flograttarola.com | GitHub github.com/bienflorencia | @flograttarola | ecoevo.social/@flograttarola

First steps towards your Open Science journey

Cutting vEdge tools #2
IAVS EcoInformatics Seminar Series

Florencia Grattarola | 15th March 2024

Agenda for today

First steps towards your Open Science journey

Agenda

- **WHY: Three fundamental premises**
  - Why do we do science?
  - ‘Everything is a remix’.
  - Reform research evaluation and implement rewarding schemes.
- **WHAT: Some OS concepts in a nutshell**
  - Research outputs, open research data, FAIR data, and licenses.
- **HOW: Tools to come on board**
  - Authorship and credit, persistent IDs, data-sharing (repositories, standards and licensing).

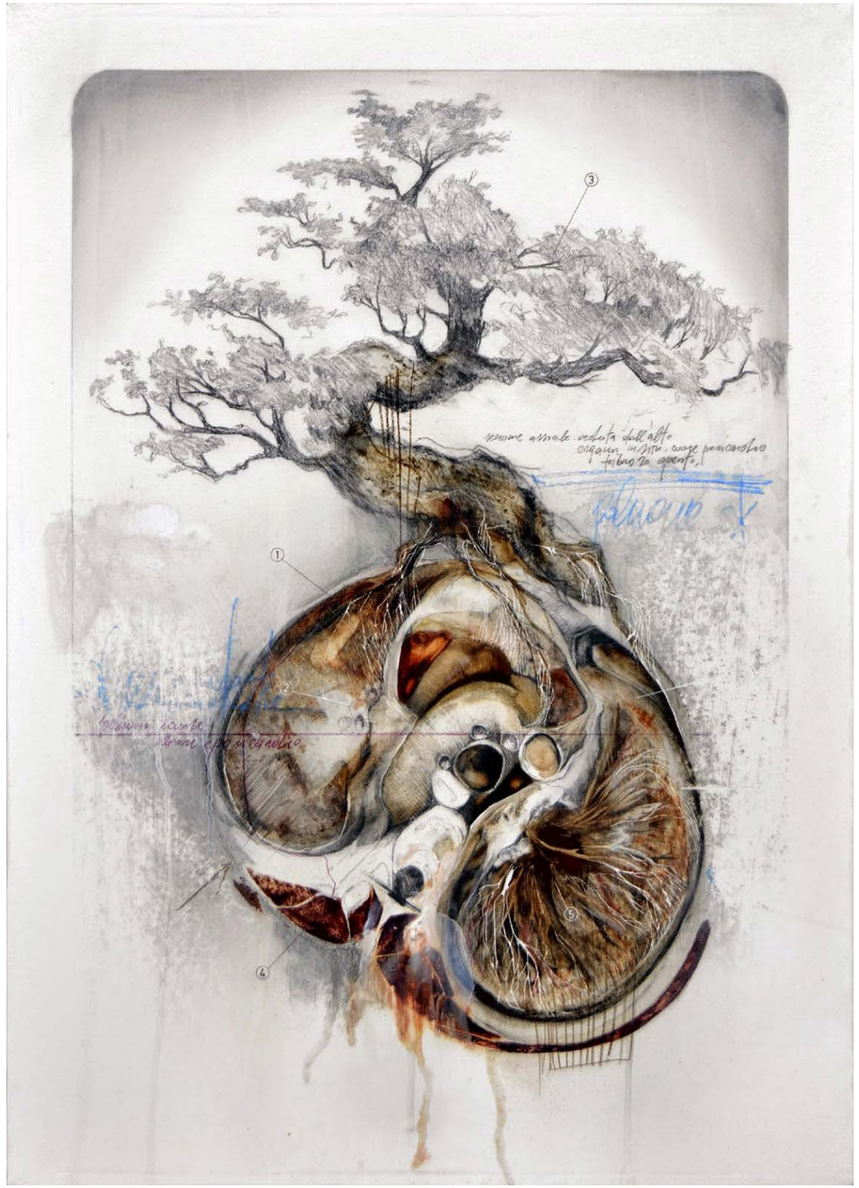
WHY

Three fundamental premises

Three grounding premises

Why do we do science?

- **science** is something we learn and do in society. As such, it is not a right. What is a right is to participate freely and actively in the benefits generated by the collective construction of **science**
- We don’t work for ourselves. Our successes are those of society as a whole. And that’s why we should also share them.



