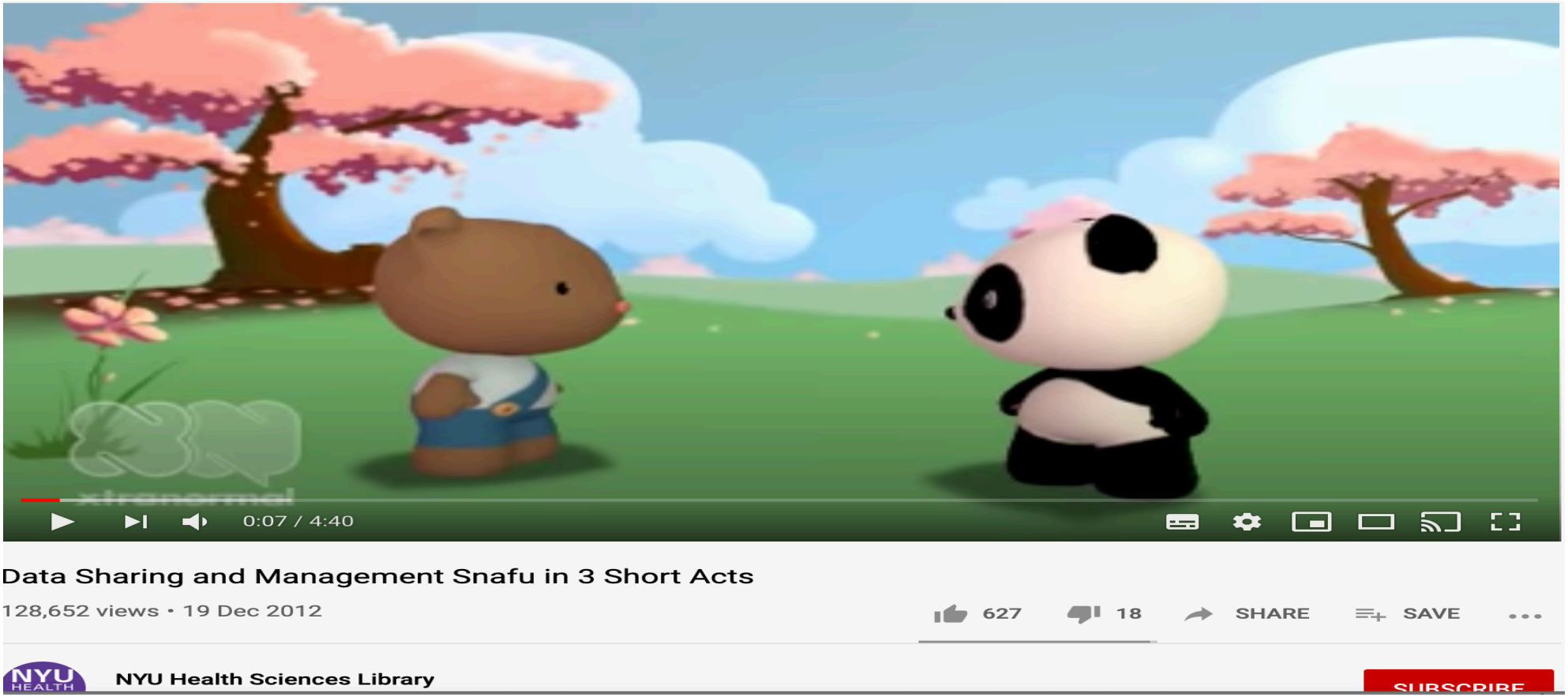
National Initiatives for Open Science in Europe

How to make your data FAIR?

National End-Users Training for North Macedonia

Vladimir Trajkovik, FINKI (FCSE), UKIM

Can you relate to this?



