TAMPERE UNIVERSITY OF TECHNOLOGY

Open science and data management

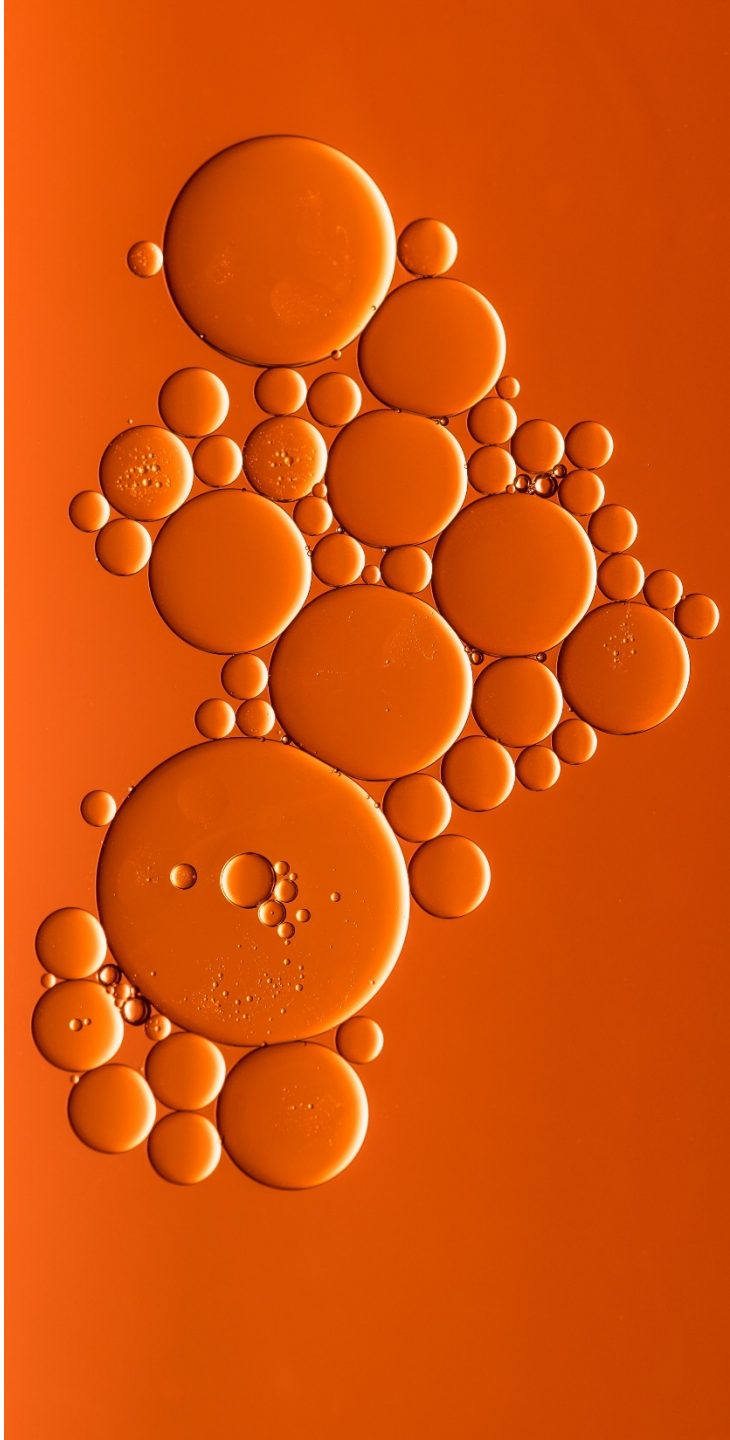
Transferable skills/Horizon 2020

Katja Fält, PhD
(ORCID: 0000-0002-6172-5377)
&
Kaisa Kulkki
(ORCID: 0000-0002-9161-6454)

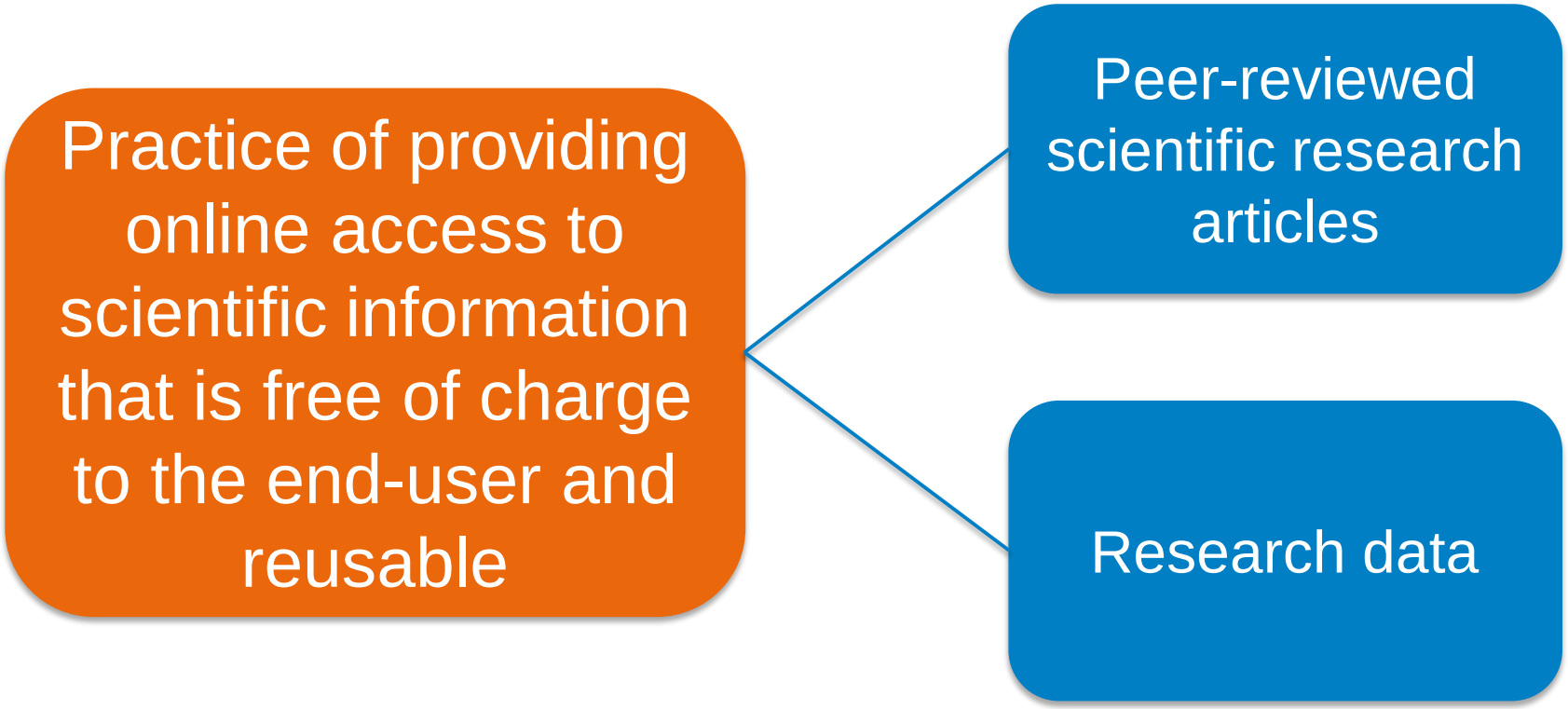
OVERVIEW

- Open Science
- Open Access Publishing
- Data Management
- Open Data
- Altmetrics
- ORCID ID
- Summary

OPEN SCIENCE



WHAT IS OPEN SCIENCE?

