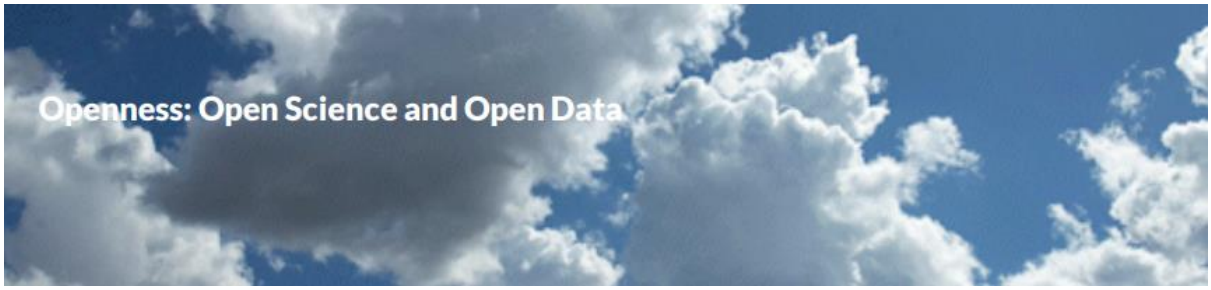


This is a short, simple tutorial that introduces Open Science and Open Data and how this relates to the responsibilities of academic librarians in advising academic staff on how to share their research data with others.

Contents

Open Science

Why should African Librarians care about Open Science?

AfLIA

What is Open Science?

Open science involves opening up the processes of creating, evaluating, sharing, exploring and storing scientific knowledge, practices and perspectives.

UNESCO Recommendation on Open Science

The UNESCO Recommendation on Open Science provides the first internationally agreed definition of Open Science which is defined as an inclusive construct that combines various movements and practices, aiming to:

- make multilingual scientific knowledge openly available, accessible and reusable for everyone,
- increase scientific collaborations and sharing of information for the benefits of science and society, and to
- open the processes of scientific knowledge creation, evaluation and communication to societal actors beyond the traditional scientific community.

All scientific disciplines and aspects of scholarly practices, including basic and applied sciences, natural and social sciences and the humanities, are included.

Open science builds on key pillars: open scientific knowledge, open science infrastructures, science communication, open engagement of societal actors and open dialogue with other knowledge systems.

Watch the UNESCO Recommendation on Open Science video (not a Creative

Commons licence) by visiting this link on YouTube:
https://www.youtube.com/watch?v= I3Wkvx_ZaFo

Activity

Think about the benefit of open science during the creation of the COVID-19 vaccines.

Would the process have been accelerated if the scientific information and research were made freely available to the scientific community?

How could African Researchers have more equitably contributed, both in terms of accessing, generating and collaborating on research?

Did the communities that were involved in the trial studies for vaccines benefit as much as the large pharmaceuticals?

Lisbeth Levey has some thoughts on Open Science and colonialism in Open Knowledge Primer for African Universities Revised and Updated Edition:

Is Open Science open to everyone?

Open science may be open, but the questions are to whom and to what. Scientific research, including research about Africa, is primarily carried out by scholars in the Global North. This is true for all science, but it seems an anomaly to discuss open science when it is not entirely open.

Open Science in English

Publishing, including open access publishing, is equally disproportionately focused on researchers from the Global North, even when the research is about Africa.

There is also a slant towards the North in terms of access and publishing. Most scientific journals are published in English. According to Pisana Ferrar, in a 2021 article, more than 90% of the indexed articles in the natural sciences are now published in English. This fact is a disadvantage for both readers and authors for whom English is a foreign language.

Resources

Understanding Open Science.. UNESCO Open Science Toolkit factsheet. https://unesdoc.unesco.org/ark:/48223/pf0000383323

Open Knowledge Primer for African Universities Revised and Updated Edition. May 2023. Lisbeth Levey. https://www.oerafrica.org/system/files/12591/open-knowledge-primer-june-2023.pdf?file=1&type=node&id=12591&force=1

Open Data

What is open data?

Open data is data that can be freely used, re-used and redistributed by anyone - subject only, at most, to the requirement to attribute and sharealike.

Read the video transcript below which explains, in a nutshell, what Open Data is.