https://www.youtube.com/watch?v=N2zK3sAtr-4

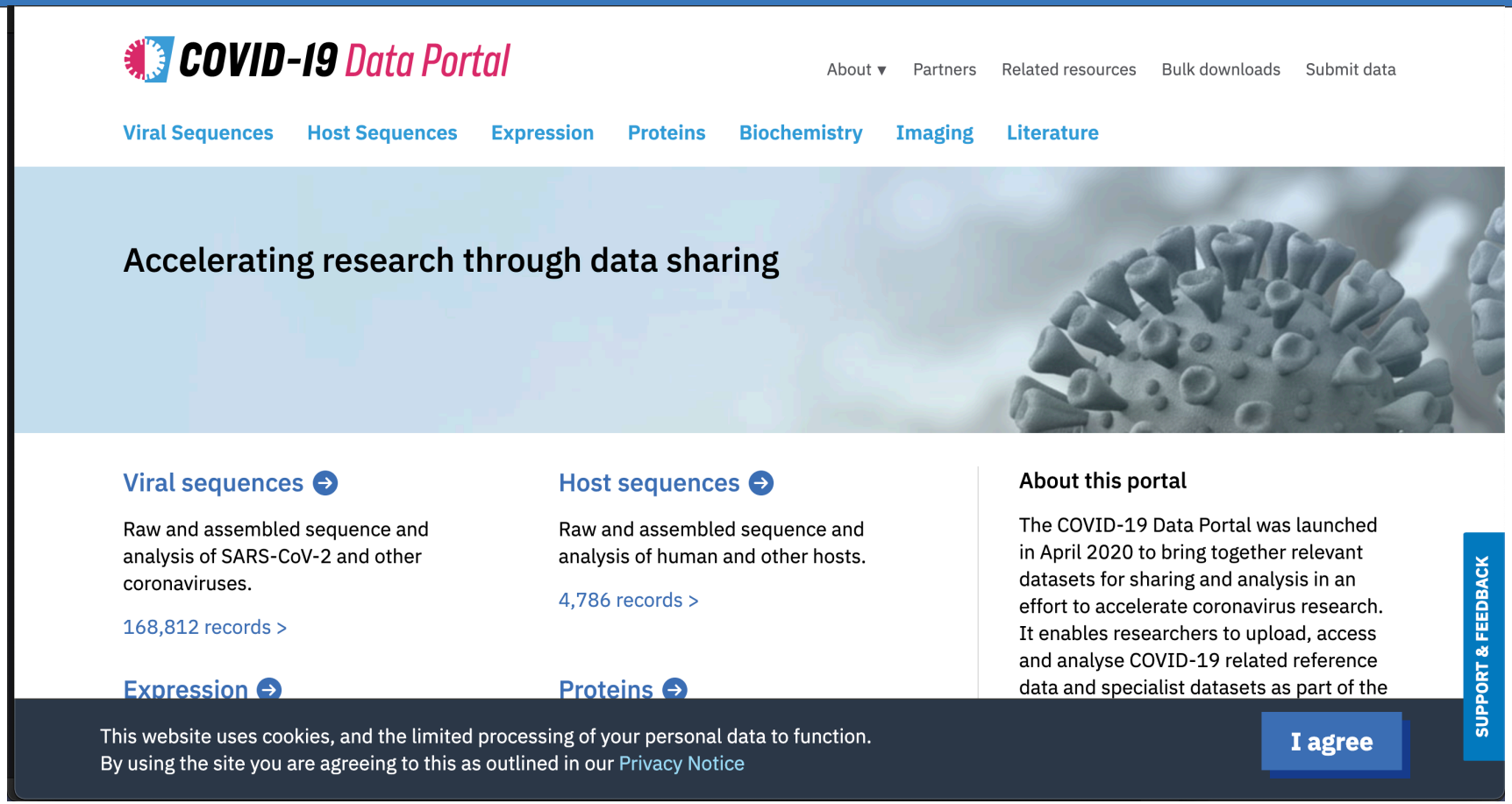
Why you should be FAIR with your data?

The FAIR principles describe how research outputs should be organized so they can be more easily accessed, understood, exchanged and reused.

Major funding bodies, including the European Commission, promote FAIR data to maximize the integrity and impact of their research investment. If you are in receipt of H2020 funding the EC requires a Data Management Plan (DMP) as part of the H2020 data pilot.

The EC supports FAIR data not as a standard but as a framework to follow when designing a Data Management Plan. It has produced a set of Guidelines for FAIR data management.

Than this becomes possible!



https://www.covid19dataportal.org/

FAIR data is about principles not standards!

Providing other researchers with access to your data facilitates knowledge discovery and improves research transparency.

Data can be FAIR but not open.

Open data may not be FAIR.

What is FAIR after all?

'Findable' as discoverable with metadata, identifiable and locatable by means of a standard identification mechanism

'Accessible' as always available and obtainable; even if the data is restricted, the metadata is open

'Interoperable' as both syntactically parse-able and semantically understandable, allowing data exchange and reuse between researchers, institutions, organizations or countries

'Reusable' as sufficiently described and shared with the least restrictive licenses, allowing the widest reuse possible and the least cumbersome integration with other data sources

Findable

Has a persistent identifier (PID); DOI is example of PID.

