generation

Data processing  
to prepare inputs  
for analysis

Data  
analysis

Communicating  
results

## Procedures

data protection,  
ethics permit,  
infrastructure,  
standards,  
protocols,  
data dictionaries,  
data access, ...

## Biosamples and instruments

populations (statistical) and inclusion criteria,  
physical processing steps,  
working storage conditions,  
long-term storage location,  
sample quality assessment,  
sample annotations,  
reagents, instruments, kits, ...

## Data and computational workflows

digital processing steps,  
working storage conditions,  
long-term storage location,  
data quality assessment,  
sample/data annotations,  
reference data,  
analysis method...

## Outputs

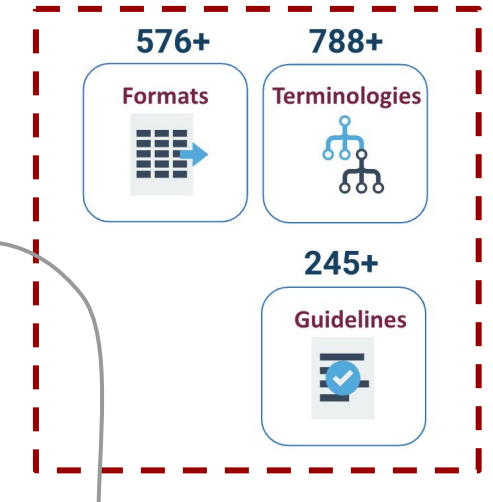
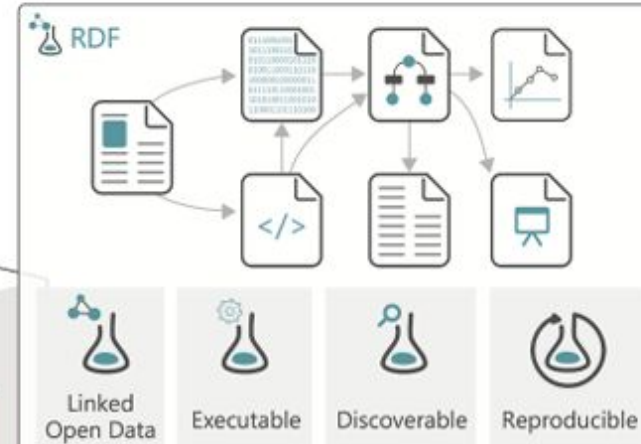
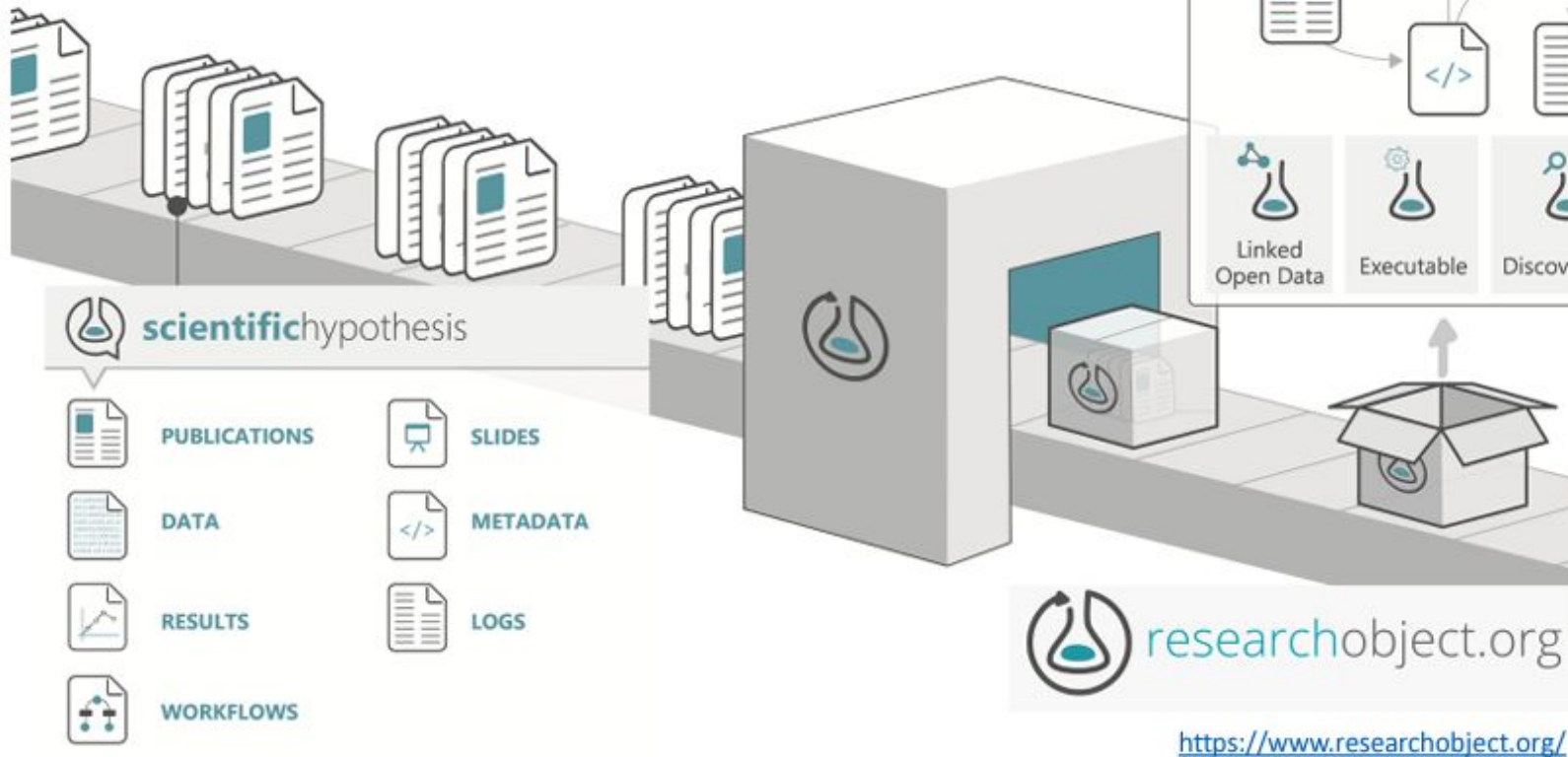
publications,  
data,  
tools,  
workflows,  
reports,  
dashboards, ...