Video transcript: simpleshow foundation. (2016). Open Data – explained in a nutshell (CC BY)

Open Data explained simply:

Every second we create huge amounts of data, some see just a lot of numbers. At the same time that data represents our daily experiences: doctor's visits, commutes, weather conditions and even how many parking tickets have been issued on our street.

Most of the data is stored in ways which make it inaccessible whereas some information means a huge opportunity. With open data we can use these numbers and learn from them together.

Is there any connection between rainfall and doctors appointments or between commute routes and parking tickets? Using open data we have a chance to find out.

Thanks to open data, entrepreneurs who find new ways to develop their businesses. Scientists get a better understanding of our world while citizens and politicians can make more informed decisions like with England's project: where does my money go, showing people where their taxes get spent. |

Anyone can combine open data with private data. Should you discover something interesting or create new data that you want to share, you can do so using tools such as Wikidata.

There are countless benefits to making information more accessible. Some of the data that is free to access today by rights is stored in places that make it in practice almost inaccessible. Open data aims to change this.

With standards for handling, storing and accessing data we can avoid these problems. Furthermore, data owners need to take privacy aspects into account. They have to ensure that all open data has been made anonymous and that no one's privacy is threatened. There has never been more data any time in our history than today, so why not benefit from it together by turning numbers into open data.

FAIR

The FAIR (Findable, Accessible, Interoperable and Reusable) principles refer to three types of entities: data (or any digital object), metadata (information about that digital object), and infrastructure. For instance, principle F4 defines that both metadata and data are registered or indexed in a searchable resource (the infrastructure component).

- **Findable**

The first step in (re)using data is to find them. Metadata and data should be easy to find for both humans and computers. Machine-readable metadata are essential for automatic discovery of datasets and services, so this is an essential component of the FAIRification process.

- F1. (Meta)data are assigned a globally unique and persistent identifier

- F2. Data are described with rich metadata (defined by R1 below)
- F3. Metadata clearly and explicitly include the identifier of the data they describe
- F4. (Meta)data are registered or indexed in a searchable resource

- **Accessible**

Once the user finds the required data, she/he/they need to know how they can be accessed, possibly including authentication and authorisation.

- A1. (Meta)data are retrievable by their identifier using a standardised communications protocol
- A1.1 The protocol is open, free, and universally implementable
- A1.2 The protocol allows for an authentication and authorisation procedure, where necessary
- A2. Metadata are accessible, even when the data are no longer available

- **Interoperable**

The data usually need to be integrated with other data. In addition, the data need to interoperate with applications or workflows for analysis, storage, and processing.

- I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.

- I2. (Meta)data use vocabularies that follow FAIR principles
- I3. (Meta)data include qualified references to other (meta)data

- **Reusable**

The ultimate goal of FAIR is to optimise the reuse of data. To achieve this, metadata and data should be well-described so that they can be replicated and/or combined in different settings.

- R1. (Meta)data are richly described with a plurality of accurate and relevant attributes
- R1.1. (Meta)data are released with a clear and accessible data usage license
- R1.2. (Meta)data are associated with detailed provenance
- R1.3. (Meta)data meet domain-relevant community standards

Activity

How would you communicate the FAIR principles to a staff member wanting to make their data openly available?