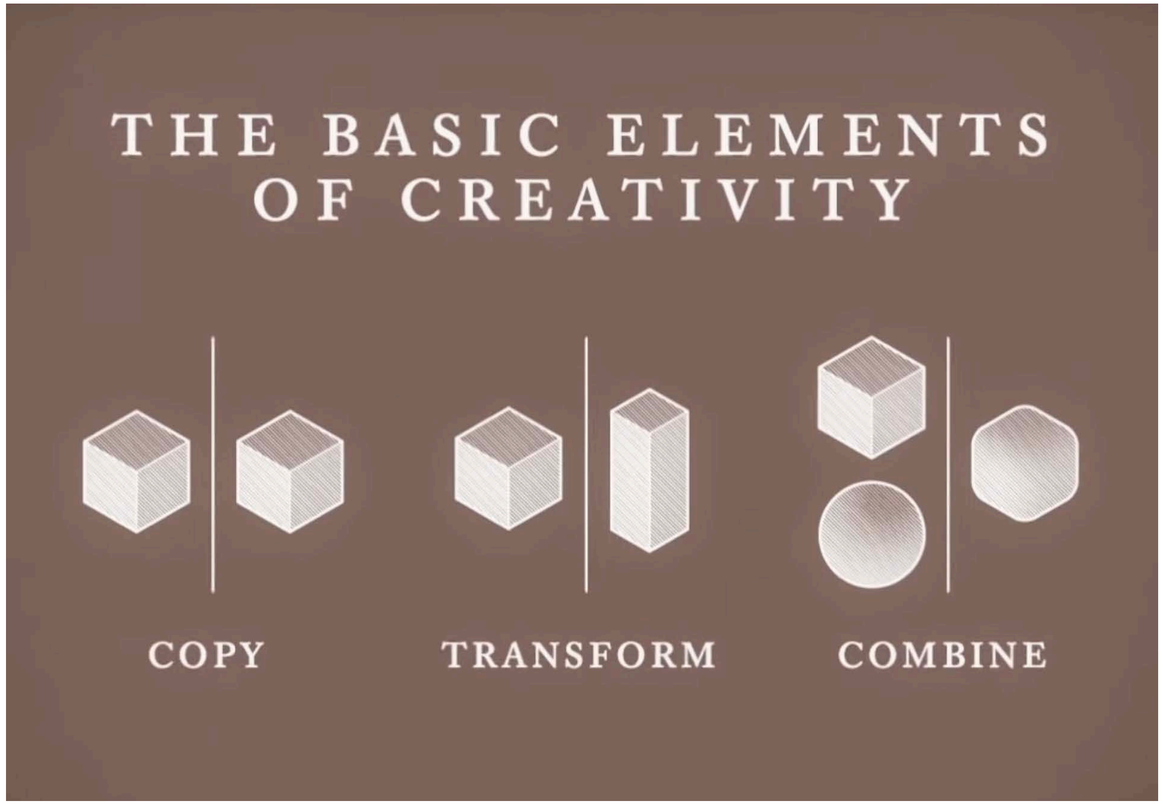
Why do we pay for public domain in Uruguay? Alejandro Gortázar (2018)
Nº 23 de Hemisferio Izquierdo, dedicated to the Commons.

Three grounding premises

‘Everything is a remix’

- Science is a **cumulative process** that builds on previously discovered knowledge.
- We can do this because **knowledge** becomes accessible and we are able to access it (*knowledge commons*).

Elinor Ostrom. Governing the Commons: The Evolution of Institutions for Collective Action



Kirby Ferguson. Everything is a Remix (www.everythingisaremix.info)

Three grounding premises

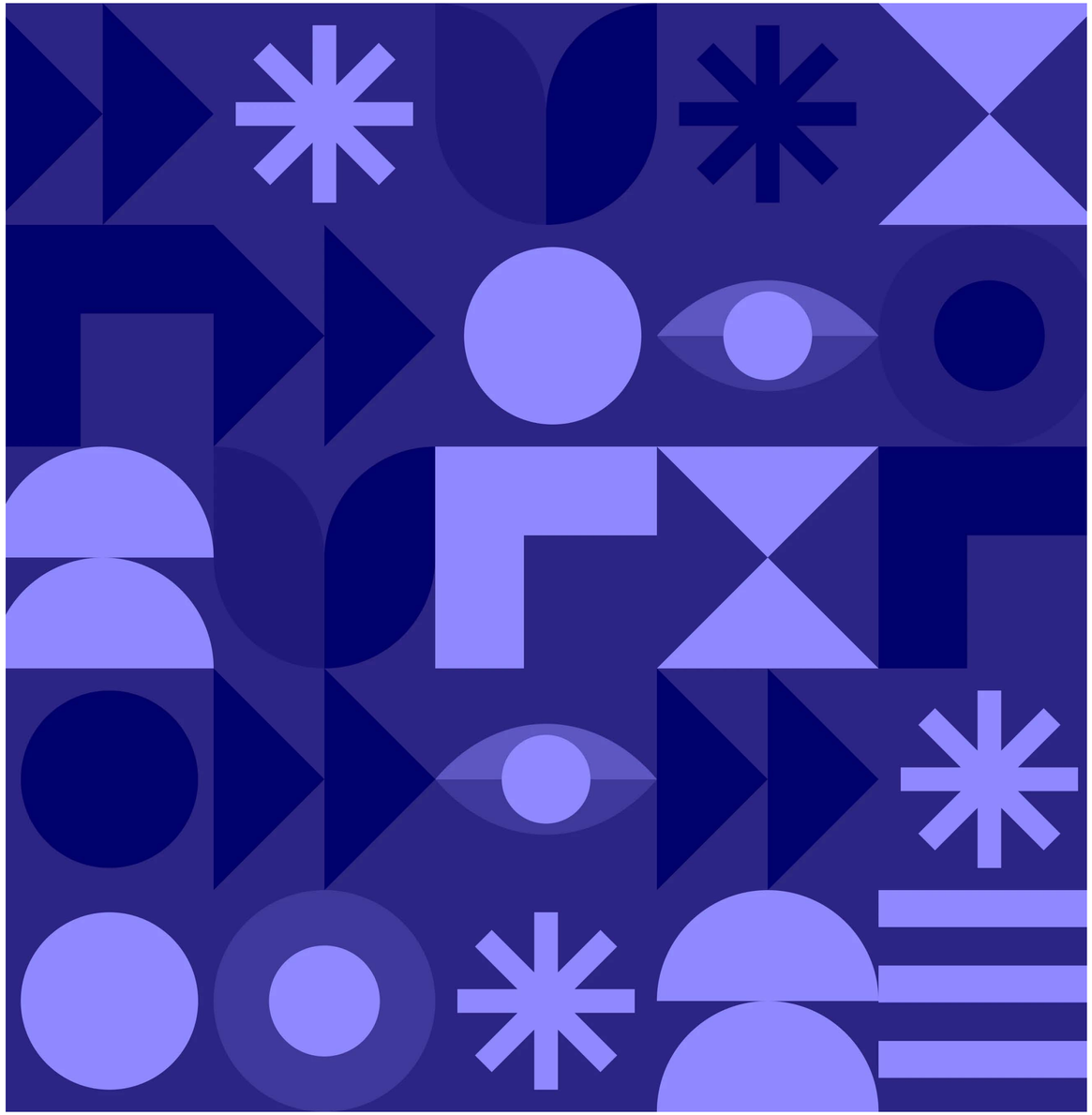
‘Everything is a remix’