When depositing your data in a repository, make sure you select a repository that assigns a persistent identifier (for example Zenodo).

Has rich metadata.

Follow standard metadata schemes, general ones such as Dublin Core, or discipline specific (Consult the DCC metadata directory, the RDA Metadata Directory and a portal of data standards at FAIRsharing) .

Is searchable and discoverable online.

Check the how discoverable (indexed) is the repository you plan to use.

Accessible - as open as possible, as closed as necessary

if access is allowed, data should be retrievable without the need for specialized protocols

even if the full content is not made openly available, the data must be as findable as possible

Findable + Accessible = Data Repository

Stores the data safely

Make sure the data is findable

Describes the data appropriately (metadata)

Adds license information

You can deposit data to a general repository (e.g. Zenodo, Harvard Dataverse) or a subject-specific repository (e.g. Dryad).

What about your discipline? Search www.re3data.org for more suitable data repositories. A demonstration of searching for research data repositories using the re3data directory will be given after this presentation.

Interoperable = Common (open) formats and standards + Controlled vocabularies

So it can be integrated with other data, applications and workflows.

Try NOT to create data with proprietary software

Try to use community agreed schemas, keywords, thesauri or ontologies where possible

Reusable = well-documented + clear licence information

README file to help ensure that your data can be correctly interpreted and reanalyzed:

- for each filename, a short description of what data it includes, optionally describing the relationship to the tables, figures, or sections within the accompanying publication;
- for tabular data: definitions of column headings and row labels; data codes (including missing data); and measurement units;
- any data processing steps, especially if not described in the publication, that may affect interpretation of results;
- a description of what associated datasets are stored elsewhere, if applicable;
- whom to contact with questions.

Reusable = well-documented + clear licence information

Data should have a **clear license** to govern the terms of its reuse.

The Open Access guidelines under Horizon 2020 recommend CC-0 or CC-BY as a straightforward and effective way to make it possible for others to mine, exploit and reproduce the data.

EUDAT provides a wizard to help you choose an appropriate license

So, how FAIR are your data?

Jones, S. & Grootveld, M. (2017, November).

How FAIR are your data? (CC-BY)

Zenodo. *http://doi.org/10.5281/zenodo.1065991*

How FAIR are your data? (1/2)

Findable

It should be possible for others to discover your data. Rich metadata should be available online in a searchable resource, and the data should be assigned a persistent identifier.

- ☐ A persistent identifier is assigned to your data
- ☐ There are rich metadata, describing your data
- ☐ The metadata are online in a searchable resource e.g. a catalogue or data repository
- ☐ The metadata record specifies the persistent identifier

Accessible

It should be possible for humans and machines to gain access to your data, under specific conditions or restrictions where appropriate. FAIR does not mean that data need to be open! There should be metadata, even if the data aren't accessible.

- ☐ Following the persistent ID will take you to the data or associated metadata
- ☐ The protocol by which data can be retrieved follows recognised standards e.g. http
- ☐ The access procedure includes authentication and authorisation steps, if necessary
- ☐ Metadata are accessible, wherever possible, even if the data aren't

How FAIR are your data? (2/2)

Interoperable

Data and metadata should conform to recognised formats and standards to allow them to be combined and exchanged.

- ☐ Data is provided in commonly understood and preferably open formats
- ☐ The metadata provided follows relevant standards
- ☐ Controlled vocabularies, keywords, thesauri or ontologies are used where possible
- ☐ Qualified references and links are provided to other related data

Reusable

Lots of documentation is needed to support data interpretation and reuse. The data should conform to community norms and be clearly licensed so others know what kinds of reuse are permitted.