“Protocol” & “project plan” icons by Justin Blake, and “infrastructure” icon by Eko Purnomo, from thenounproject.com

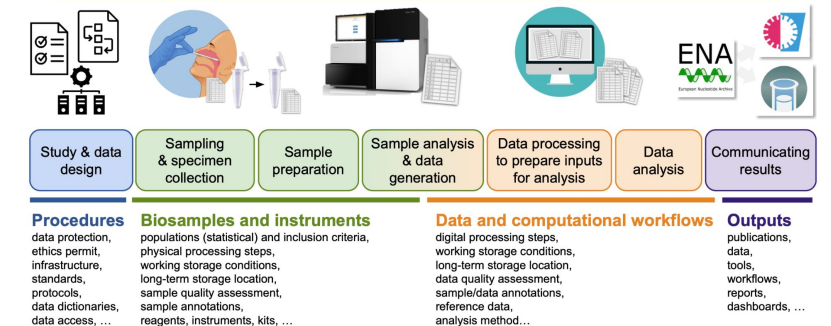
# Research outputs and metadata in context (II)



 Enabling **reproducible**, transparent research.



## Research outputs and metadata in context (I)



\*"Protocol" & "project plan" icons by Justin Blake, and "infrastructure" icon by Eko Purnomo, from thenounproject.com

43

# Making sense of your group's data

---



“ data should be readable for machines without the need for specialised or ad hoc algorithms, translators, or mappings”



How about another group's data?

# Making sense of another group's data

---



“ data should be readable for machines without the need for specialised or ad hoc algorithms, translators, or mappings”

“ data should be assessable so that judgments can be made about their reliability and the competence of those who created them”.



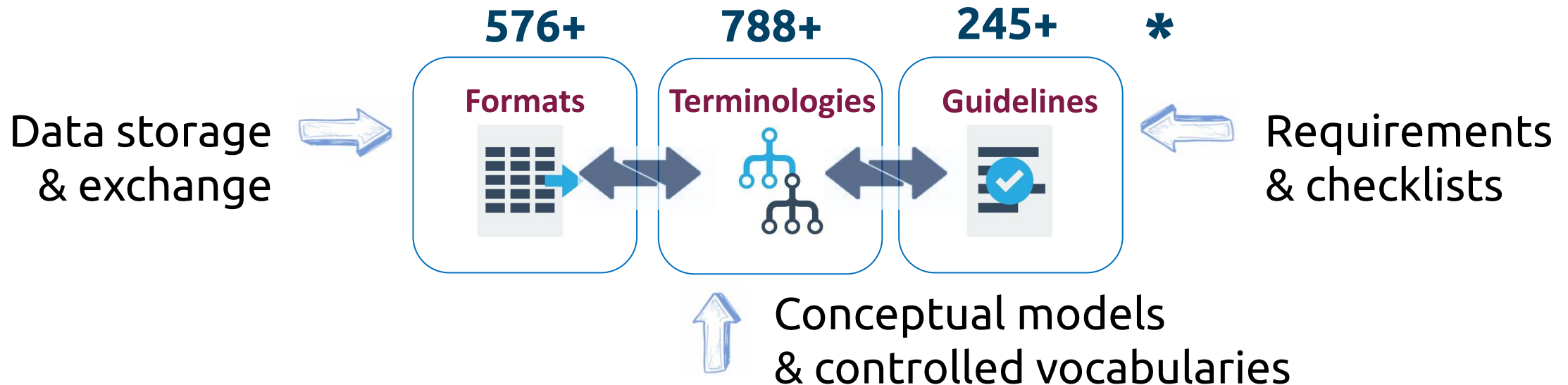
How about FAIR research data?

# Making sense of FAIR research data



“ data should be readable for machines without the need for specialised or ad hoc algorithms, translators, or mappings”

“ data should be assessable so that judgments can be made about their reliability and the competence of those who created them”.



\* Numbers from the FAIRsharing Registry

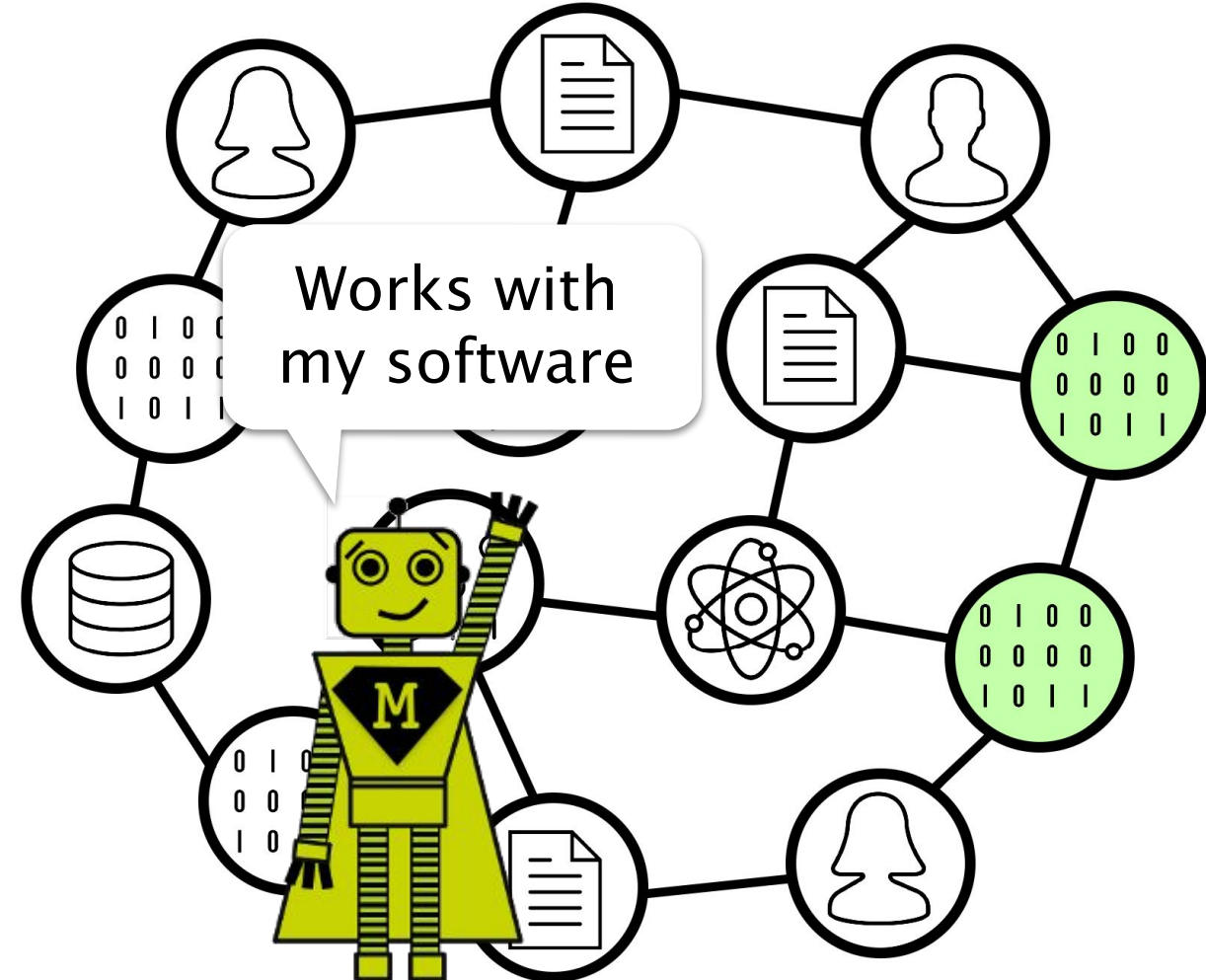
# Towards a web of FAIR data and services



Graph: "PID Graph" from the FREYA project

## "From Gutenberg to Berners-Lee"

- What types of research assets?
- Where will people be looking?
- Who should have access?
- Which standards/specifications?
- What cross-references and what documentation can you provide? (Rich metadata!)