Kirby Ferguson. Everything is a Remix (www.everythingisaremix.info)

Three grounding premises

Reform research evaluation

- How researchers are assessed today (mainly **impact factor**) doesn’t align with coming on board with open science.
- We need to recognise the **diverse outcomes, practices and activities** that maximise the quality and impact of research.

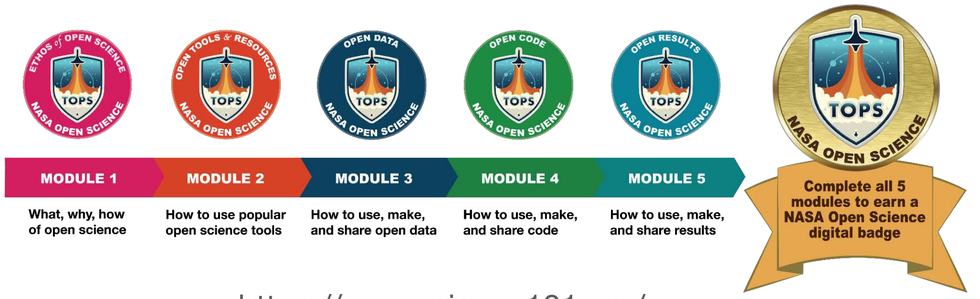


Three grounding premises

Implement rewarding schemes

- We need more **incentives**:
  - national and institutional policies,
  - tangible rewards (e.g., funds),
  - capacity building or support.

ANII
AGENCIA NACIONAL
DE INVESTIGACIÓN
E INNOVACIÓN



https://openscience101.org/

WHAT

Some OS concepts in a nutshell

OS concepts in a nutshell

OS activities

- Open science is so much more than open access of research articles.

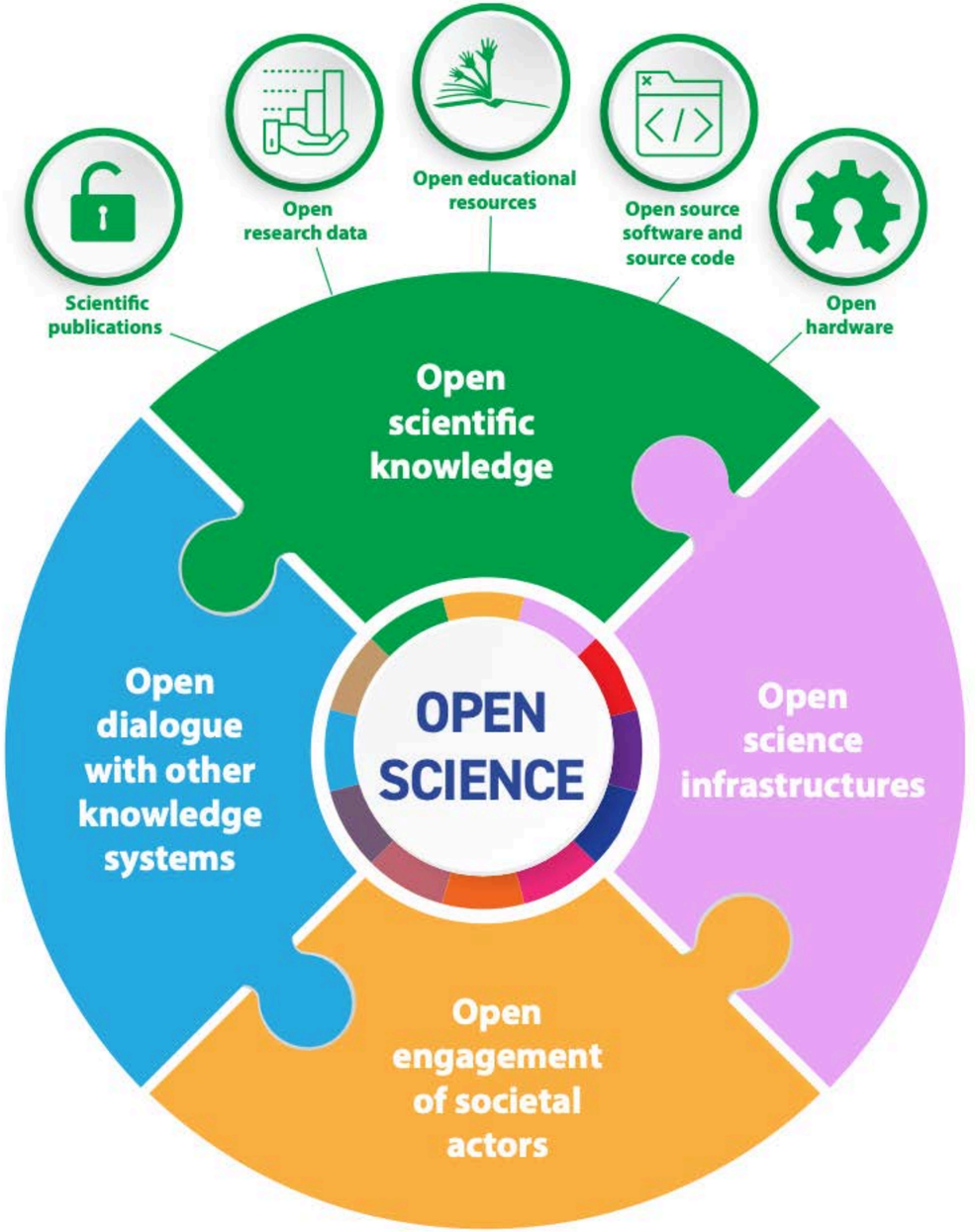
RDA-SHARC IG (SHAring Reward & Credit)
https://www.rd-alliance.org/groups/sharing-rewards-and-credit-sharc-ig