- ☐ The data are accurate and well described with many relevant attributes
- ☐ The data have a clear and accessible data usage license
- ☐ It is clear how, why and by whom the data have been created and processed
- ☐ The data and metadata meet relevant domain standards

FAIR data maturity model indicators

F1	RDA-F1-01M	Metadata is identified by a persistent identifier	●●●	Essential
F1	RDA-F1-01D	Data is identified by a persistent identifier	●●●	Essential
F1	RDA-F1-02M	Metadata is identified by a globally unique identifier	●●●	Essential
F1	RDA-F1-02D	Data is identified by a globally unique identifier	●●●	Essential
F2	RDA-F2-01M	Rich metadata is provided to allow discovery	●●●	Essential
F3	RDA-F3-01M	Metadata includes the identifier for the data	●●●	Essential
F4	RDA-F4-01M	Metadata is offered in such a way that it can be harvested and indexed	●●●	Essential

https://www.rd-alliance.org/group/fair-data-maturity-model-wg/outcomes/fair-data-maturity-model-specification-and-guidelines-0

FAIR data maturity model indicators

A1	RDA-A1-01M	Metadata contains information to enable the user to get access to the data	●●	Important
A1	RDA-A1-02M	Metadata can be accessed manually (i.e. with human intervention)	●●●	Essential
A1	RDA-A1-02D	Data can be accessed manually (i.e. with human intervention)	●●●	Essential
A1	RDA-A1-03M	Metadata identifier resolves to a metadata record	●●●	Essential
A1	RDA-A1-03D	Data identifier resolves to a digital object	●●●	Essential
A1	RDA-A1-04M	Metadata is accessed through standardised protocol	●●●	Essential
A1	RDA-A1-04D	Data is accessible through standardised protocol	●●●	Essential
A1	RDA-A1-05D	Data can be accessed automatically (i.e. by a computer program)	●●	Important
A1.1	RDA-A1.1-01M	Metadata is accessible through a free access protocol	●●●	Essential
A1.1	RDA-A1.1-01D	Data is accessible through a free access protocol	●●	Important
A1.2	RDA-A1.2-01D	Data is accessible through an access protocol that supports authentication and authorisation	●	Useful
A2	RDA-A2-01M	Metadata is guaranteed to remain available after data is no longer available	●●●	Essential

https://www.rd-alliance.org/group/fair-data-maturity-model-wg/outcomes/fair-data-maturity-model-specification-and-guidelines-0

FAIR data maturity model indicators

I1	RDA-I1-01M	Metadata uses knowledge representation expressed in standardised format	●●	Important
I1	RDA-I1-01D	Data uses knowledge representation expressed in standardised format	●●	Important
I1	RDA-I1-02M	Metadata uses machine-understandable knowledge representation	●●	Important
I1	RDA-I1-02D	Data uses machine-understandable knowledge representation	●●	Important
I2	RDA-I2-01M	Metadata uses FAIR-compliant vocabularies	●●	Important
I2	RDA-I2-01D	Data uses FAIR-compliant vocabularies	●	Useful
I3	RDA-I3-01M	Metadata includes references to other metadata	●●	Important
I3	RDA-I3-01D	Data includes references to other data	●	Useful
I3	RDA-I3-02M	Metadata includes references to other data	●	Useful
I3	RDA-I3-02D	Data includes qualified references to other data	●	Useful
I3	RDA-I3-03M	Metadata includes qualified references to other metadata	●●	Important
I3	RDA-I3-04M	Metadata include qualified references to other data	●	Useful

https://www.rd-alliance.org/group/fair-data-maturity-model-wg/outcomes/fair-data-maturity-model-specification-and-guidelines-0

FAIR data maturity model indicators

R1	RDA-R1-01M	Plurality of accurate and relevant attributes are provided to allow reuse	●●●	Essential
R1.1	RDA-R1.1-01M	Metadata includes information about the licence under which the data can be reused	●●●	Essential
R1.1	RDA-R1.1-02M	Metadata refers to a standard reuse licence	●●	Important
R1.1	RDA-R1.1-03M	Metadata refers to a machine-understandable reuse licence	●●	Important
R1.2	RDA-R1.2-01M	Metadata includes provenance information according to community-specific standards	●●	Important
R1.2	RDA-R1.2-02M	Metadata includes provenance information according to a cross-community language	●	Useful
R1.3	RDA-R1.3-01M	Metadata complies with a community standard	●●●	Essential
R1.3	RDA-R1.3-01D	Data complies with a community standard	●●●	Essential
R1.3	RDA-R1.3-02M	Metadata is expressed in compliance with a machine-understandable community standard	●●●	Essential
R1.3	RDA-R1.3-02D	Data is expressed in compliance with a machine-understandable community standard	●●	Important