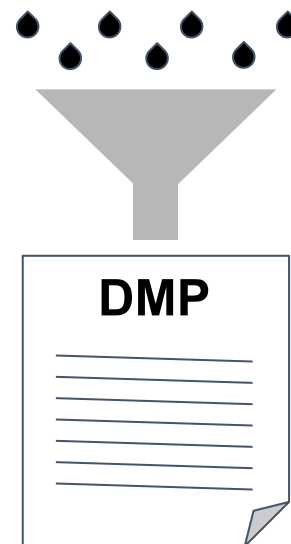
Robot: MetaManMachine by Nikola Vasiljevic (2021),  
CC BY-SA 4.0, doi:10.5281/zenodo.4471098

# Documenting your approach to FAIR practices



## Data Management Plan (DMP):

A document addressing requirements and practices for the project's data



**The Swedish Research Council:** All who are awarded a grant from the Swedish Research Council must have a data management plan if the research generates research data.

# DMP as a checklist beyond FAIR



- Identify data management **gaps**
- Establish **project-wide** standards
- Estimate **costs**
- Define **responsibilities**
- Ensure **well-managed** research data
- First step towards **FAIRness**
- Meet funder and stakeholder **demands**

Reduce time spent later on

Openness

Reproducibility

Facilitating collaboration

Return of investment

Transparency



# Sections of a common DMP template



## 1. Description of data

- What types of data will be created and/or collected?

Formats

Amount/volume of data

Instrument

Equipment

## 2. Documentation

- How will the material be documented and described?

ELN & LIMS

Collection method

Metadata standards

Versioning

## 3. Storage and backup

- How is data security, storage and backup handled?

Organization

Naming convention

Backup strategy

Access

Security

# Sections of a common DMP template (II)

---



## 4. Legal and ethical aspects

- How is data handled? Any legal requirements?

Sensitive data

Confidentiality

Intellectual property rights

## 5. Accessibility and long-term storage

- How, when, and where will research data and metadata be made accessible?

Repositories

Raw data

Code & Software

Type of storage

## 6. Responsibility and resources

- Who are the responsible persons for data management?

Organization

Naming convention

Backup strategy

Access

Security

# The Data Stewardship Wizard (DSW)



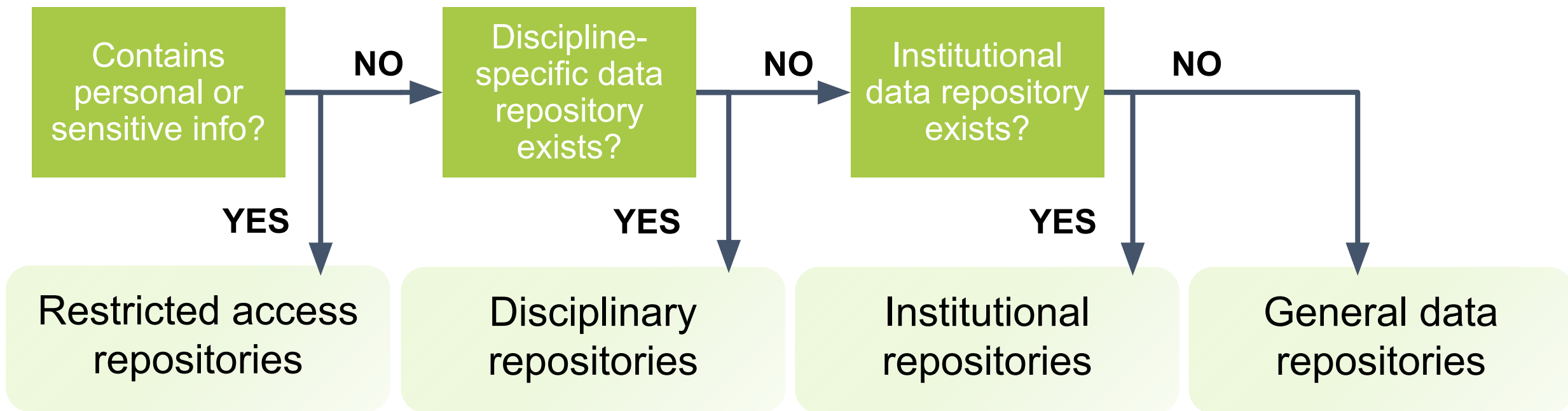
- **Interactive questionnaire** specific for life sciences.
- **Adapted templates** to national funders e.g. **Swedish Research Council** and **H2020**.
- Login by using your university account (SWAMID).
- Do you **need help**? We provide support and guidance!

Learn more:  
<https://dsw.scilifelab.se>  
[youtu.be/HY2DVnNGkAs](https://youtu.be/HY2DVnNGkAs)  
(short DSW introduction)

The screenshot displays the SciLifeLab DSW interface. On the left is a sidebar menu with the following items: 'Projects' (with a folder icon), 'Guide - Write a DMP' (with a question mark icon), 'Guide - NBIS Support Checklist' (with a question mark icon), 'Short intro DSW' (with a video icon), 'DSW workshop' (with a video icon), and 'About project templates' (with an information icon). The main content area on the right is titled 'Log In' and contains an 'Email' input field, a 'Password' input field, a 'Forgot your password?' link, and a 'Log In' button. Below the login fields, there is a section for 'Or connect with' which includes a link for 'Life Science RI (university)' with a small icon.



# Tip #1: Find a repository that promotes FAIR



<https://www.re3data.org/>

# Tip #2: Make the most of your DMP



Borghi, J. et al (2018). Support your Data. <https://doi.org/10.3897/rio.4.e26439>

## **Ad Hoc**

When it comes to my data, I have a "way of doing things" but no standard or documented plans.

## **One-Time**

I create some formal plans about how I will manage my data at the start of a project, but I generally don't refer back to them.