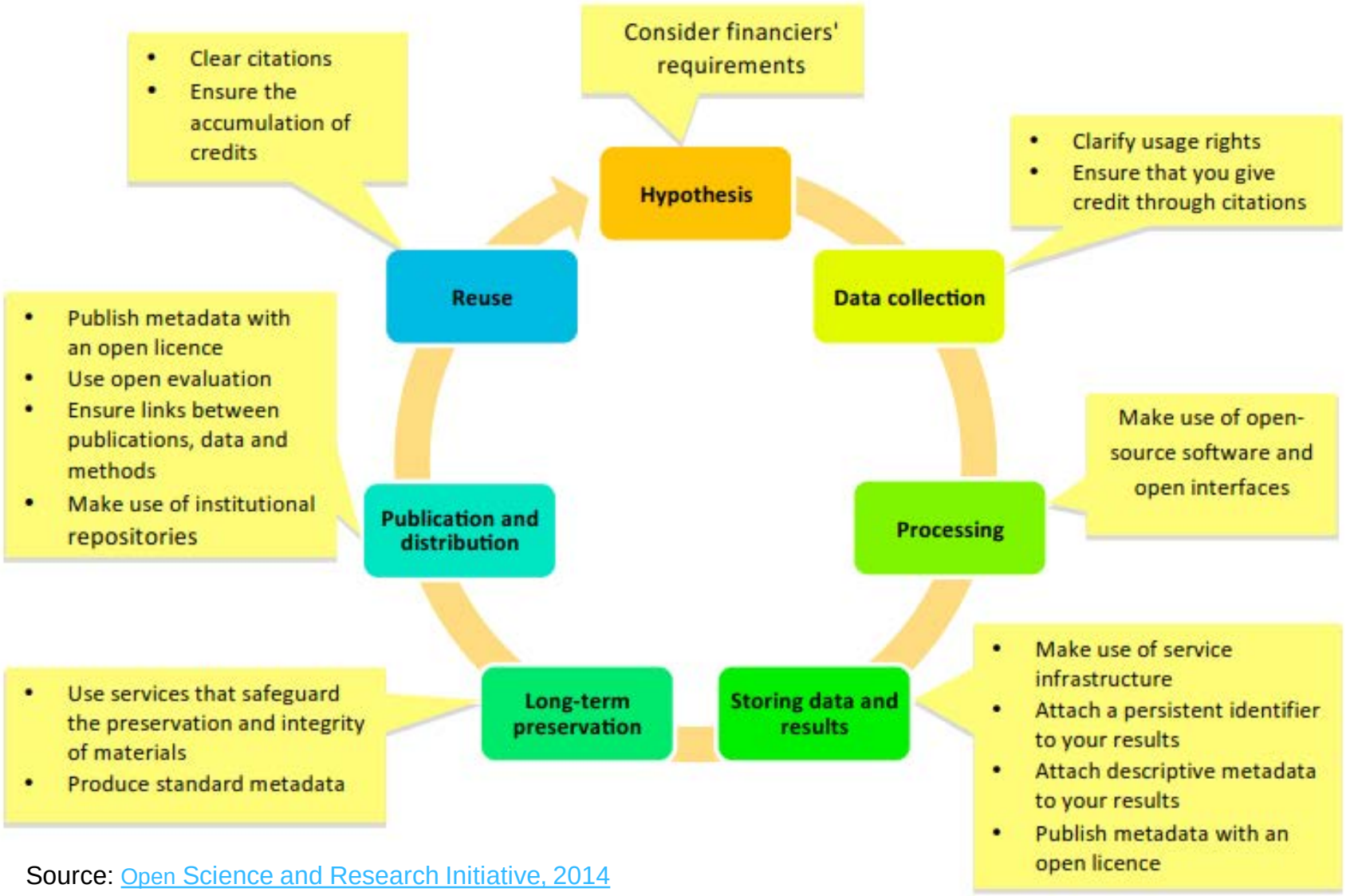


Definition by EU OA Pilot Guide

BENEFITS OF OPEN SCIENCE (e.g.)

- Better quality of research results
- Increases visibility, improves impact
- Reproducibility of research results
- Faster transfer of information
- Equal access to research

Promoting openness at different stages of the research process



Source: Open Science and Research Initiative, 2014

OA AND HORIZON 2020

- **Scientific publications**: open access to all peer-reviewed scientific publications

- **Research data**: open access is the default setting for research data generated in Horizon 2020 but under certain conditions an opt-out is possible

OPEN ACCESS PUBLISHING

TAMPERE UNIVERSITY OF TECHNOLOGY

OPEN ACCESS PUBLISHING

Open Access is the **free, immediate, online availability of research articles** coupled with the rights to use these articles fully in the digital environment

Definition by SPARC

WHY PUBLISH OPEN ACCESS?

- Wider audience: policy makers, companies, general public, researchers…
- Citation advantage for OA articles
- Requirements of research funders

OPEN ACCESS MODELS

Gold Open Access

- Publishing an article in an OA journal
- Article processing charge (APC) possible

Green Open Access

- Archiving a copy of an article in a repository after it has been published commercially
- Self-archiving

Hybrid OA: a subscription-based journal allows authors to make individual articles open access on payment of an article processing charge

Points to consider in selecting a (OA) journal

In Ulrichsweb catalog? In Directory of Open Access Journals?

The field and scope of the journal?

Do researchers on your field publish in this journal? Do they read it?

Indexed in Web of Science or Scopus?

Do you reach your target group through this journal?

Copyright and Open Access policy?

Is the publisher well-known?

Research data policy of the journal?

Is the journal refereed?

Does your funder require OA publishing?

Who is on the editorial board?

An ISSN number?

Impact factor?

DOIs provided for articles?

THINK CAREFULLY IF...

- Journal scope statement absent or vague
- Journal web site difficult to locate or identify
- Publisher direct marketing obtrusive
- Information on peer-review, copyright absent or unclear
- Publisher has negative reputation
- More indicators: https://www.gvsu.edu/library/sc/open-access-journal-quality-indicators-5.htm

SELF-ARCHIVING

- Usually only the accepted manuscript version (e.g. Elsevier, Springer, IEEE, Wiley)
- Many publishers apply embargo periods: self-archived version will be available after 1-2 years
- Keep all versions!