Publishing Open Data

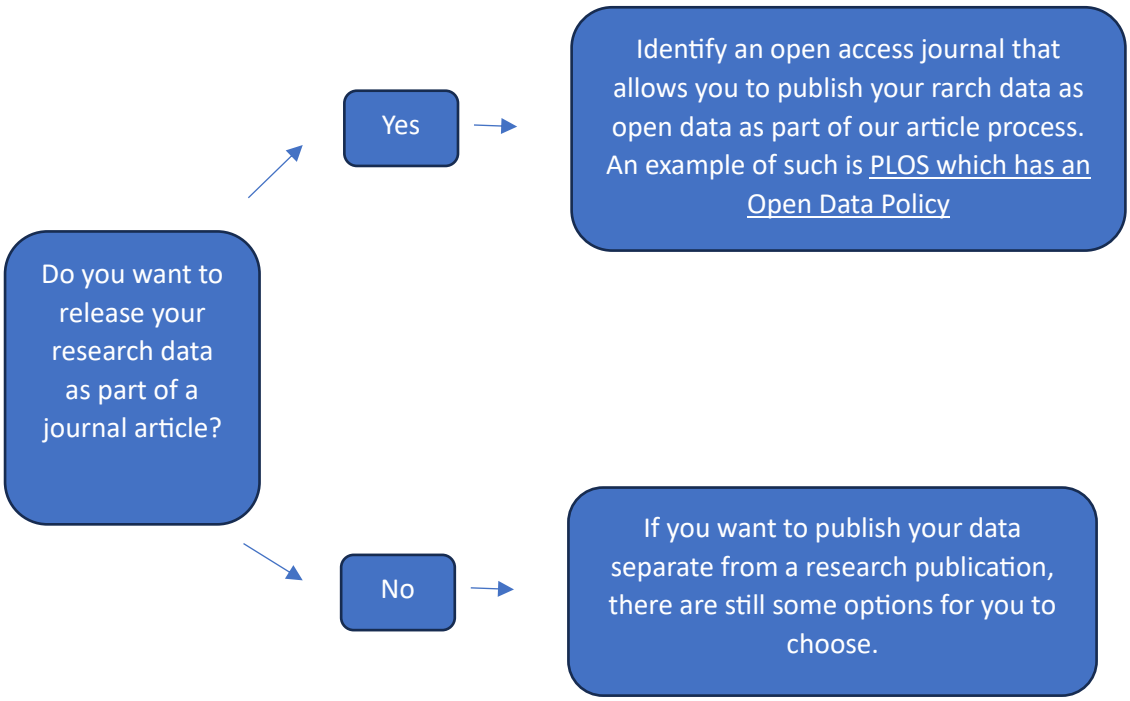
So, you want to make your data open. There are a few options to do this:

1. Release your data as part of your research article in a journal
2. Release your data as a standalone resource
    a. This could be hosted on your institution's Open Access repository
    b. Hosted on a third party open data repository

So, you want to make your data open? Open Data ensures reproducibility, inspires trust, receives credit, makes a contribution and preserves the scientific record.

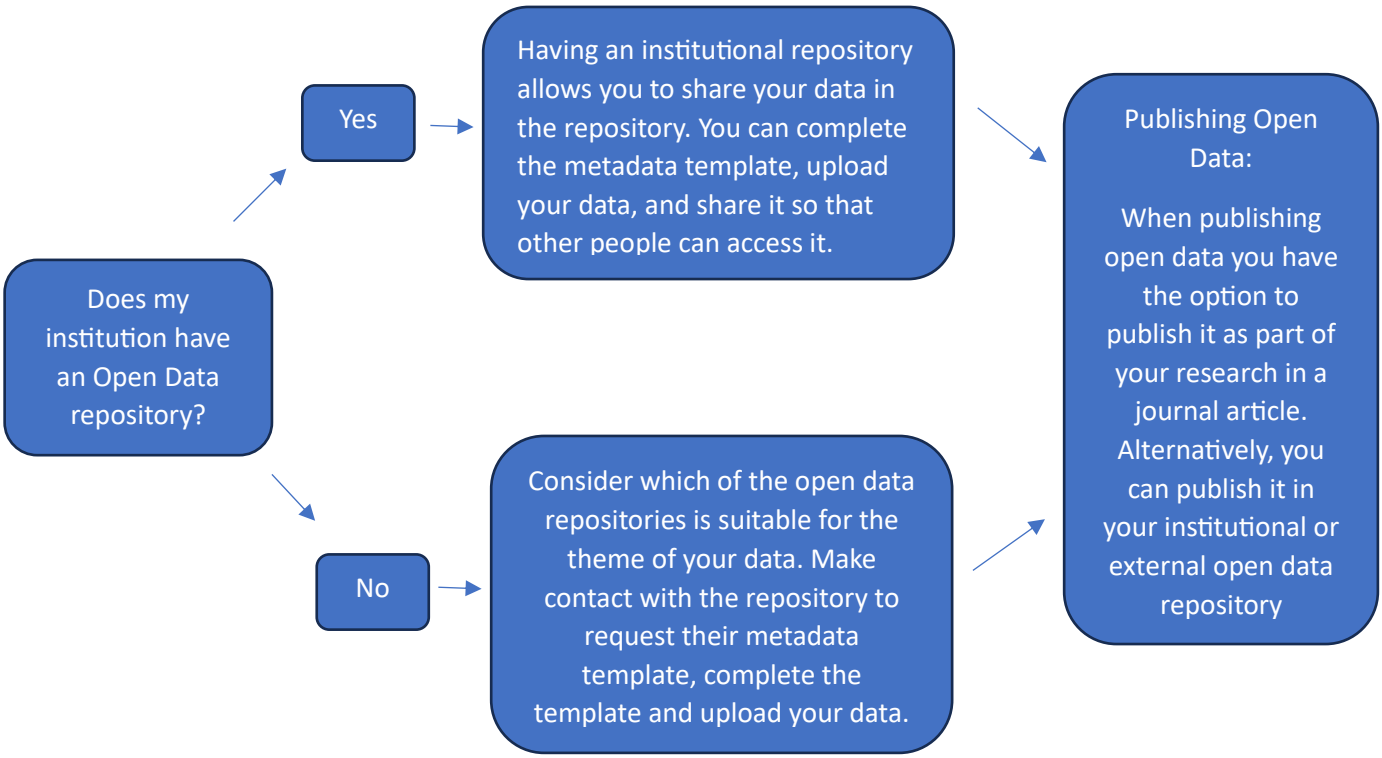
Review the two scenarios detailed below:

Scenario 1



Scenario 2

If your answer to the above question is 'No', continue with this scenario.



Open Data Repositories

There are a few open data repositories for African data that you can access and explore. View any of the ones below to browse the data sets that they have available:

- **Wikidata**
  - Wikidata is a free and open knowledge base that can be read and edited by both humans and machines
  - https://www.wikidata.org/wiki/Wikidata:Main_Page
- **OpenAFRICA**
  - OpenAFRICA aims to be the largest independent repository of open data on the African Continent
  - https://africaopendata.org/
- **Open Cities Africa**
  - Open data today to build resilient and sustainable societies tomorrow
  - https://opencitiesproject.org/

- **World Bank Open Data**
    - The World Bank Open Data is a free and open access to global development data
    - https://data.worldbank.org/
- **Datadot**
    - Datadot is an accessible, trusted and actionable source for a full breadth of the world's health data
    - https://data.who.int/
- **African Development Bank Group** (AfDB) Africa Information Highway
    - Initiative in 2012, The Africa Information Highway (AIH) was created by the AfDB to scale up the collection, management, and dissemination of quality statistics relating to Africa
    - https://dataportal.opendataforafrica.org/data
- **re3data**
    - Registry of Research data and Repositories
    - https://www.re3data.org/

University Open Data Repositories

The University of Cape Town has an open data repository that contains external open data sets that can be used by their own academics and students in their studies.

In addition, they also use the repository to share their own institutional research data.

- University of Cape Town
    - The University of Cape Town has an open data portal called DataFirst
    - https://www.datafirst.uct.ac.za/dataportal

Open Data Repositories

Which is the best solution for your institution? Do you want to publish your data on an external data repository or do you have an institutional repository that can be adapted to store your open data?
There are advantages and disadvantages to where you should publish your open data.

Metadata

When publishing your data you need to think about the metadata fields that you need to complete. There are some international standards for metadata, such as DataCite, Dublin Core, DDI, EML etc. Below is an example of the metadata required for the SAEON Open Data Platform

Wiki https://odpwiki.saeon.ac.za/minimum_metadata_for_the_saeon_odp

# (1)	Property	Description	Obligation	Purpose
1	Identifier	Unique identifier in the provider's environment	Mandatory	Citation
2	Creator(s)	Usually the author of the work	Mandatory	Citation
3	Title	The title of the work	Mandatory	Citation
4	Publisher	The publisher of the work	Mandatory	Citation
5	Publication year	The year of publication	Mandatory	Citation
6	Subject	One or more keywords, which may come from authoritative lists (vocabularies)	Recommended	Discovery
8	Date	The date coverage of the dataset or digital object - in most cases not the same as the year of publication	Recommended	Discovery
10	Resource type	This assists with re-usability and discovery of resource types appropriate to the intended end use	Mandatory	Re-usability
11	Alternate identifier	An alternative identifier, such as a DOI, handle, or ARK. If an identifier is not provided, SAEON will assign a DOI on behalf of the requestor	Optional	Re-usability
12	Related identifier	Critical for version management - references previous versions of the same digital object, amongst other uses	Recommended	Re-usability
16	Rights	This is an important determinant of re-usability.	Optional	Re-usability