## **Active and Informative**

I develop detailed plans about how I will manage my data that I actively revisit and revise over the course of a project.

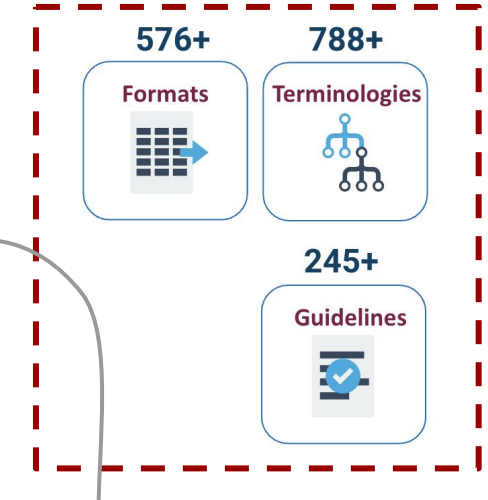
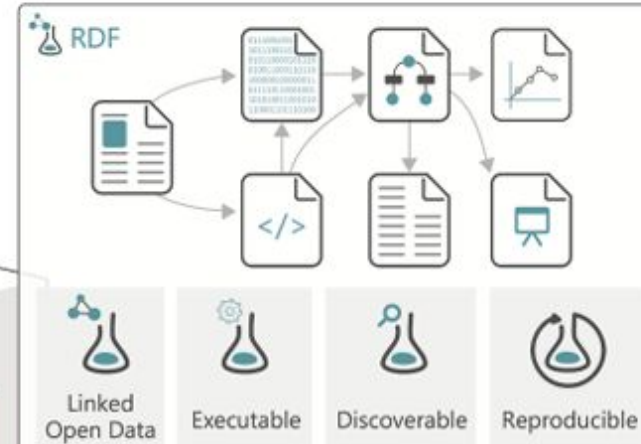
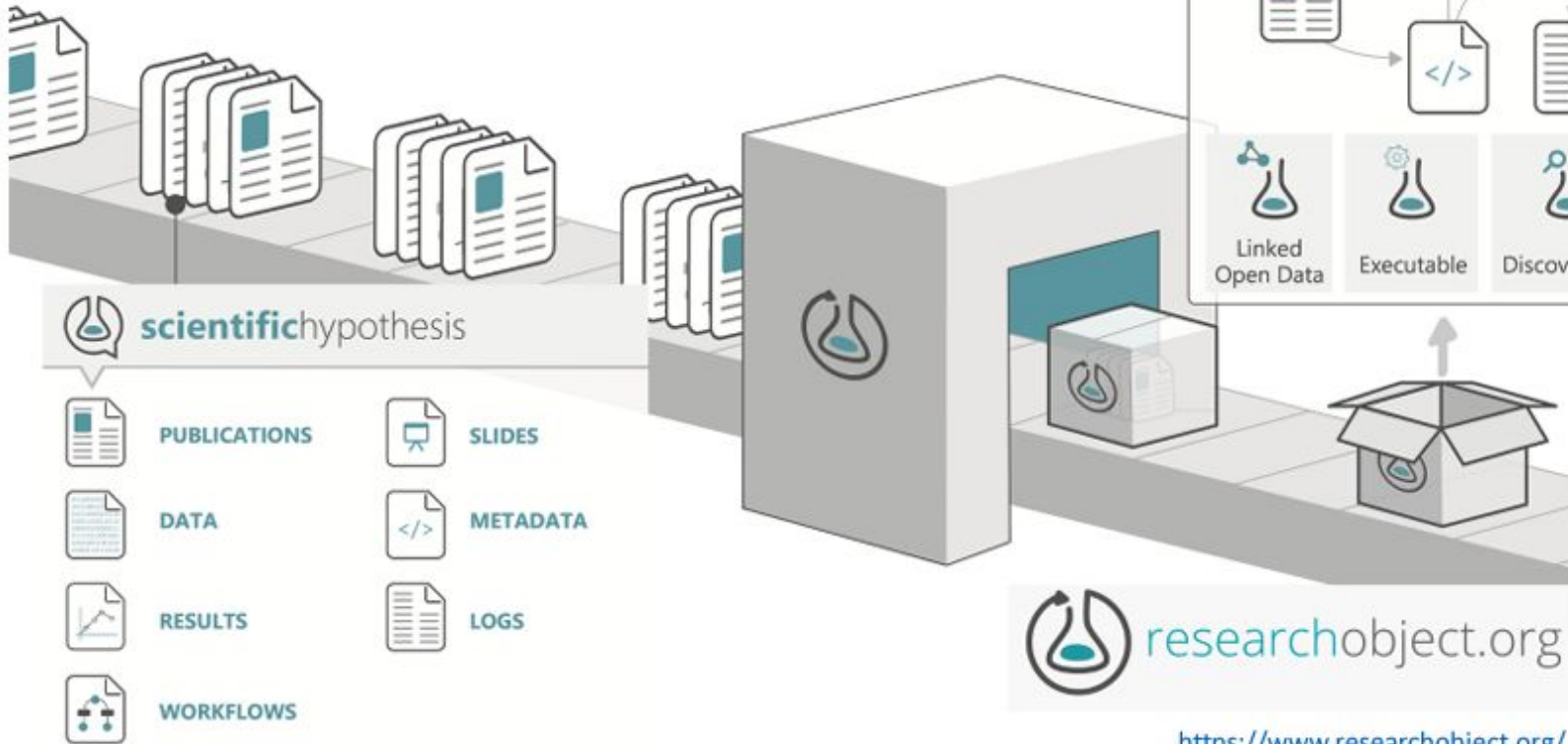
## **Optimized for Re-Use**

I have created plans for managing my data that are designed to streamline its future use by myself or others.

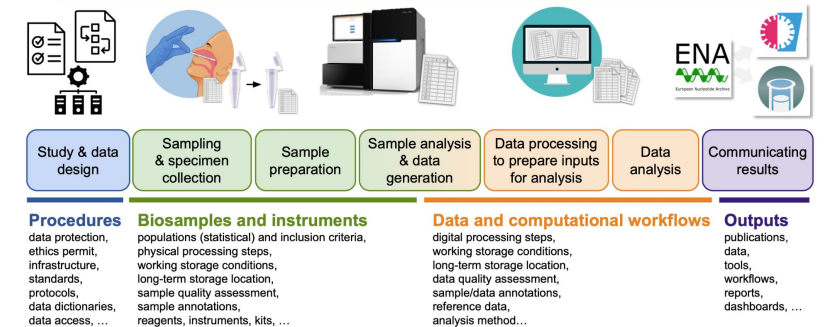
# Tip #3: Keep your research outputs organised



 Enabling **reproducible**, transparent research.