Figure: Public Domain

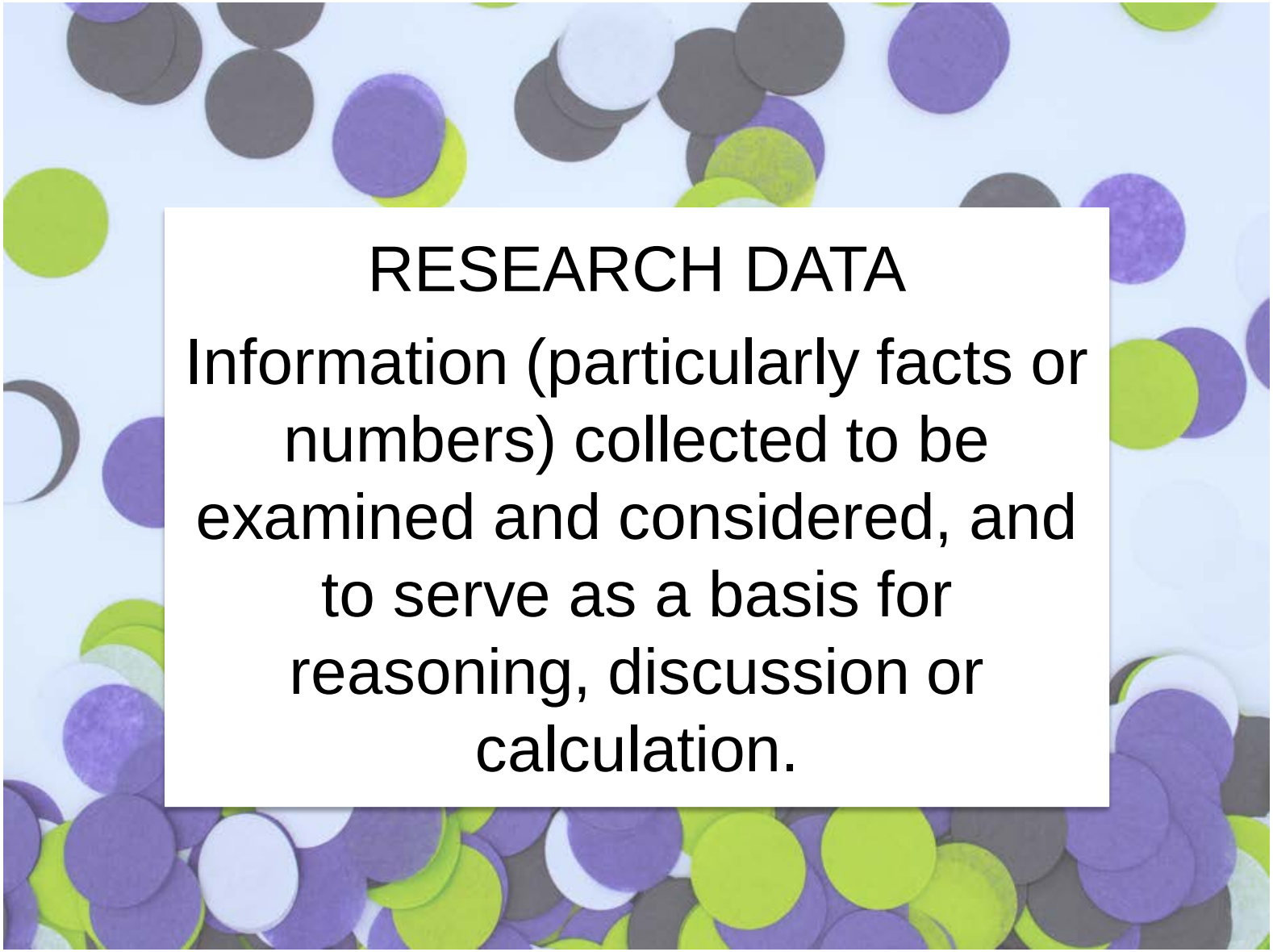
WHERE TO SELF-ARCHIVE?

- Check the Open Access policy of your University
- Use primarily the institutional repository or an established, non-commercial archive on your field
- Use social networking sites cautiously for self-archiving
- Check publisher policies re social networking sites!

TOOLS

- Think Check Submit – tool for choosing journal
- Sherpa/RoMEO – publisher's copyright and archiving policies
- How Can I Share It – tool for checking sharing options