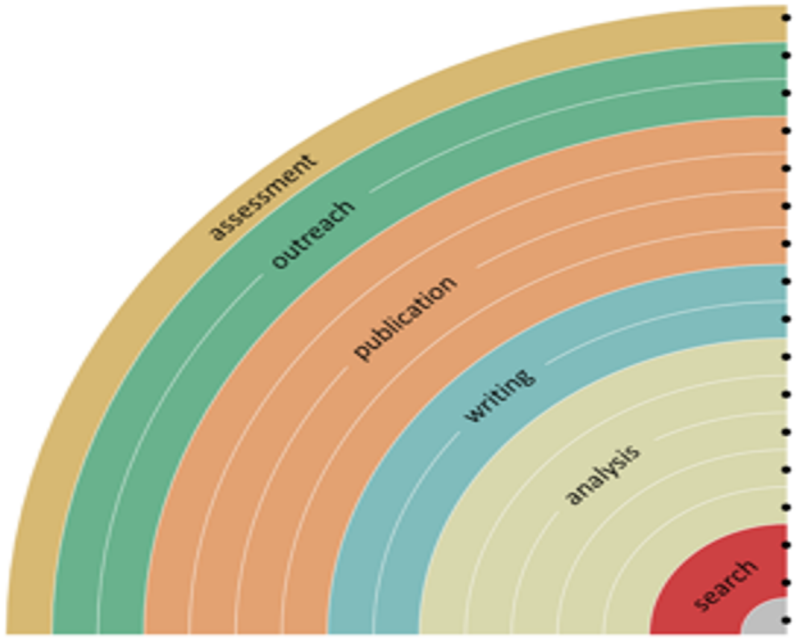
https://www.rd-alliance.org/group/fair-data-maturity-model-wg/outcomes/fair-data-maturity-model-specification-and-guidelines-0

More resources

- European Commission. Action Plan for FAIR data recommendations.
- The EC expert group on FAIR data
- EC/H2020 - Guidelines on FAIR Data Management in Horizon 2020
- Hodson, S. (2018). Making FAIR data a reality… and the challenges of interoperability and reusability. Open Science Conference 2018.
- https://www.rd-alliance.org/group/fair-data-maturity-model-wg/outcomes/fair-data-maturity-model-specification-and-guidelines-0

Is this all that I need?
FAIR data is only a part of open science

You can make your workflow more open by ...



- adding alternative evaluation, e.g. with altmetrics
- communicating through social media, e.g. Twitter
- sharing posters & presentations, e.g. at FigShare
- using open licenses, e.g. CC0 or CC-BY
- publishing open access, 'green' or 'gold'
- using open peer review, e.g. at journals or PubPeer
- sharing preprints, e.g. at OSF, arXiv or bioRxiv
- using actionable formats, e.g. with Jupyter or CoCalc
- open XML-drafting, e.g. at Overleaf or Authorea
- sharing protocols & workfl., e.g. at Protocols.io
- sharing notebooks, e.g. at OpenNotebookScience
- sharing code, e.g. at GitHub with GNU/MIT license
- sharing data, e.g. at Dryad, Zenodo or Dataverse
- pre-registering, e.g. at OSF or AsPredicted
- commenting openly, e.g. with Hypothes.is
- using shared reference libraries, e.g. with Zotero
- sharing (grant) proposals, e.g. at RIO