## Research outputs and metadata in context (I)



\*"Protocol" & "project plan" icons by Justin Blake, and "infrastructure" icon by Eko Purnomo, from thenounproject.com

# A life cycle perspective on research outputs!



<https://data-guidelines.scilifelab.se/>

- **Get Support**—for anyone involved in life science research that is affiliated with a Swedish university or research institute.
- **Events & Training**—upcoming conferences, webinars, workshops, and training opportunities.
- **Topics**—covering the entire data life cycle with links to relevant resources.



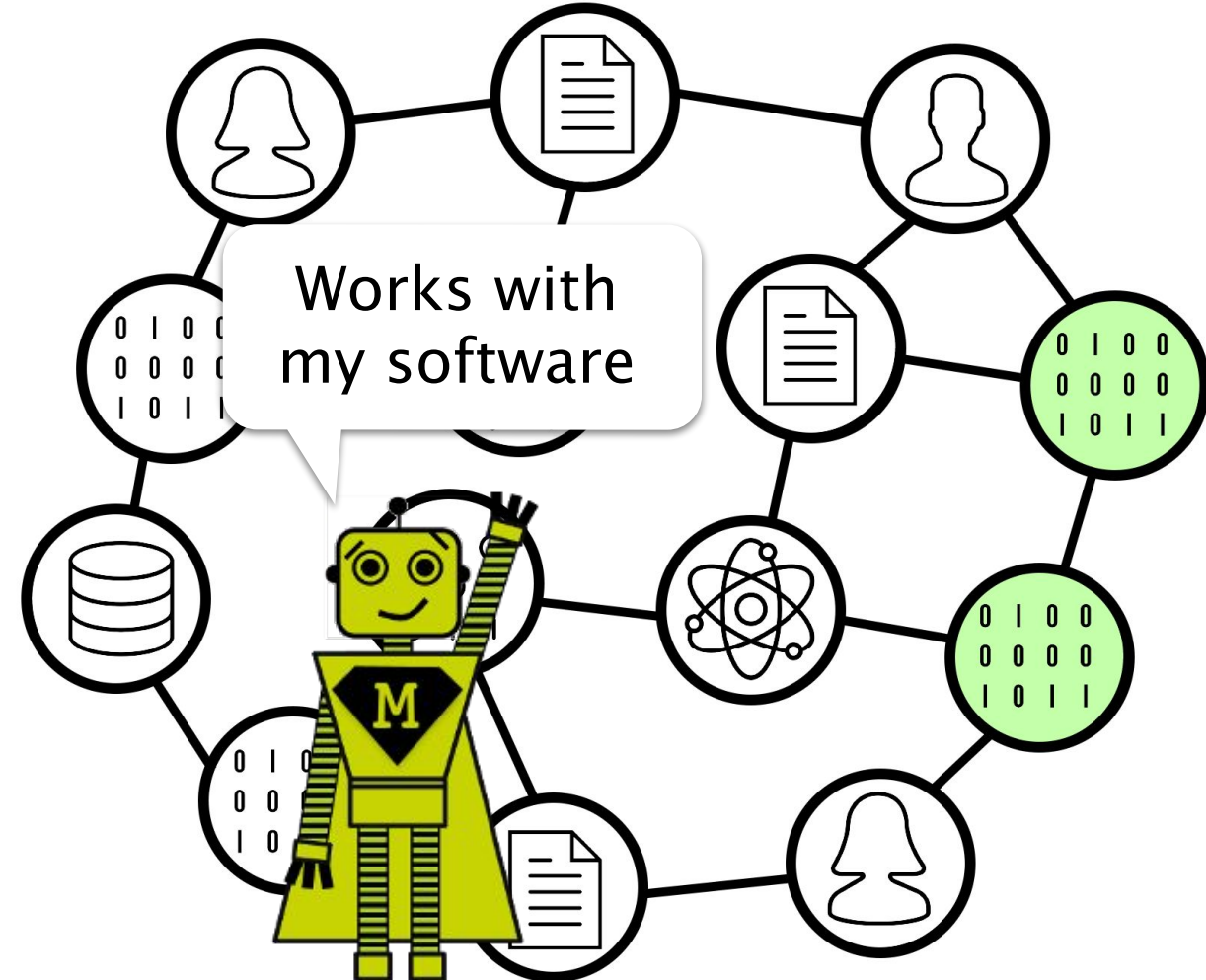
# Towards a web of FAIR data and services



Graph: "PID Graph" from the FREYA project

## "From Gutenberg to Berners-Lee"

- What types of research assets?
- Where will people be looking?
- Who should have access?
- Which standards/specifications?
- What cross-references and what documentation can you provide? (Rich metadata!)



Robot: MetaManMachine by Nikola Vasiljevic (2021),  
CC BY-SA 4.0, doi:10.5281/zenodo.4471098



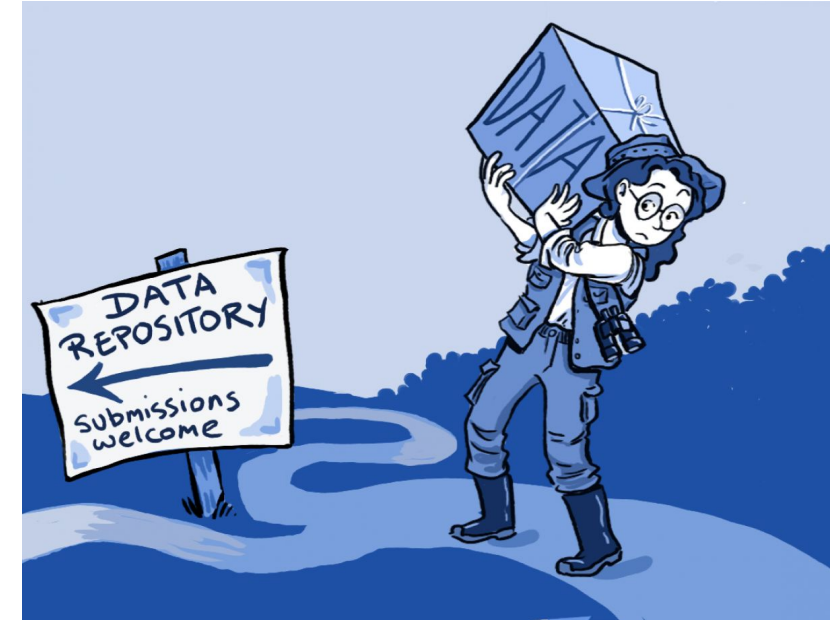
# Sharing code and machine learning models

By Arnold

# Publishing artifacts from your research project



- Open science/FAIR has expanded: not only data but any artifacts from your research project
- What do we need to consider?
  - Scientific outcome
  - Data (raw and processed)
  - **Analysis code/pipeline/workflow**
  - **Models**
  - **Tools**
  - Much more: data management plan, lab notebooks, ethical review, grant proposal, peer reviews, etc.
- Default is publishing everything, there needs to be a good reason not to do it.