DATA MANAGEMENT



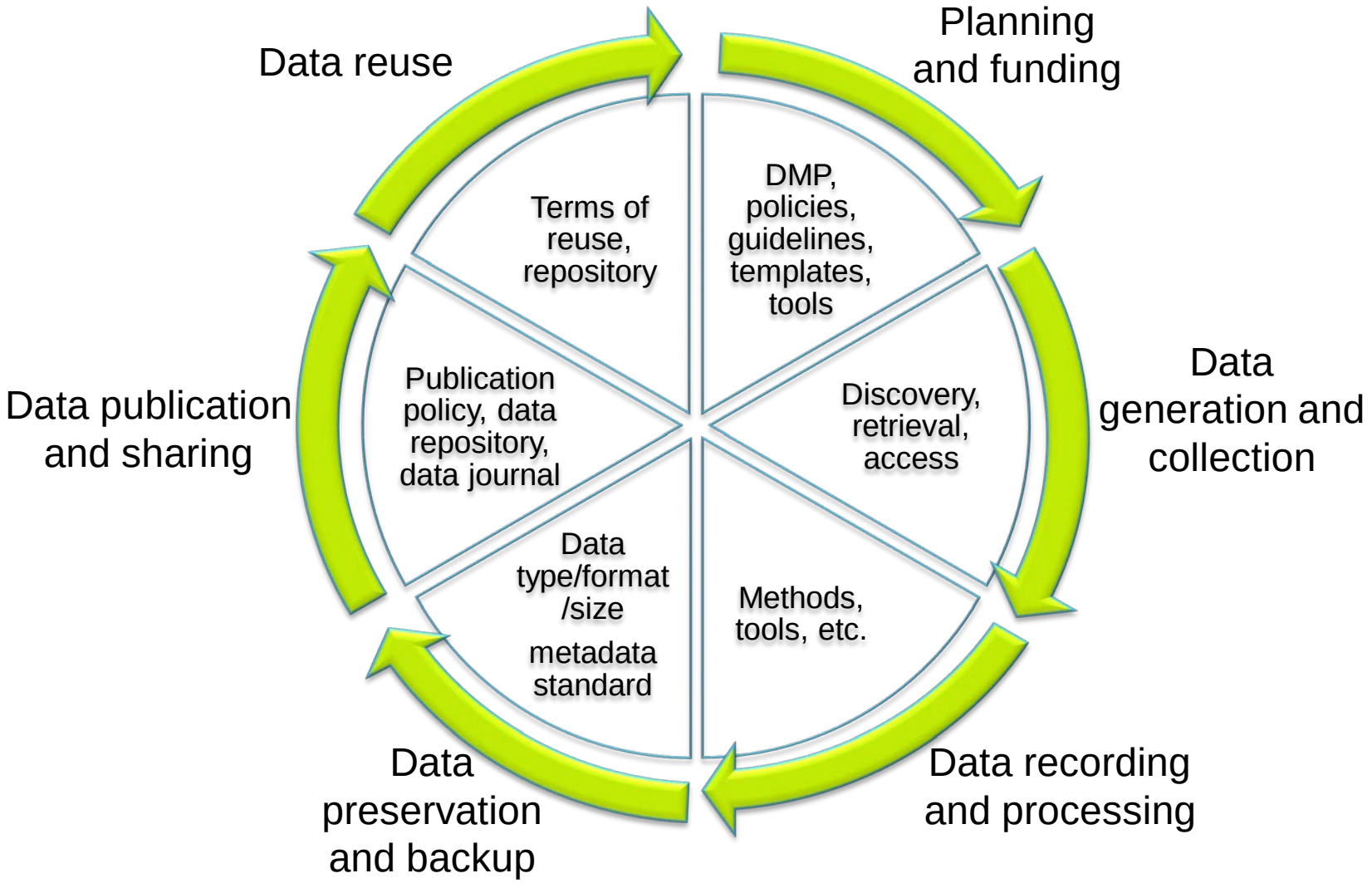
Horizon2020 definition

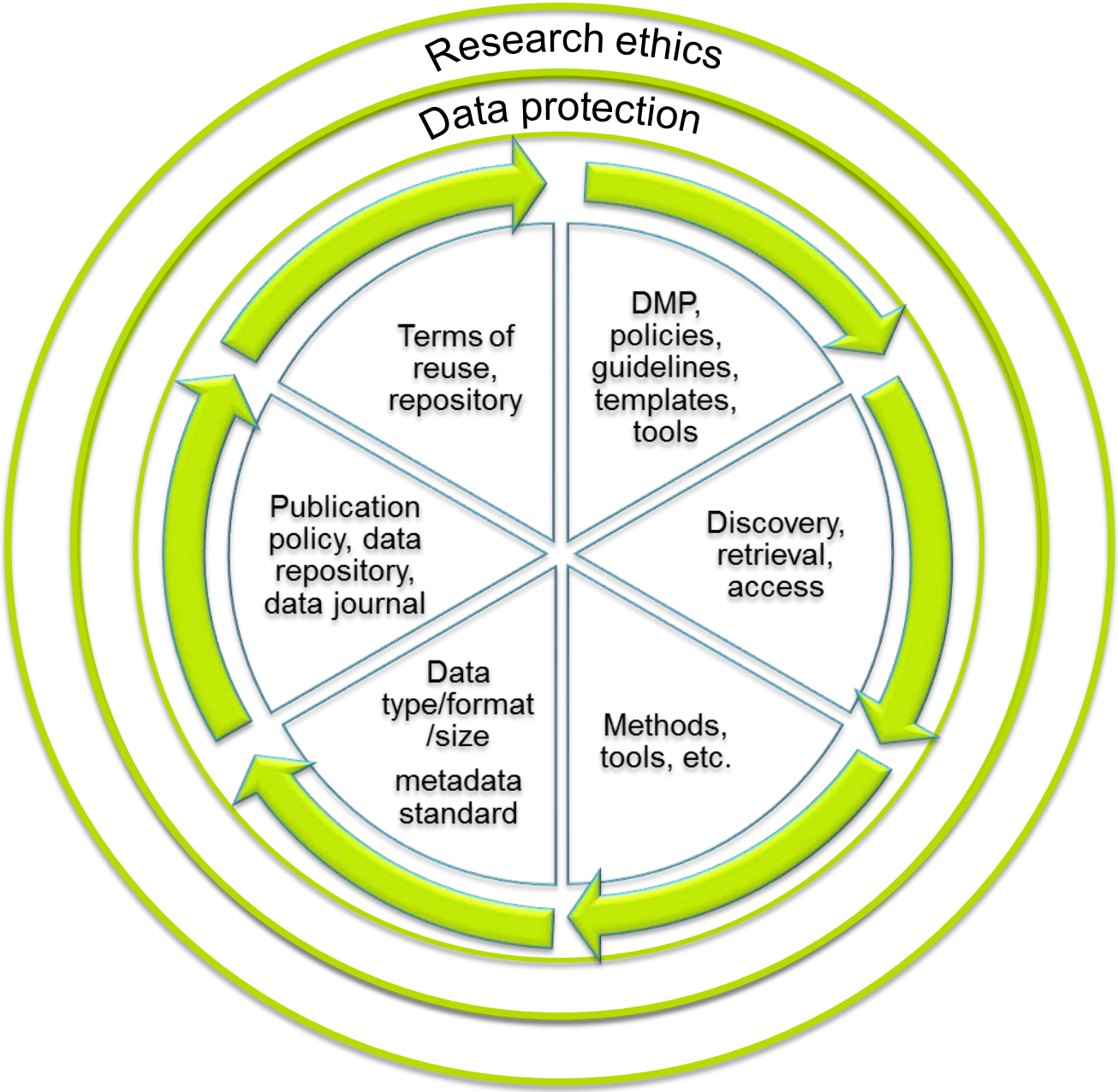
MANAGING RESEARCH DATA = DATA MANAGEMENT

- Data and related metadata are created, preserved, documented and organised systematically
- Data lifecycle and terms of use have been planned and agreed upon

- Responsible conduct of research has been followed in managing data
- Data is not compromised at any level
- It is possible to reuse the data