Open Science activities
Publishing a paper or monograph book as open access
Sharing a research manuscript as a preprint
Preregistration of the study design, methods, hypothesis etc., prior to commencing the research
Open or FAIR data management and sharing (for research data, software, models, algorithms, workflows etc.)
Participation in open peer review (being reviewed or the reviewer)
Participation in public engagement, including citizen or community science
Collaboration via virtual research environments or virtual laboratories

OS concepts in a nutshell

Research data

- It's the **evidence** most commonly accepted by the specific academic community as necessary to support research scientifically and technically.
- They are **generated through research activities** such as experiments, measurements, surveys, interviews, observations, etc.
- For example, they may include field notes, texts, illustrations, photographs, sounds, databases, and code.

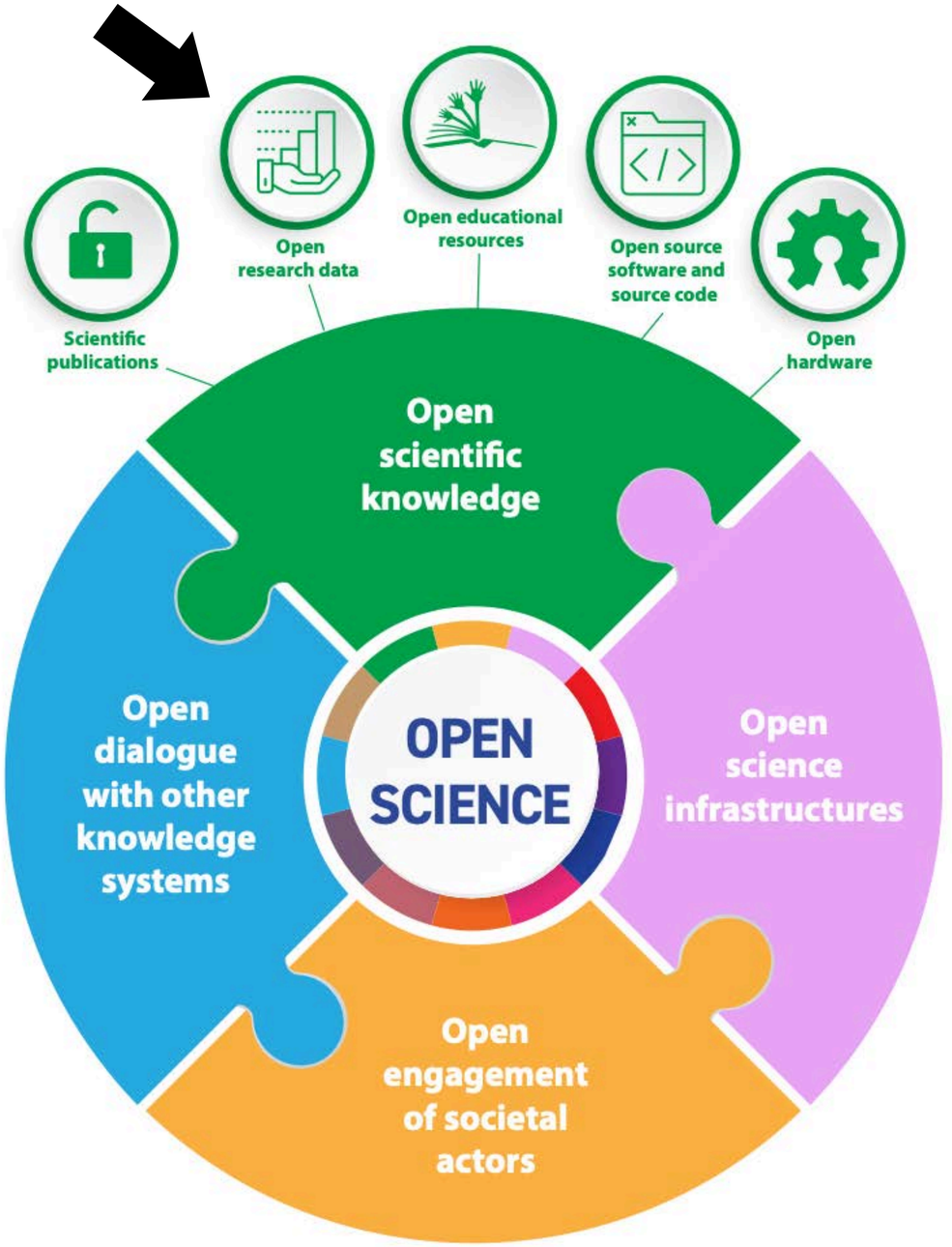


https://unesdoc.unesco.org/ark:/48223/pf0000379949

OS concepts in a nutshell

Open research data

- They can be **openly used, reused, retained and redistributed** by anyone, subject to acknowledgement.
- They are available in a user-friendly, human- and machine-readable format, in accordance with principles of good data management, such as **the FAIR principles.**