Credit: <http://www.openaire.eu/blogs/research-data-management-rdm-support-at-the-university-of-vienna> License: CC ATTRIBUTION 4.0 INTERNATIONAL



# Publishing artifacts from your research project

---



- Whatever you share as part of a single project needs to be interconnected - when people find one, they should also find another
- Same principles apply to all artifacts shared

- globally unique persistent identifier (DOI or others)
- versioned
- in a (community-created) standard format
- described by (community-created) standard metadata
- with a clear license
- shared where the community will find it (possibly refer from places)
- different components work with each other
- ensured long-term availability and archiving

# Sharing code



- Short analysis script/workflow/pipeline - no matter how small
- Simple script files or a notebook or report (Jupyter Notebook, R Notebook) →
- Important to share the environment as well (packages & versions; [renv](#), requirements.txt)
- Together with your data should produce results in article/preprint/presentation etc.
- If you cannot share data, create a dummy/fake dataset with same columns and column formats.

The image shows two side-by-side screenshots of code notebooks. The left screenshot is a Jupyter Notebook titled 'The Lorenz Differential Equations'. It shows a file explorer on the left with a list of files including 'audio', 'images', 'Cpp.ipynb', 'Data.ipynb', 'Fasta.ipynb', 'Julia.ipynb', 'Lorenz.ip...', 'lorenz.py', and 'R.ipynb'. The main area displays the title and introductory text, followed by a code cell with the following content:

```
[1]: %matplotlib inline
from ipywidgets import interactive, fixed

We explore the Lorenz system of differential equations:


$$\dot{x} = \sigma(y - x)$$

```

The right screenshot is an R Notebook titled 'nb-demo.Rmd'. It shows a code cell with the following content:

```
9
10 {r}
11 summary(iris)
12
13
14 {r}
15 library(ggplot2)
16 qplot(Sepal.Length, Petal.Length, data = iris, color = Species, size =
Petal.Width)
17
```

Below the code, the R Notebook displays the output of the `summary(iris)` function as a table:

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
Min.	:4.300	:2.000	:1.000	:0.100	setosa :50
1st Qu.:	:5.100	:2.800	:1.600	:0.300	versicolor:50
Median	:5.800	:3.000	:4.350	:1.300	virginica :50
Mean	:5.843	:3.057	:3.758	:1.199	
3rd Qu.:	:6.400	:3.300	:5.100	:1.800	
Max.	:7.900	:4.400	:6.900	:2.500	

Below the table, the R Notebook displays a scatter plot of Petal.Length vs. Sepal.Length, colored by Species and sized by Petal.Width. The plot shows three distinct clusters of points corresponding to the three species: setosa (red), versicolor (green), and virginica (blue). The size of the points represents the Petal.Width, with a legend on the right showing sizes for 0.5, 1.0, 1.5, 2.0, and 2.5.

Credit:  
<https://docs.jupyter.org/en/latest/>  
License: BSD-3-Clause license

Credit:  
<https://bookdown.org/yihui/rmarkdown/notebook.html>  
License: CC BY-NC-SA 4.0 Deed

# Sharing code



- GitHub is convenient but not sufficient for FAIR
  - not an archival repository
  - does not provide a persistent identifier.
- Deposit a copy of your code to a repository that will provide a persistent identifier.
- Zenodo offers integration with GitHub, you can get a DOI for each release (Figshare offers similar functionality).
- Provide the DOI or another persistent identifier when referring to the code.

The screenshot shows the Zenodo release page for python-pillow/Pillow version 10.2.0. The page includes a header with the publication date (January 2, 2024) and version number (10.2.0). It features a list of authors, a link to the release notes, and a 'Changes' section detailing updates. On the right, there are statistics for views (70K) and downloads (4K), a list of previous versions, and a section for external resources. A blue arrow points from the text in the list to the Zenodo release page.

Published January 2, 2024 | Version 10.2.0

python-pillow/Pillow: 10.2.0

Andrew Murray<sup>1</sup>; Hugo van Kemenade<sup>1</sup>; wiredfool; Jeffrey A. Clark (Alex)<sup>2</sup>; Alexander Karpinsky<sup>3</sup>; Ondrej Baranovič; Christoph Gohlke; Jon Dufresne<sup>4</sup>; Yay295; Matthew Brett<sup>5</sup>; DWesl; David Schmidt; Konstantin Kopachev<sup>6</sup>; Alastair Houghton<sup>7</sup>; REDxEYE; Sandro Mani<sup>8</sup>; Steve Landey<sup>9</sup>; Aarni Koskela<sup>10</sup>; Josh Ware<sup>11</sup>; vashek; Piolie; Jason Douglas<sup>12</sup>; Stanislaw T.; David Caro; Uriel Martinez<sup>13</sup>; Steve Kossouho<sup>14</sup>; Riley Lahd; Antony Lee

<https://pillow.readthedocs.io/en/stable/releasenotes/10.2.0.html>

**Changes**

- Add `keep_rgb` option when saving JPEG to prevent conversion of RGB colorspace #7553 [ @bgilbert ]
- Trim negative glyph offsets in `ImageFont.getmask()` #7672 [ @nulano ]
- Removed unnecessary "pragma: no cover" #7668 [ @radarhere ]
- Trim glyph size in `ImageFont.getmask()` #7669 [ @radarhere ]
- Fix loading IPTC images and update test #7667 [ @nulano ]
- Allow uncompressed TIFF images to be saved in chunks #7650 [ @radarhere ]
- Concatenate multiple JPEG EXIF markers #7496 [ @radarhere ]
- Changed IPTC tile tuple to match other plugins #7661 [ @radarhere ]
- Do not assign new fp attribute when exiting context manager #7566 [ @radarhere ]
- Support arbitrary masks for uncompressed RGB DDS images #7589 [ @radarhere ]
- Support setting ROWSPERSTRIP tag #7654 [ @radarhere ]
- Apply `ImageFont.MAX_STRING_LENGTH` to `ImageFont.getmask()` #7662 [ @radarhere ]
- Optimise `ImageColor` using `functools.lru_cache` #7657 [ @hugovk ]
- Restricted environment keys for `ImageMath.eval()` #7655 [ @radarhere ]
- Optimise `ImageMode.getmode` using `functools.lru_cache` #7641 [ @hugovk ]
- Added trusted PyPI publishing #7616 [ @radarhere ]
- Compile FriBidi for Windows ARM64 #7629 [ @nulano ]
- Fix incorrect color blending for overlapping glyphs #7497 [ @ZachNagengast ]
- Add .git-blame-ignore-revs file #7528 [ @akx ]
- Attempt memory mapping when tile args is a string #7565 [ @radarhere ]
- Fill identical pixels with transparency in subsequent frames when saving GIF #7568 [ @radarhere ]
- Removed unnecessary string length check #7560 [ @radarhere ]