RESEARCH DATA LIFECYCLE



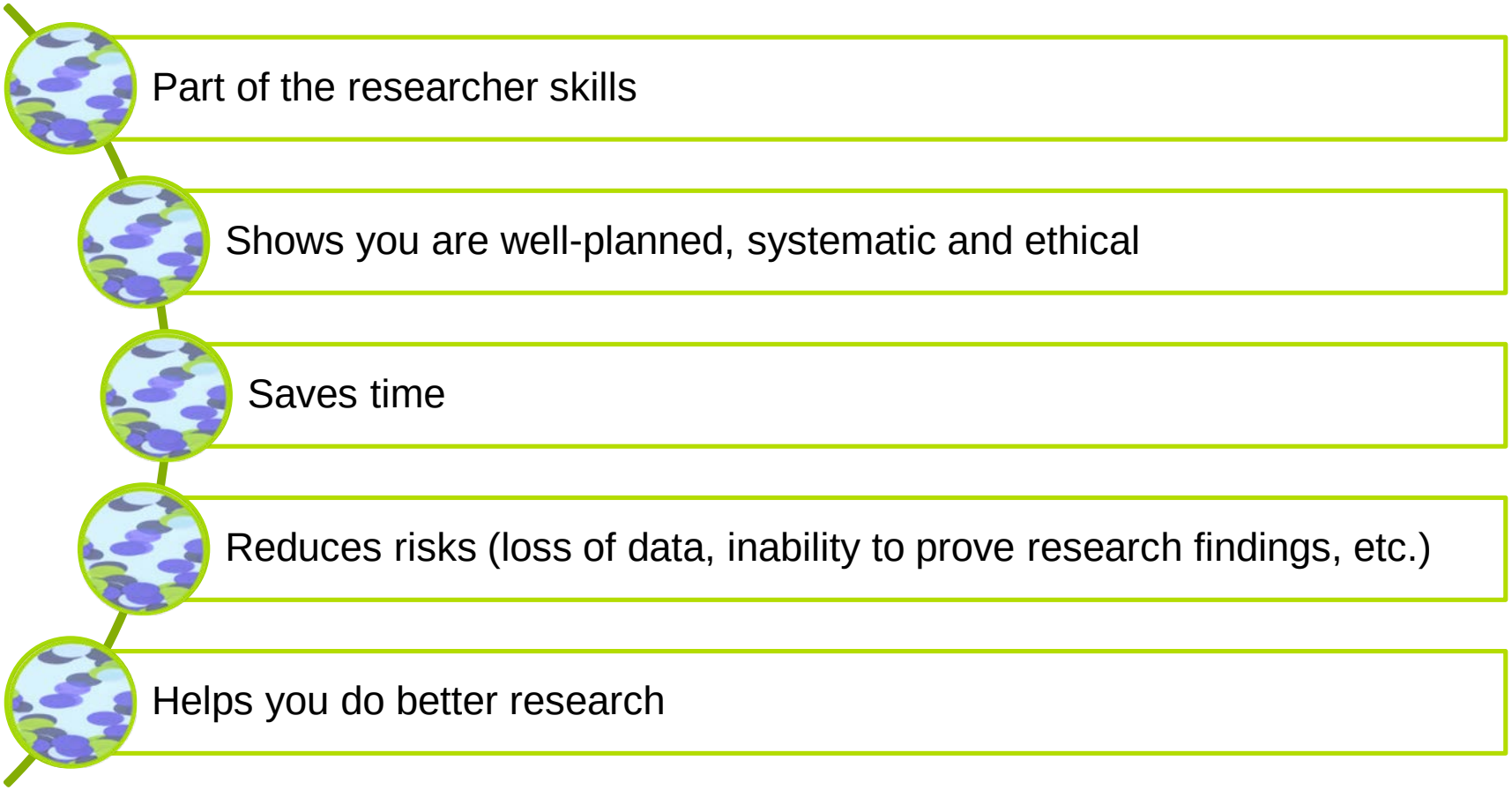

Research ethics
Data protection
Terms of reuse, repository
DMP, policies, guidelines, templates, tools
Discovery, retrieval, access
Methods, tools, etc.
Data type/format /size
metadata standard
Publication policy, data repository, data journal

DATA – THE FOUNDATION OF RESEARCH RESULTS

"Research data are the evidence that underpins the answer to the research question, and can be used to validate findings regardless of its form (e.g. print, digital, or physical)."

Concordat on Open Research Data, published on 28 July 2016
https://www.ukri.org/files/legacy/documents/concordatonopenresearchdata-pdf/

REASONS TO MANAGE YOUR DATA



OPEN DATA

WHAT DOES OPEN ACCESS TO RESEARCH DATA MEAN?

Free access to research data (by anyone)

Data and metadata are deposited onto repositories in format that anyone can find, access and reuse

Horizon2020: The right to access and reuse digital research data under the terms and conditions set out in the Grant Agreement.

PUBLICATIONS AND DATA

THE FUTURE?

RDA Plenary Cartoons by Auke Herrema:https://rd-alliance.org/plenary-meetings/fourth-plenary/plenary-cartoons.html

OPEN ACCESS TO RESEARCH DATA (HORIZON 2020)

"Good research data management is not a goal in itself, but rather the key conduit leading to knowledge discovery and innovation, and to subsequent data and knowledge integration and reuse."

"As open as possible, as closed as necessary"

Encouraging sound data management as an essential part of research best practice.

WHY OPEN DATA?

Access to data is at the heart of research integrity.	Analysis and results can be evaluated and verified if data (and methods) have been opened
Access to open and reusable data has transformed research.	Many research domains rely on the open availability of data from multiple sources in order to operate effectively.
There are broader economic and societal benefits to open research.	The Open and FAIR principles allow data to be used for innovation beyond academia.

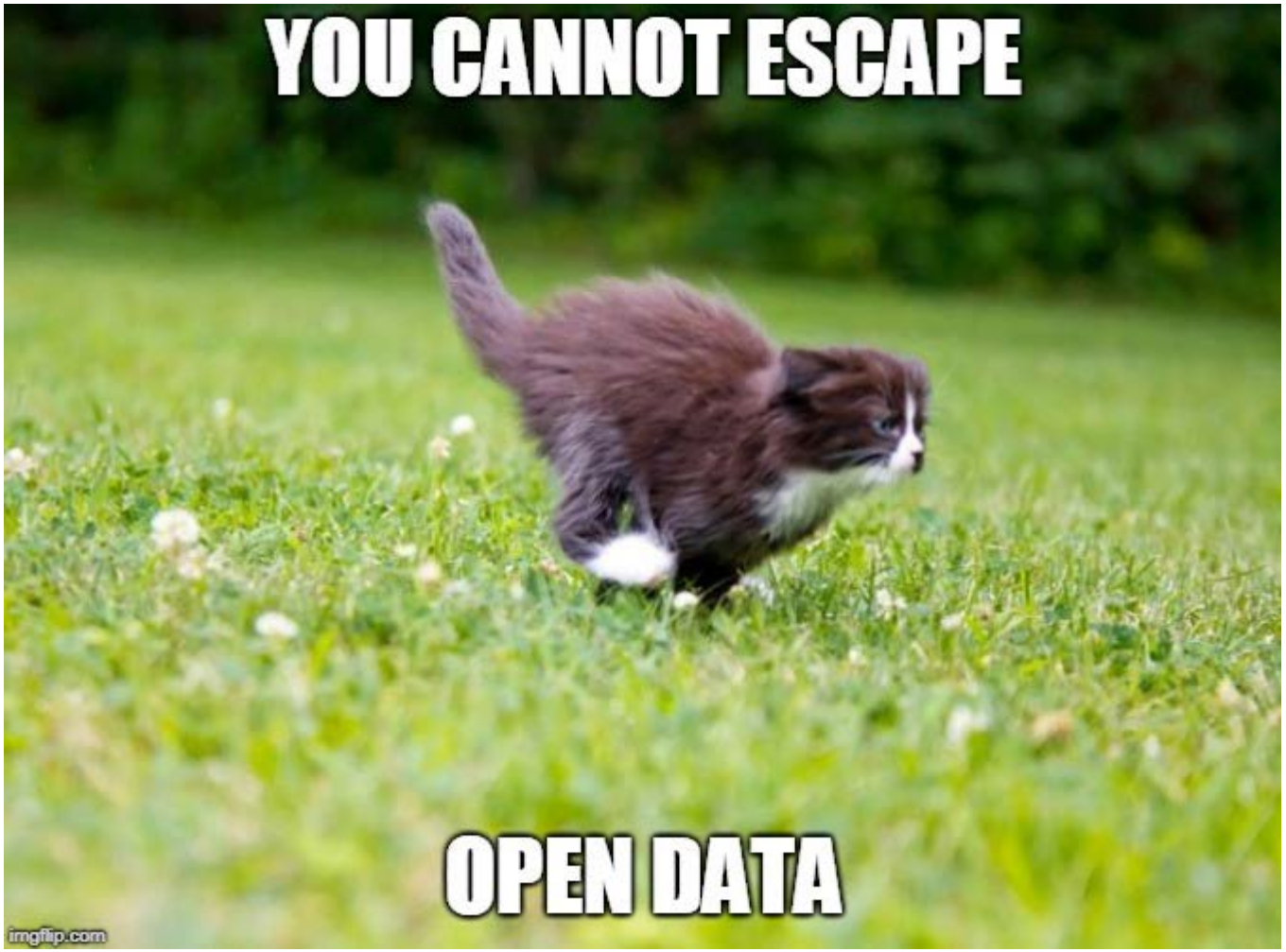
Source: Hodson, Simon, Jones, Sarah, Collins, Sandra, Genova, Françoise, Harrower, Natalie, Laaksonen, Leif, … Wittenburg, Peter. (2018). Turning FAIR data into reality: interim report from the European Commission Expert Group on FAIR data (Version Interim draft).
http://doi.org/10.5281/zenodo.1285272