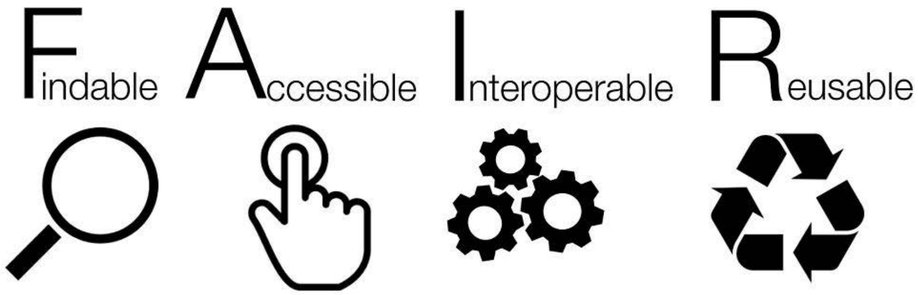


https://unesdoc.unesco.org/ark:/48223/pf0000379949

OS concepts in a nutshell

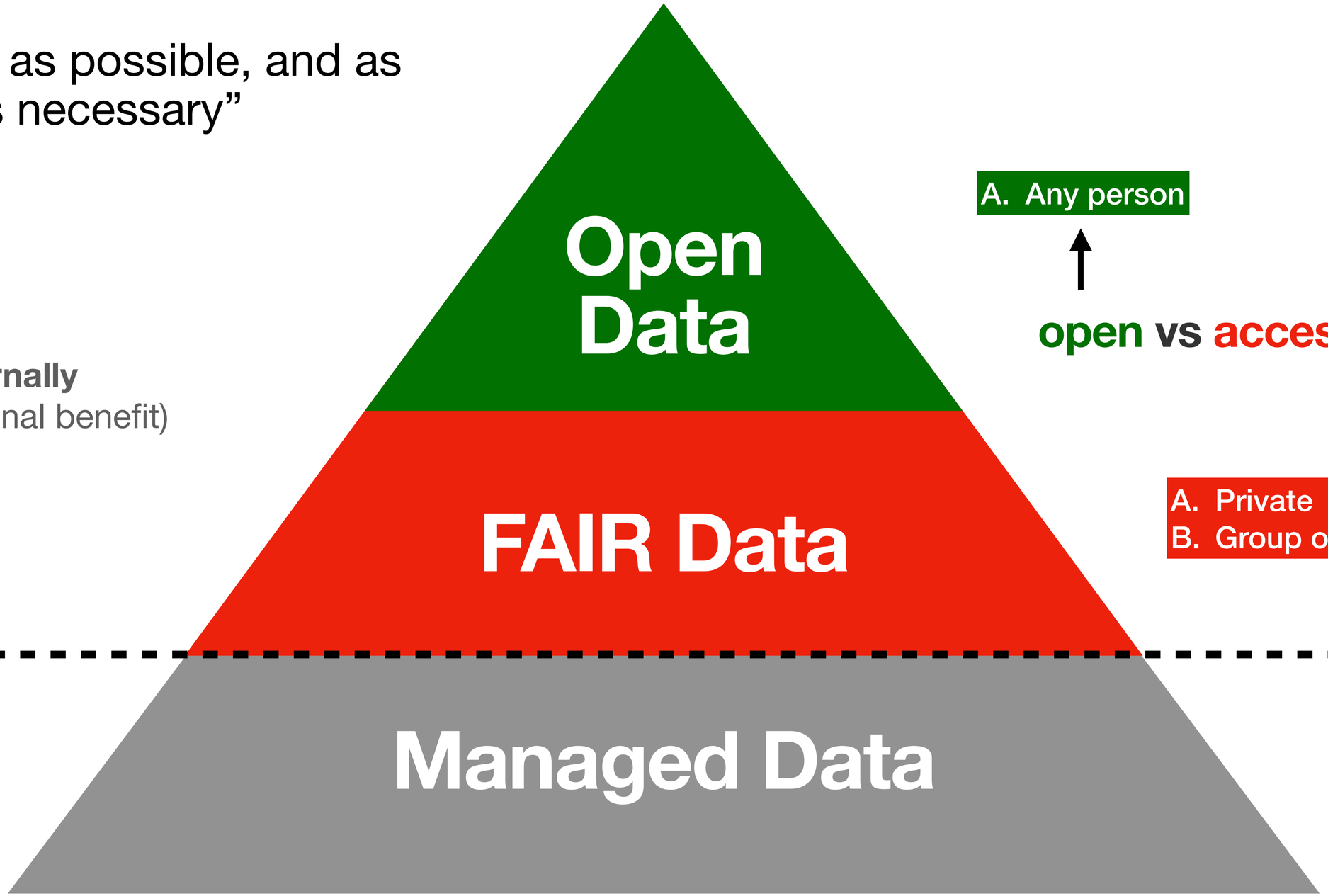
FAIR data

- These principles guide us in making research data **easy to find**, **accessible**, **interoperable**, and **reusable**.



https://www.go-fair.org/fair-principles/

"As open as possible, and as closed as necessary”



Externally
(for communal benefit)

Internally