**Versions**

Version	Date
Version 10.2.0	Jan 2, 2024
Version 10.0.1	Sep 15, 2023
Version 10.0.0	Jul 1, 2023
Version 9.5.0	Apr 1, 2023
Version 9.4.0	Jan 2, 2023

[View all 51 versions](#)

**Cite all versions?** You can cite all versions by using the DOI [10.5281/zenodo.596518](https://doi.org/10.5281/zenodo.596518). This DOI represents all versions, and will always resolve to the latest one. [Read more.](#)

**External resources**

Available in

[python-pillow/Pillow](#)  
Release: 10.2.0

DOI [10.5281/zenodo.596518](https://doi.org/10.5281/zenodo.596518)



# Sharing tools

## FAIR Principles for Research Software (FAIR4RS Principles)

<https://doi.org/10.15497/RDA00068>

### **F: Software, and its associated metadata, is easy for both humans and machines to find.**

F1. Software is assigned a globally unique and persistent identifier.

- F1.1. Components of the software representing levels of granularity are assigned distinct identifiers.
- F1.2. Different versions of the software are assigned distinct identifiers.

F2. Software is described with rich metadata.

F3. Metadata clearly and explicitly include the identifier of the software they describe.

F4. Metadata are FAIR, searchable and indexable.

### **A: Software, and its metadata, is retrievable via standardized protocols.**

A1. Software is retrievable by its identifier using a standardized communications protocol.

- A1.1. The protocol is open, free, and universally implementable.
- A1.2. The protocol allows for an authentication and authorization procedure, where necessary.

A2. Metadata are accessible, even when the software is no longer available.

### **I: Software interoperates with other software by exchanging data and/or metadata, and/or through interaction via application programming interfaces (APIs), described through standards.**

I1. Software reads, writes and exchanges data in a way that meets domain-relevant community standards.

I2. Software includes qualified references to other objects.

### **R: Software is both usable (can be executed) and reusable (can be understood, modified, built upon, or incorporated into other software).**

R1. Software is described with a plurality of accurate and relevant attributes.

- R1.1. Software is given a clear and accessible license.
- R1.2. Software is associated with detailed provenance.

R2. Software includes qualified references to other software.

R3. Software meets domain-relevant community standards.



# Sharing machine learning models

Share:

- training dataset as data
- pre-processing and training scripts as code
- resulting models in a way that will allow others to build on them
- ... or use them for inferences:
  - ◆ an inference script + good description for how to use it
  - ◆ Docker Image with everything included
- No community standard for describing models (different attempts now).
- Some examples for inspiration:
  - Google Model cards <https://modelcards.withgoogle.com/about>
  - HuggingFace <https://huggingface.co/docs/hub/model-cards>
  - BioImage.io [https://bioimage.io/docs/#/bioimageio\\_model\\_spec](https://bioimage.io/docs/#/bioimageio_model_spec)

## Model Card for {{ model\_id | default("Model ID", true) }}

{{ model\_summary | default("", true) }}

### Model Details

#### Model Description

{{ model\_description | default("", true) }}

- **Developed by:** {{ developers | default("[More Information Needed]", true) }}
- **Funded by [optional]:** {{ funded\_by | default("[More Information Needed]", true) }}
- **Shared by [optional]:** {{ shared\_by | default("[More Information Needed]", true) }}
- **Model type:** {{ model\_type | default("[More Information Needed]", true) }}
- **Language(s) (NLP):** {{ language | default("[More Information Needed]", true) }}
- **License:** {{ license | default("[More Information Needed]", true) }}
- **Finetuned from model [optional]:** {{ base\_model | default("[More Information Needed]", true) }}

#### Model Sources [optional]

- **Repository:** {{ repo | default("[More Information Needed]", true) }}
- **Paper [optional]:** {{ paper | default("[More Information Needed]", true) }}
- **Demo [optional]:** {{ demo | default("[More Information Needed]", true) }}

#### Uses

##### Direct Use

{{ direct\_use | default("[More Information Needed]", true) }}

##### Downstream Use [optional]

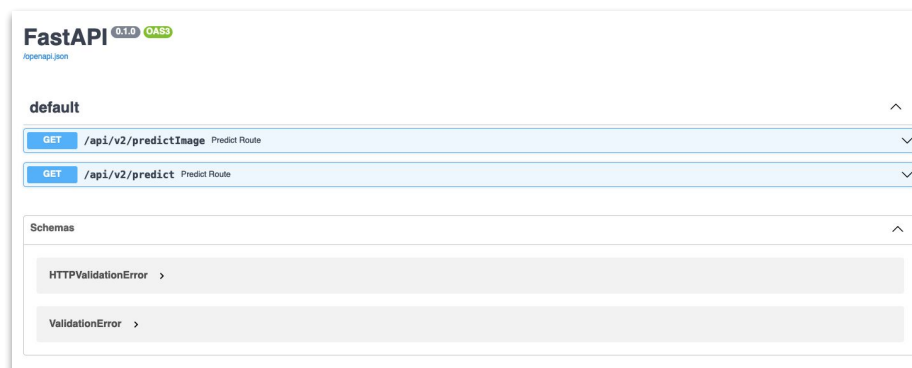
Excerpt from HuggingFace model card template:

[https://github.com/huggingface/huggingface\\_hub/blob/main/src/huggingface\\_hub/templates/modelcard\\_template.md](https://github.com/huggingface/huggingface_hub/blob/main/src/huggingface_hub/templates/modelcard_template.md)