WHY OPEN DATA – THE RESEARCHER PERSPECTIVE

Ensures preservation and accessibility of data in the future	Facilitates later reuse	Increases the reproducibility of published findings and the ease with which other researchers can use, extend, and cite that work -> more citations
Signals that researchers value transparency and have confidence in their own research	Makes it easier for researchers to connect with one another by increasing the discoverability and visibility of one's work	Can lead to new project and employment opportunities

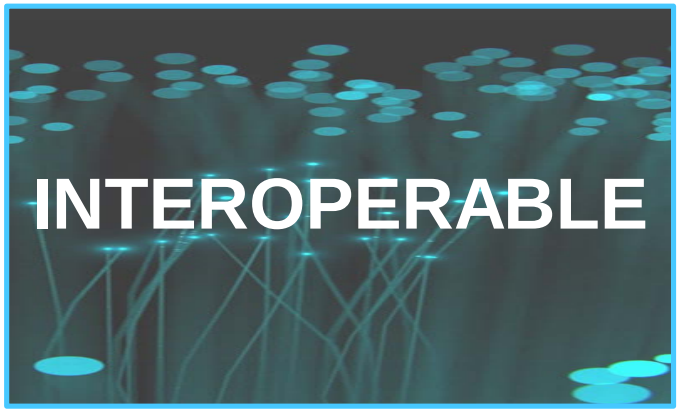
Point of View: How open science helps researchers succeed

eLife 2016;5:e16800 DOI: 10.7554/ELIFE.16800

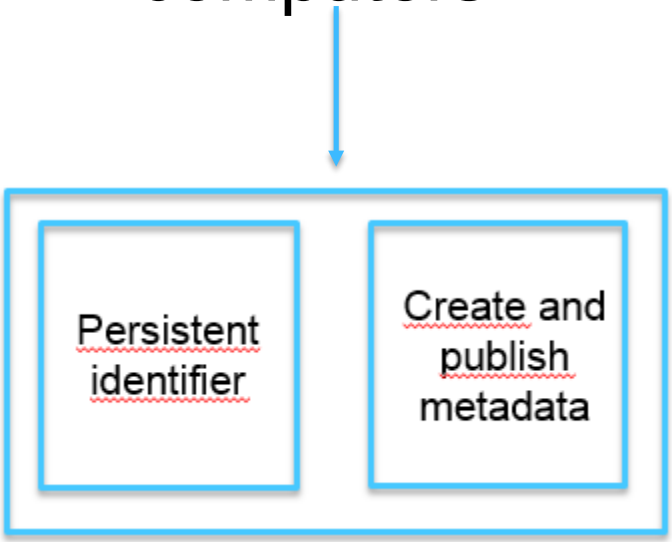


Created with Meme Generator: https://imgflip.com/memegenerator

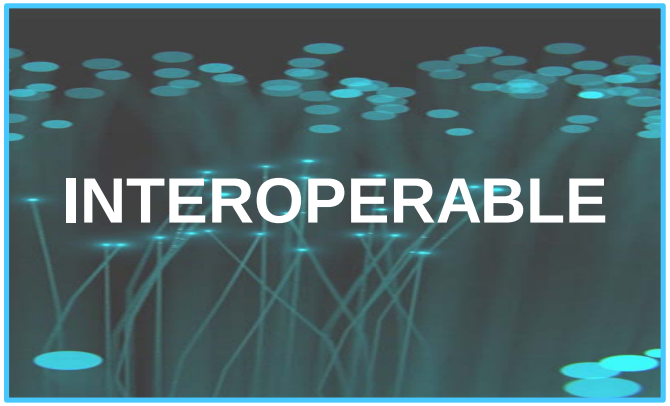
FAIR DATA



Data and metadata are easy to find by both humans and computers



How the data and the metadata can be accessed (incl. authentication and authorization)

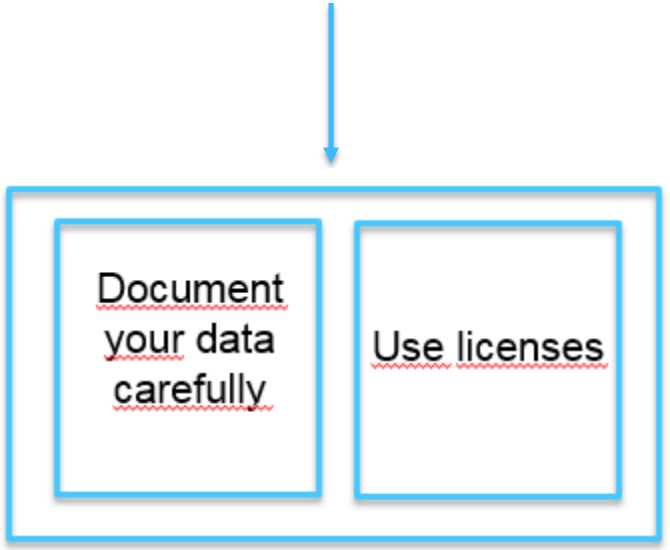


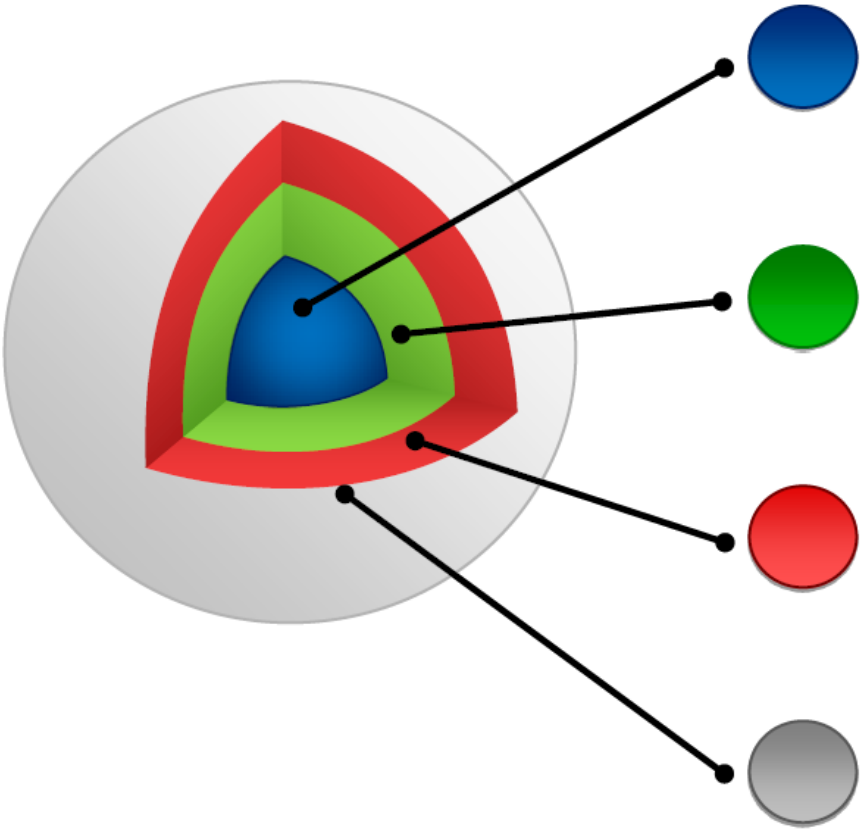
Data need to be integrated with other data + to interoperate with applications/ workflows

Vocabularies, ontologies, thesauri, metadata models

Linking and citing data

Data and metadata should be well-described in order to be replicated





DATA

The core bits

At its most basic level, data is a bitstream or binary sequence. For data to have meaning and to be FAIR, it needs to be represented in standard formats and be accompanied by Persistent Identifiers (PIDs), metadata and code. These layers of meaning enrich the data and enable reuse.