DOI: 10.5281/zenodo.1147025

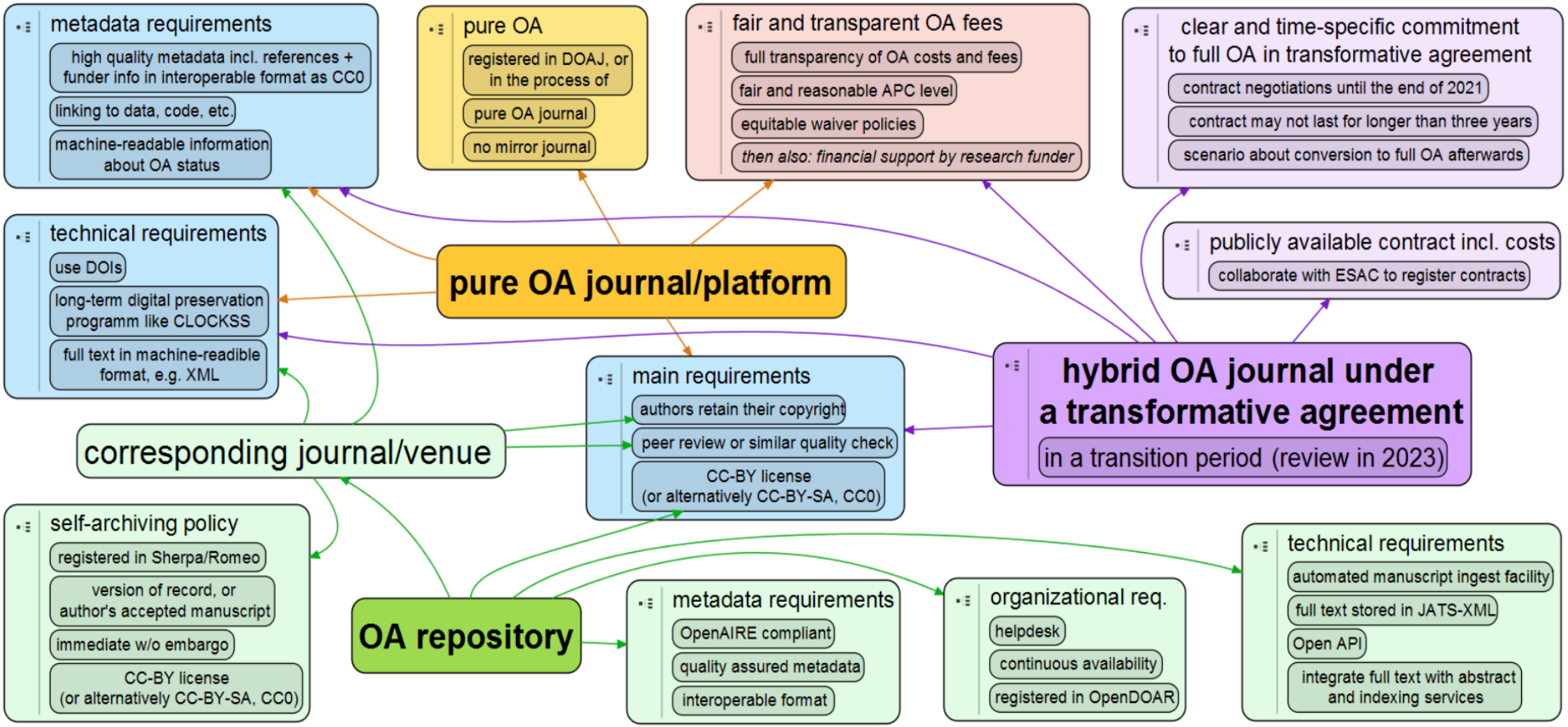
Plan S(hock)

“Plan S” is an initiative for open-access science publishing launched in 2018 by "cOAlition S", a consortium of national research agencies and funders from twelve European countries.

The plan requires scientists and researchers who benefit from state-funded research organizations and institutions to publish their work in open repositories or in journals that are available to all by 2021.

This will come into effect in funding calls issued from 1 January 2021.

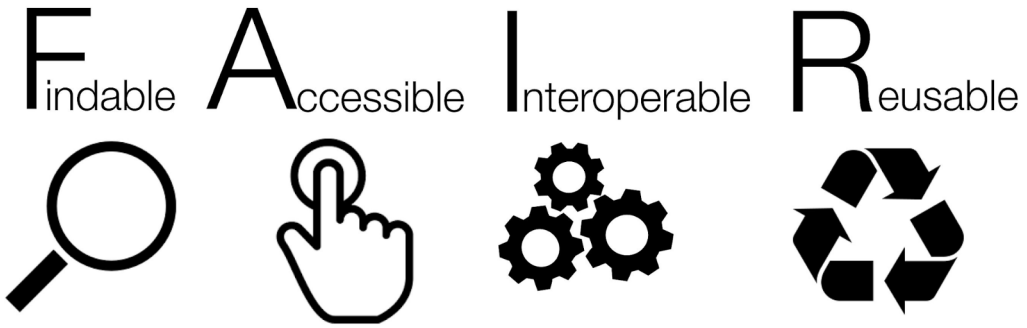
Requirements for Plan S compliance in one page



Legend: blue = requirement for all; red = requirement for pure + hybrid OA; orange = requirement for pure OA; green = requirement for OA repository; purple = requirement for hybrid OA

Based on the Guidance on the Implementation of Plan S (2018-11-27) https://www.coalition-s.org/feedback/

Thank you!



https://ni4os.eu

@NI4OS_eu @NI4OS

Source: https://en.wikipedia.org/wiki/FAIR_data#/media/File:FAIR_data_principles.jpg