(for own use)

A. Any person

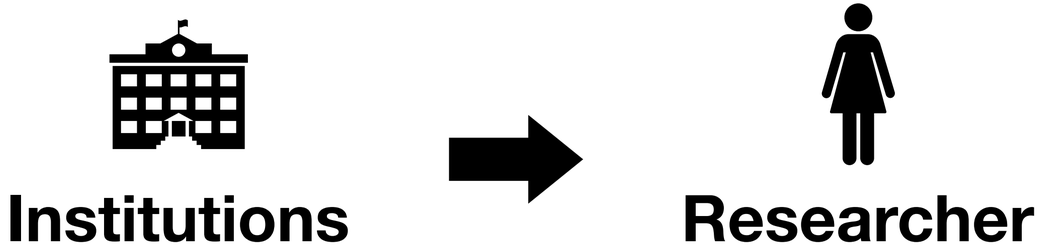
open vs accessible

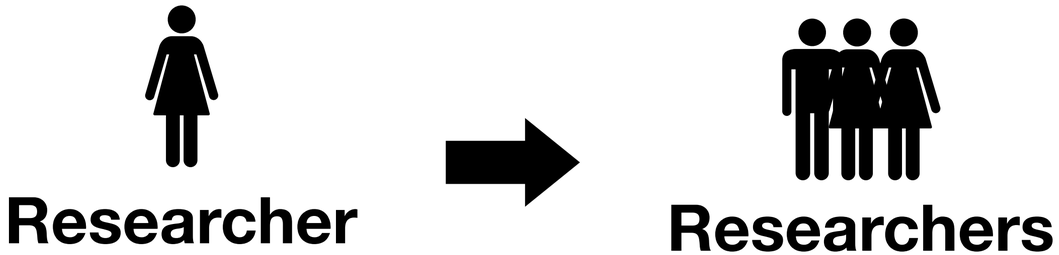
A. Private
B. Group of people

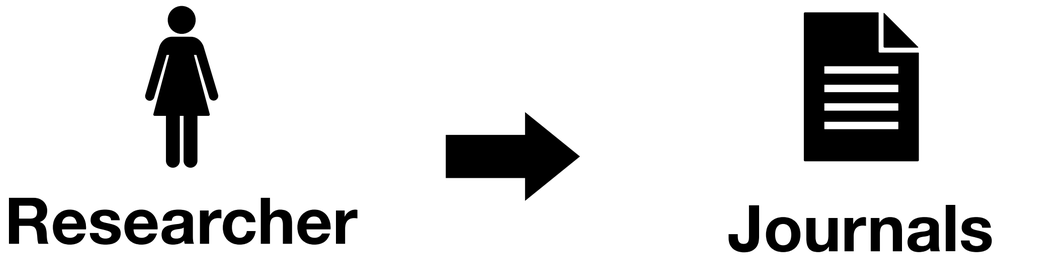
OS concepts in a nutshell

Data ownership

- **Who owns the data?**
- **Ownership** refers to the ability to access, create, modify, derive benefit from, or remove data, and also to the right to assign these access privileges to others.