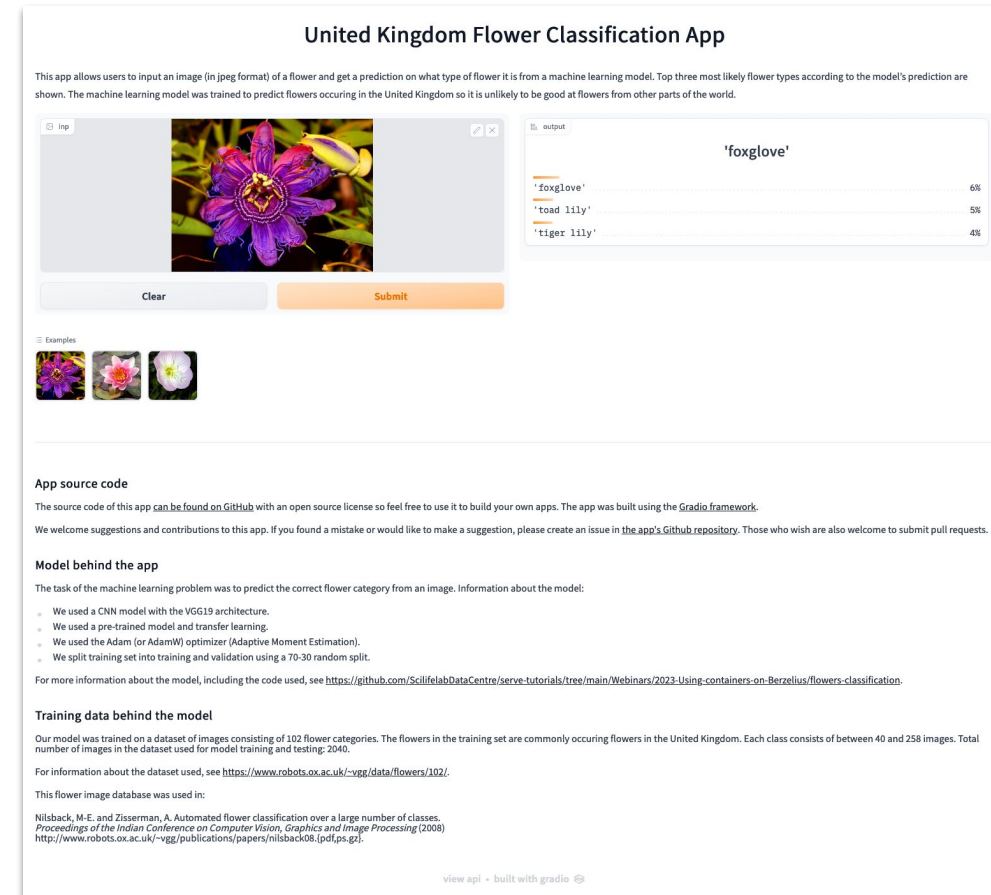


# Sharing machine learning models

- You might want to make your model available online with a URL
  - API endpoint for programmatic access
  - An app with a graphical user interface
- Useful as additional material for a publication
- To allow the research community to test out and explore
- (Prototype of) a public service



Example description of API endpoints, <https://cplogdapi.serve.scilifelab.se/docs>

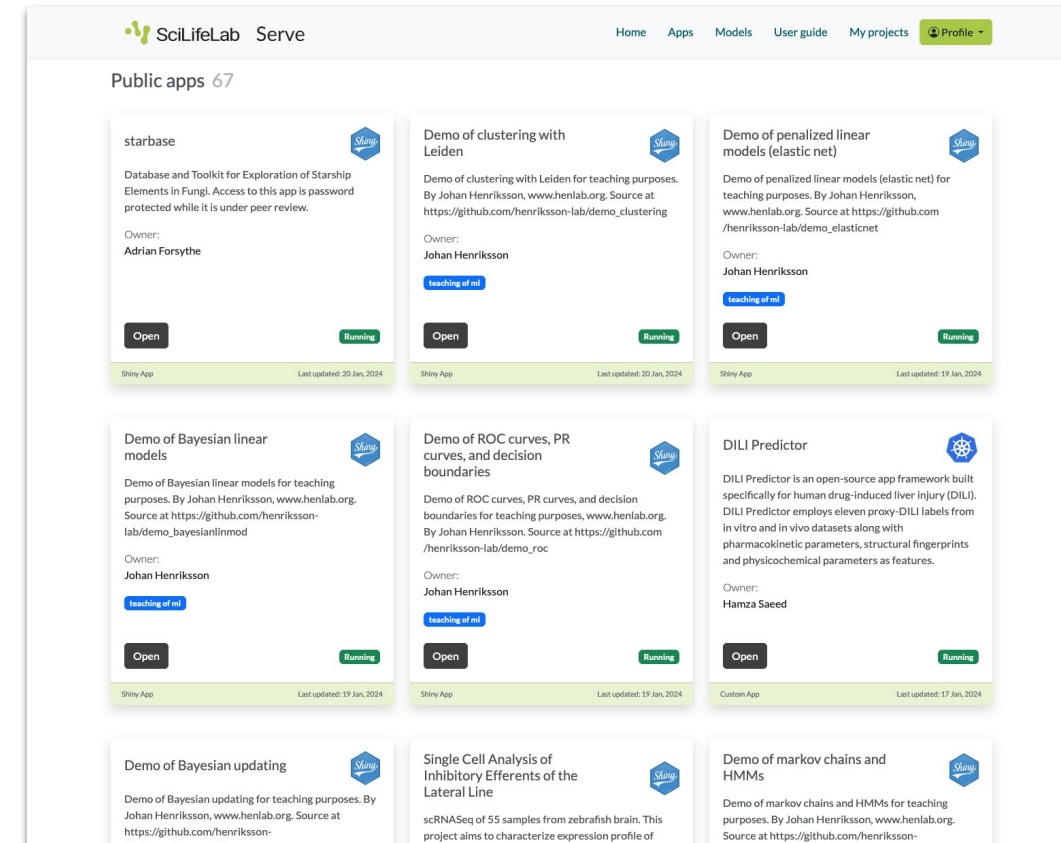


# SciLifeLab Serve



## - Hosting ML models and applications + FAIR compliance

- **Machine learning model serving**
  - Serving of trained machine learning models using dedicated tools and receive API endpoints for inference – e.g. PyTorch Serve, Tensorflow Serving, MLFlow Serve, etc.
- **Application hosting**
  - Host applications built with Shiny, Dash, Gradio, Streamlit, Flask, etc. with user interfaces or API endpoints to allow others to explore your data or results, or to provide tools based on your deployed machine learning models.
- **Example: Boreal Rhizospheric Atlas**  
<https://boreal-atlas.serve.scilifelab.se/>
- Dedicated team offers consultations, support, and training.  
Contact: [serve@scilifelab.se](mailto:serve@scilifelab.se)



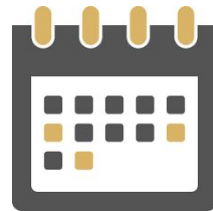
Screenshot of <https://serve.scilifelab.se/apps/>

# Thank you! Tack!



- **Contact us throughout the entire data life cycle!**  
<https://data-guidelines.scilifelab.se/> or  
[data-management@scilifelab.se](mailto:data-management@scilifelab.se)

## Upcoming seminars/workshop:



- February 28:** Data sharing (SciLifeLab Data Repository and Serve)
- March 27:** Data Management Plans
- April 17:** Sensitive data
- April 23–25:** Introduction to Data Management Practices (NBIS workshop)
- May 15/22:** Open Science



**Conclusions / Q&A (10 min)**