		SAEON advocates the use of open licences		
17	Description	The element can be used for both an abstract and a methodology or lineage description	Recommended	Re-usability
18	Geo-Location	Spatial region or named place where the data was gathered or about which the data is focused. This may not be applicable to all datasets, but very useful for discovery in cases where it does apply	Recommended	Discovery
A	Online resource	A link to the location of the digital object in the web		Discovery
B	Original metadata	A link to the original metadata record in the web		Re-usability
C	Supplementary material	A link to supplementary material in the web. This can be a formal; published methodology or protocol		Re-usability
D	Research ethics	A link to an ethics protocol or guideline in the web		Re-usability
E	Download link	A download link in cases where this is different from (A)		Re-usability
F	Grant number	If research is grant funded, this attribute is mandatory to prove deposit (this condition applies to NRF funded research)		Re-usability
7	Contributor(s)	Additional contributions to the work	Recommended	Re-usability
9	Language	In South Africa, the language is almost exclusively English, but the element should be used if it is not	Optional	Re-usability
13	Size	Size of the work - assists with download decisions	Optional	Re-usability
14	Format	Format of the work - assists with discovery and download decisions	Optional	Re-usability

15	Version	The version assigned by the owner of the work. This is not the preferred method of indicating versioning - see Related Identifier above	Optional	Re-usability
19	Funding reference	Information about financial support (funding) for the resource being registered. It is a best practice to supply funding information when financial support has been received	Optional	Re-usability

Think about the metadata fields that you would need to complete for your open data. How would you advise staff on the importance of having accurate metadata?

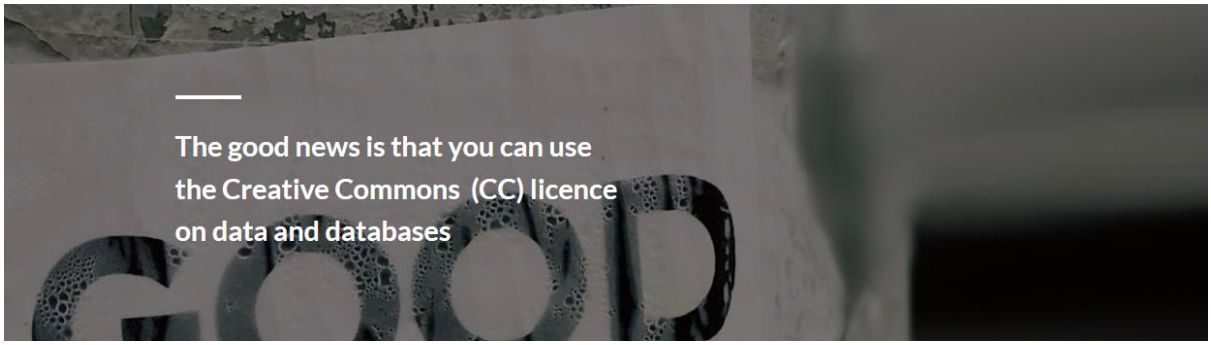
Metadata is not only important, but it is often a mandatory requirement in publishing open data.

AfLIA has started introducing African librabrians to linked data through Wikidata. It has received a grant from the Wikimedia Foundation to design and run training for librarians in English and in French.