Toolkit for Researchers on Legal issues

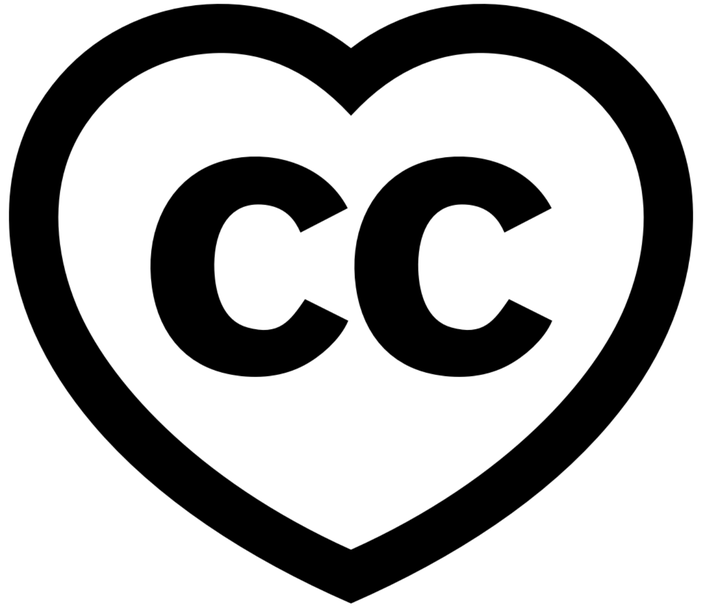
OpenAIRE ADVANCE

Margoni, T., & Tsiavos, P. (2018). Toolkit for Researchers on Legal Issues. Zenodo. https://doi.org/10.5281/zenodo.2574619

OS concepts in a nutshell

Licenses

- Licenses provide a simple, standardised way to **give the public permission to share and use** your creative work on conditions of your choice.
- They are not automatic. You need to **declare your chosen license**.



Katsushika Hokusai

HOW

Tools, tips and how-tos

Tools to come on board

Authorship and credit

- **Authorship** confers credit and has important implications for the academic career.
- A good idea is to **establish a standard** for your group or project to define **how to handle authorships**.

Handling co-authorships and author order in MOBI lab

Author: *Petr Keil* With contributions from: *Carmen Soria, Gabriel Ortega, Francois Leroy, Flo Grattarola, Kaca Tschernosterova, Frieda Wolke, Manuele Bazzichetto*

On Thu 21 September 2023, during our lab-meeting, we had a discussion about good practice concerning co-authorships in MOBI lab. Here is what we came up with:

Relevant or interesting resources

- A paper on the subject by Logan et al. (2017) PLoS ONE
- CRediT - Contributor Roles Taxonomy
- ICMJE guidelines for defining role of authors and contributors
- Ecological Society of America (ESA) code of conduct, section "Publication"

Who is a co-author and when is co-authorship deserved

A rough criterion is that co-authorship is deserved if at least 1 role on the CRediT list is clearly substantial, or if the person has at least 2 roles. This is, however, still vague. Hence, if in doubt, follow the next rule.

When in doubt if someone deserves to be a co-author on your paper, offer them an opportunity to contribute and deserve co-authorship in the upcoming phases of the writing and peer-review process.

When still in doubt, be inclusive. MOBI lab default policy is the "opt-out" policy.

Whoever contributes substantially to conception of ideas or study design, or execution of the study, or collection of the data, or to analyses, or their interpretation, should be given the chance to contribute to the main text at some point.

If you are listed on our manuscript and you have doubts if your own co-authorship is justified, try to justify it during the upcoming phases of peer-review, e.g. by helping with the revision, commenting on the reviewers comments, etc. This is usually a tedious work where every extra help to the lead author is much appreciated.