IDENTIFIERS

Persistent and unique (PIDs)

Data should be assigned a unique and persistent identifier such as a DOI or URN. This enables stable links to the object and supports citation and reuse to be tracked. Identifiers should also be applied to other related concepts such as the data authors (ORCIDs), projects (RAIDs), funders and associated research resources (RRIDs).

STANDARDS & CODE

Open, documented formats

Data should be represented in common and ideally open file formats. This enables others to reuse the data as the format is in widespread use and software is available to read the files. Open and well-documented formats are easier to preserve. Data also need to be accompanied by the code use to process and analyse the data.

METADATA

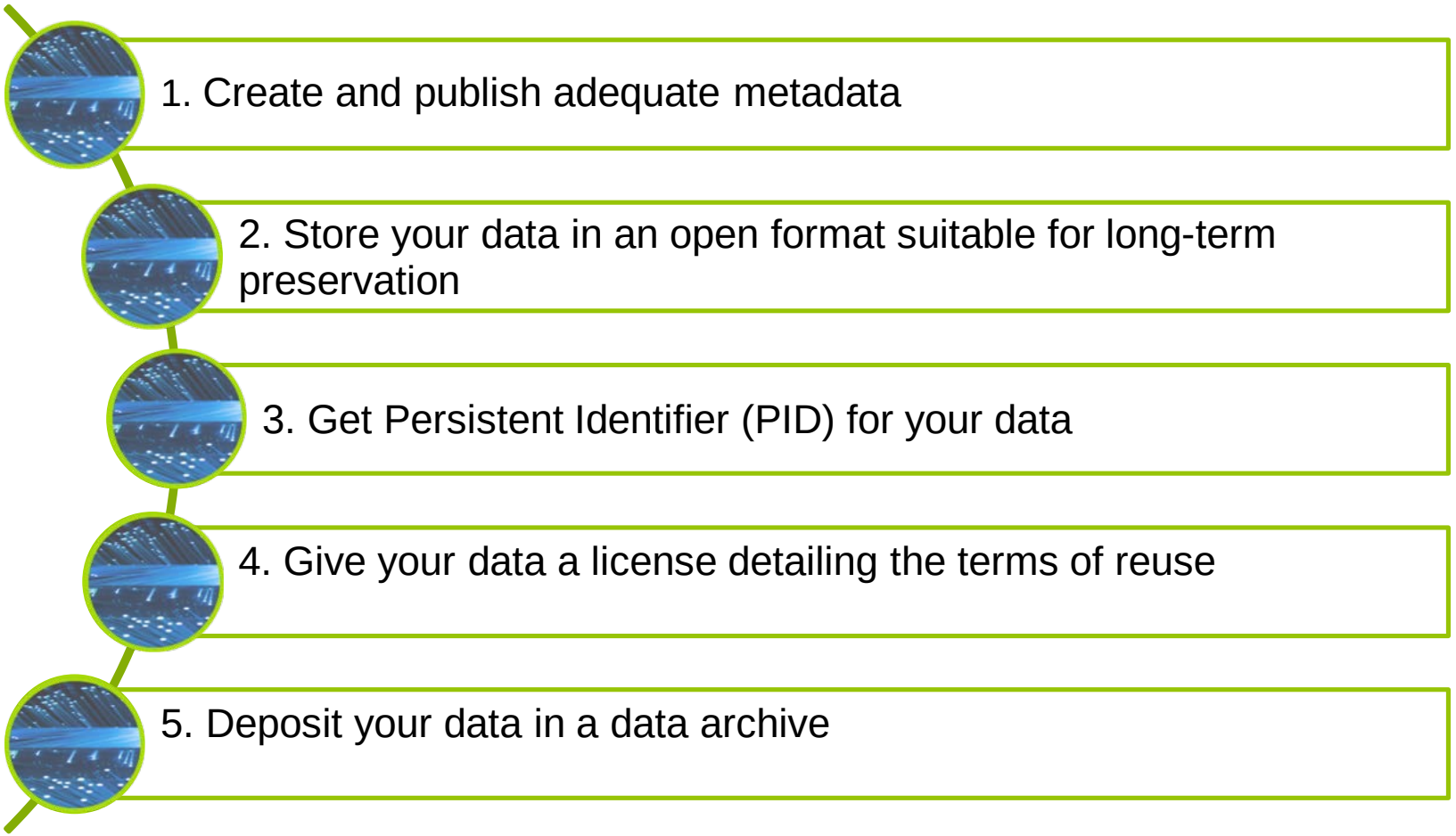
Contextual documentation

In order for data to be assessable and reusable, it should be accompanied by sufficient metadata and documentation. Basic metadata will enable data discovery, but much richer information and provenance is required to understand how, why, when and by whom the data were created. To enable the broadest reuse, data should be accompanied by a 'plurality of relevant attributes' and a clear and accessible data usage license.

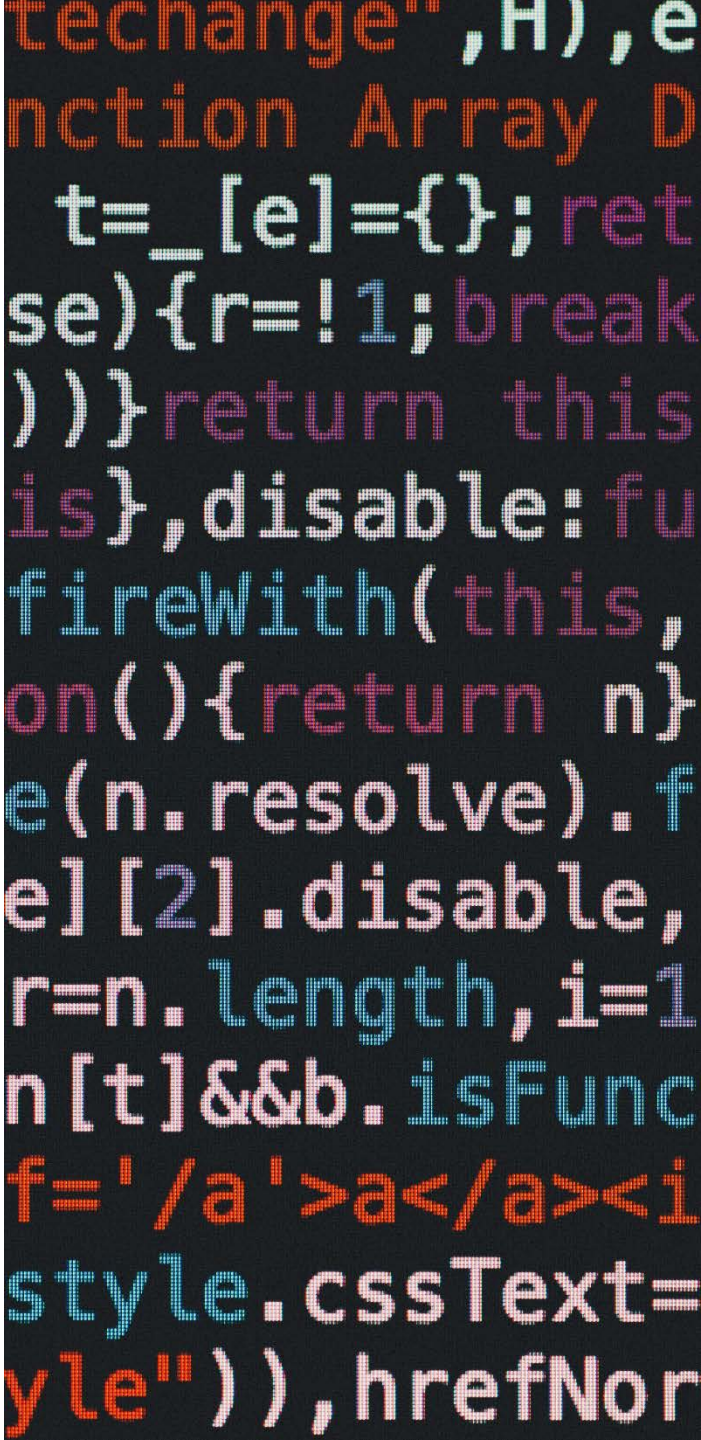
Figure 6: A model for FAIR Data Objects, noting the elements that need to be in place for data to be Findable, Accessible, Interoperable and Reusable.