AfLIA Wikidata online course

https://training.aflia.net/course/index.php?categoryid=13&lang=en

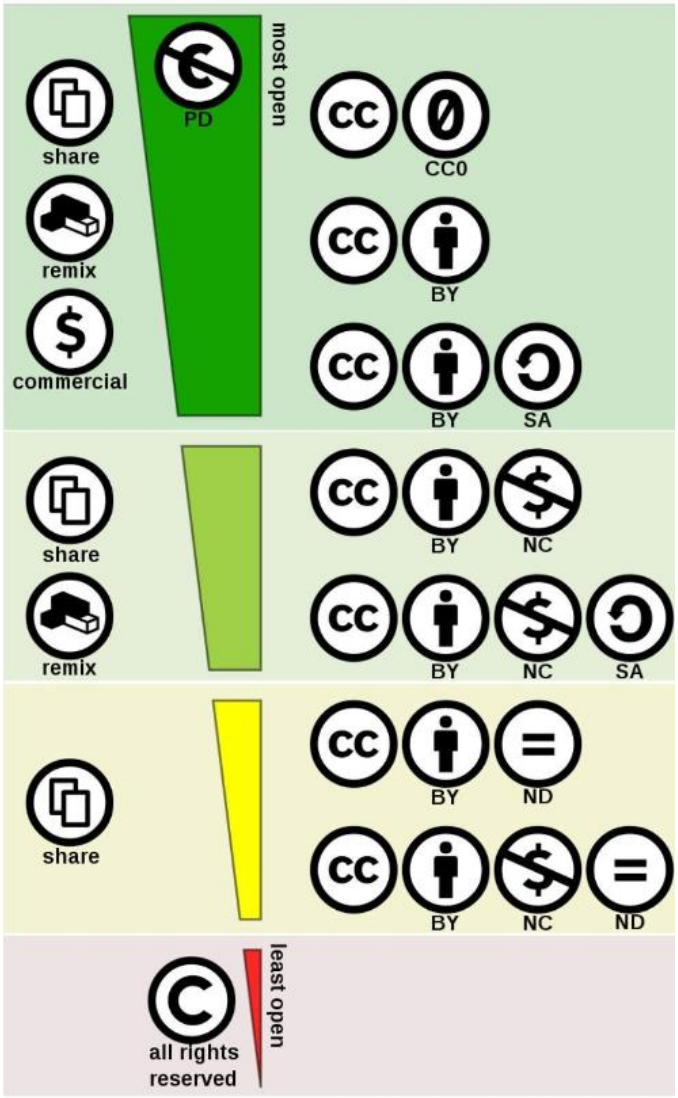
Open Data Licences



CC licences can be used to license databases. The most recent version (4.0) may be used to license databases subject to copyright and, where applicable, sui generis database rights. *Sui generis database rights prevent copying and reusing of substantial parts of a database (including frequent extraction of insubstantial parts).* Unlike copyright, database rights protect the maker's investment and not their originality.

CC does not recommend the use of its NonCommercial (NC) or NoDerivatives (ND) licences on databases intended for scholarly or scientific use.

In addition to CC licences, the CC0 Public Domain Dedication may be used to maximize re-use of databases. When applied, the effect is to waive all copyright and related rights in the database and its contents, placing it as close as possible into the worldwide public domain. In certain domains, such as science and government, there are important reasons to consider using CC0. Waiving copyright and related rights eliminates all uncertainty for potential users, encouraging maximal reuse and sharing of information.



There are additional open data licences which you can explore:

1. Creative Commons. Frequently Asked Questions about Data and CC licences. https://wiki.creativecommons.org/wiki/Data#Frequently_asked_questions_about_data_and_CC_licenses
2. Open Data Comments. Legal Tools for Open Data. https://opendatacommons.org/

Licence

Attributions and references

The following resources were used as part of the tutorial and are listed here for your reference:

1. Data. Creative Commons. 2019. https://wiki.creativecommons.org/wiki/Data CC-BY. Date Accessed: 20 September 2023
2. DataCite Metadata Schema 4.4. DataCite. 2021. https://schema.datacite.org/meta/kernel-4.4/ Date Accessed: 20 September 2023
3. Minimum Metadata. SAEON Open Data Platform Wiki. https://odpwiki.saeon.ac.za/minimum_metadata_for_the_saeon_odp C C-BY-SA. Date Accessed: 20 September 2023
4. Open Data Commons. Open Knowledge. CC-BY https://opendatacommons.org/ Date Accessed: 20 September 2023
5. Open Data Handbook. Open Knowledge. https://opendatahandbook.org/ Date Accessed: 20 September 2023
6. Open Data. Creative Commons. https://creativecommons.org/about/program-areas/open-data/ Date Accessed: 20 September 2023

7. Open Data. PLOS. https://plos.org/open-science/open-data/ Date Accessed: 20 September 2023
8. Open Knowledge Primer for African Universities Revised and Updated Edition. May 2023. Lisbeth Levy. CC BY. https://www.oerafrica.org/system/files/12591/open-knowledge-primer-june-2023.pdf?file=1&type=node&id=12591&force=1 Date Accessed: 20 September 2023
9. Understanding Open Science. UNESCO. 2022. CC BY-SA https://unesdoc.unesco.org/ark:/48223/pf0000383323 Date Accessed: 20 September 2023
10. UNESCO Recommendation on Open Science. 20 February 2023. https://www.unesco.org/en/open-science/about?hub=686 Date Accessed: 20 September 2023