Source: Hodson, Simon, Jones, Sarah, Collins, Sandra, Genova, Françoise, Harrower, Natalie, Laaksonen, Leif, … Wittenburg, Peter. (2018). Turning FAIR data into reality: interim report from the European Commission Expert Group on FAIR data (Version Interim draft). http://doi.org/10.5281/zenodo.1285272

HOW TO MAKE YOUR DATA FAIR



ALTMETRICS


techange",H),e
nction Array D
t=_[e]={};ret
se){r=!1;break
))}return this
is},disable:fu
fireWith(this,
on(){return n}
e(n.resolve).f
e][2].disable,
r=n.length,i=1
n[t]&&b.isFunc
f='/a'>a<i
style.cssText=
yle")),hrefNor

ALTMETRICS

- Alternative metrics or ”article level metrics”
- Measures the online attention and impact of articles and other research outputs (datasets, code, video, presentations…)
- Data collected: clicks, downloads, favourites, readers, watchers, blog posts, tweets, shares, reviews, citations…

WHAT CAN BE MEASURED?

- Immediate impact
- Opinions of a wider audience
- Impact of non-traditional research outputs
- Possible research collaborators and partners

DISADVANTAGES OF ALTMETRICS

- Lack of standards
- Data can be manipulated
- Altmetric data is often not available for older publications
- Requires persistent identifiers (DOI, URN)

ALTMETRICS TOOLS

- PlumX (integrated in Scopus database)
- Altmetric.com (free Altmetric Bookmarklet for researchers)
- ImpactStory (free service)

ORCID ID

ORCID iD?

- Open researcher and contributor identifier
- Required by e.g.

IEEE	IOP	Wiley	PLOS
Frontiers	Springer Nature	Royal Society of Chemistry	American Chemical Society

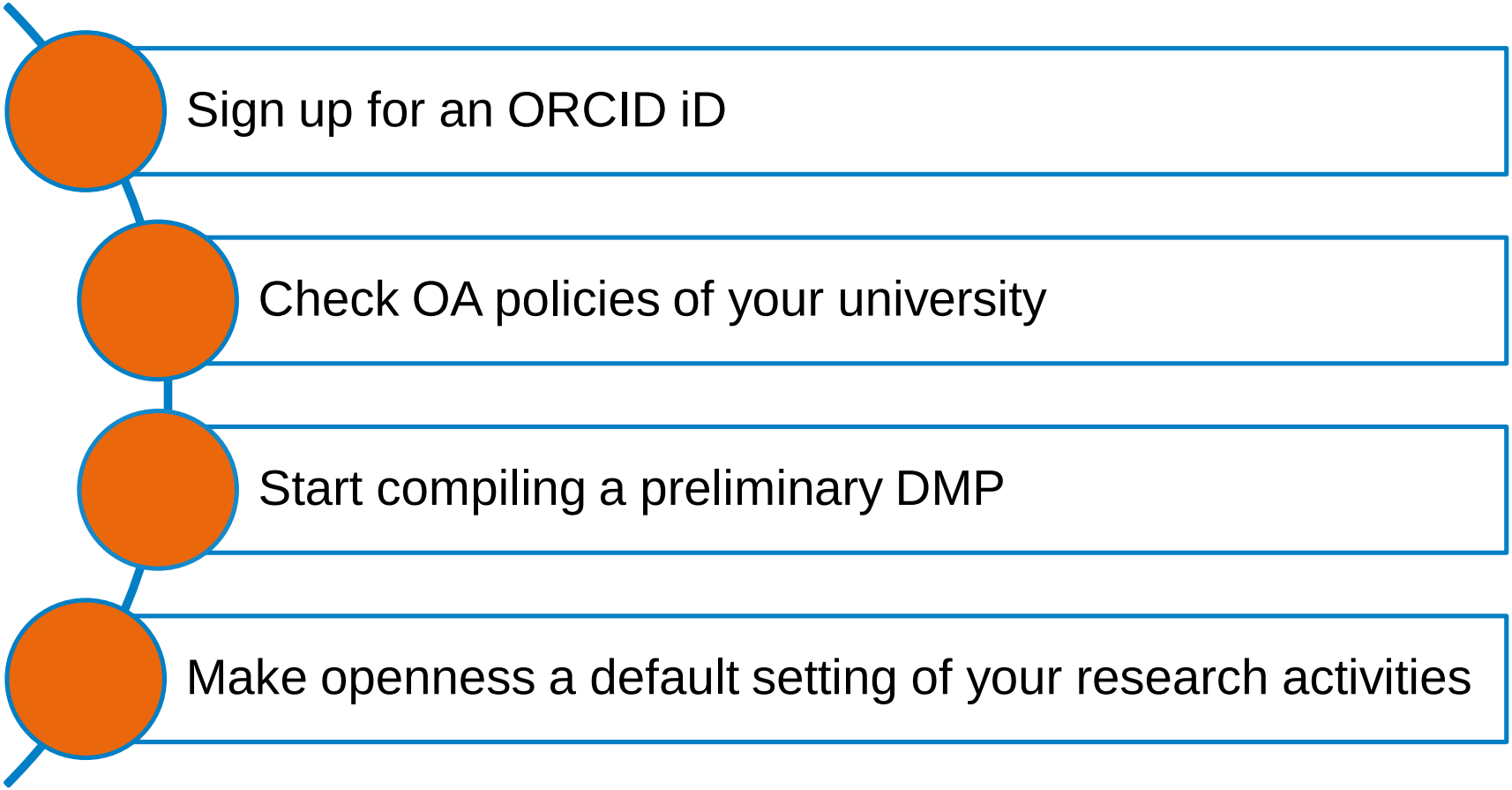
WHY USE ORCID iD?

- It helps to identify you and your research
- It improves your visibility – public profile and list of works can be shared easily
- It’s community-driven and not connected to commercial publishers or services

HOW TO GET STARTED?

- Register at https://orcid.org/
- Add your professional information on your profile
- Use your ORCID iD

SUMMARY: WHAT NEXT?



Thank you!

Images (unless otherwise stated): Unsplash