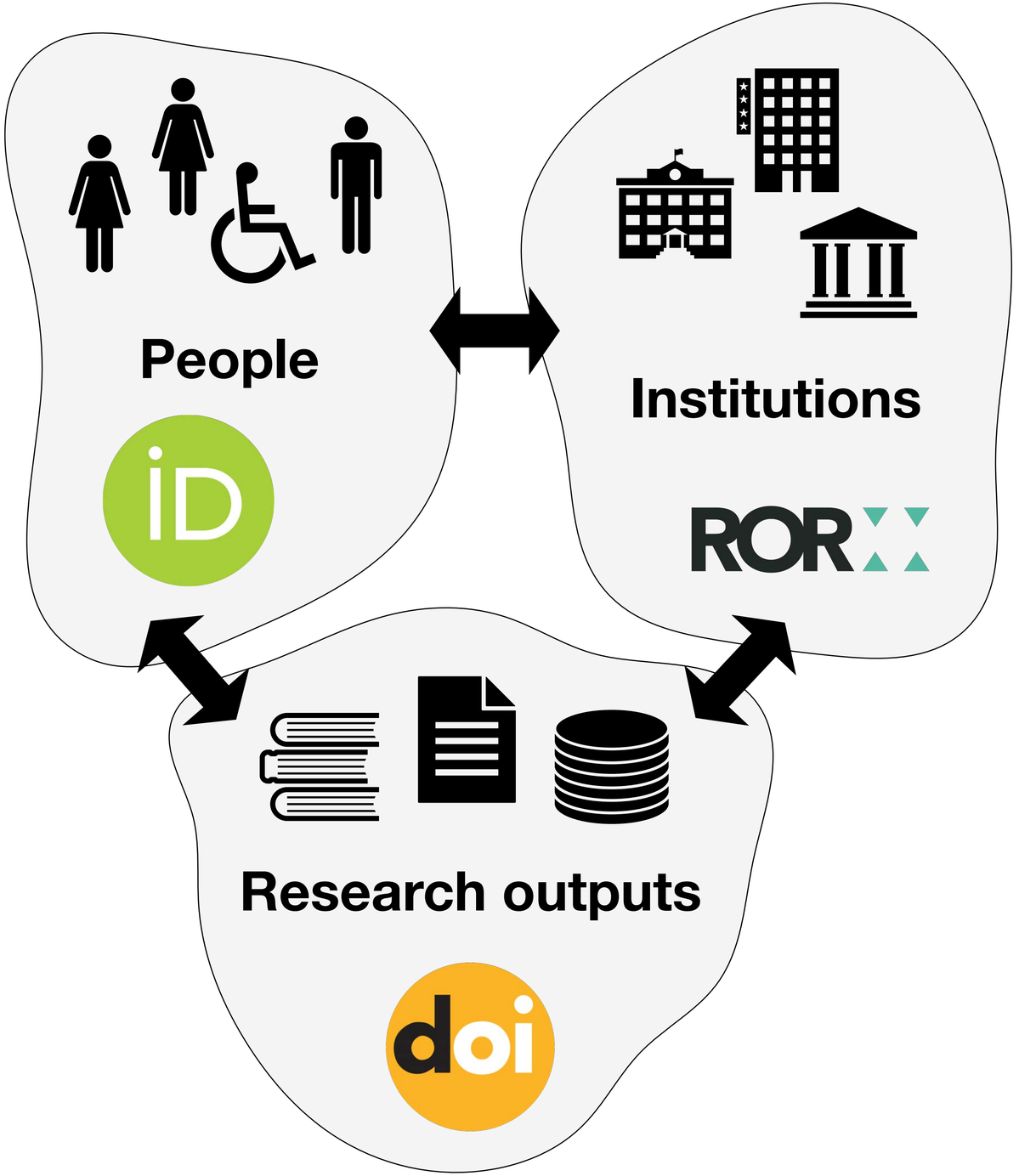
If, in MOBI lab, we publish a dataset and that dataset is under an open license, we treat these data as any other open and published data. This means that the authors of the published dataset do not have to be included as authors on future publications that will use the data. However, it is encouraged to reach out to the original authors, and involve them in the follow-up analyses as co-authors, if it benefits the science.

https://petrkeil.github.io/values/post/2023/10/05/coauthorships.html

Tools to come on board

Persistent identifiers (PIDs)

- **ORCID iD** provides a PID for researchers, and helps to distinguish the author's name
- **ROR iD** provides a PID for organisations in the research community.
- **DOI** (digital object identifier) is a PID that ensures digital objects can be permanently found online.



Tools to come on board

Persistent identifiers (PIDs)

- **Create an ORCID iD**

  https://orcid.org

- **Use it** on your articles, datasets (and metadata), code/software (and metadata), or any other object.

PUBLICATIONS: =

DATA: =

CODE: =

Stall, S., Specht, A., Amato, J. G., Corrêa, P. L. P., Curivil, F. A. L., David, R., Erdmann, C., Machicao, J., Miyairi, N., Murayama, Y., O'Brien, M., Santos, S., Wyborn, L., Vellenich, D. F., & Mabile, L. (2023). **Digital Presence Checklist**. Zenodo. https://doi.org/10.5281/zenodo.7841734

Tools to come on board

Persistent identifiers (PIDs)

- Journals will usually give you a DOI for your article.
- **Generate DOIs** (digital object identifiers) for your datasets (and metadata), code/software (and metadata), or any other object.
- **Use them** in your articles, datasets (and metadata), code/software (and metadata), or any other object.

GitHub + zenodo = doi

https://docs.github.com/en/repositories/archiving-a-github-repository/referencing-and-citing-content

https://figshare.com/

#CiteTheDOI

Tools to come on board

Data sharing

1. **Standardise** your data and make them **FAIR**.

2. Deposit your data in an appropriate **repository** and get a persistent identifier (e.g. a DOI).

3. Apply a **license** to your data that allows reuse by others (e.g., CC0 or CC-BY).

Tools to come on board

Data sharing: standards

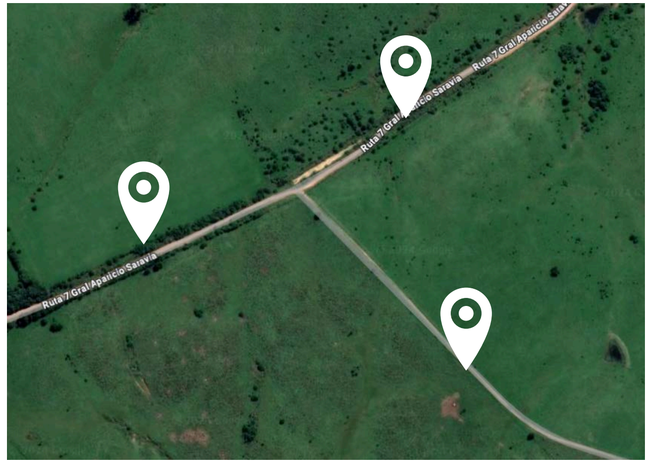
- When you **standardise our data**, you put them in a common language that can be understood by others (including machines).
- The most well-known standard for biodiversity data is the **Darwin Core** standard.

https://www.tdwg.org

https://dwc.tdwg.org/terms/

Tools to come on board

Data sharing: standards

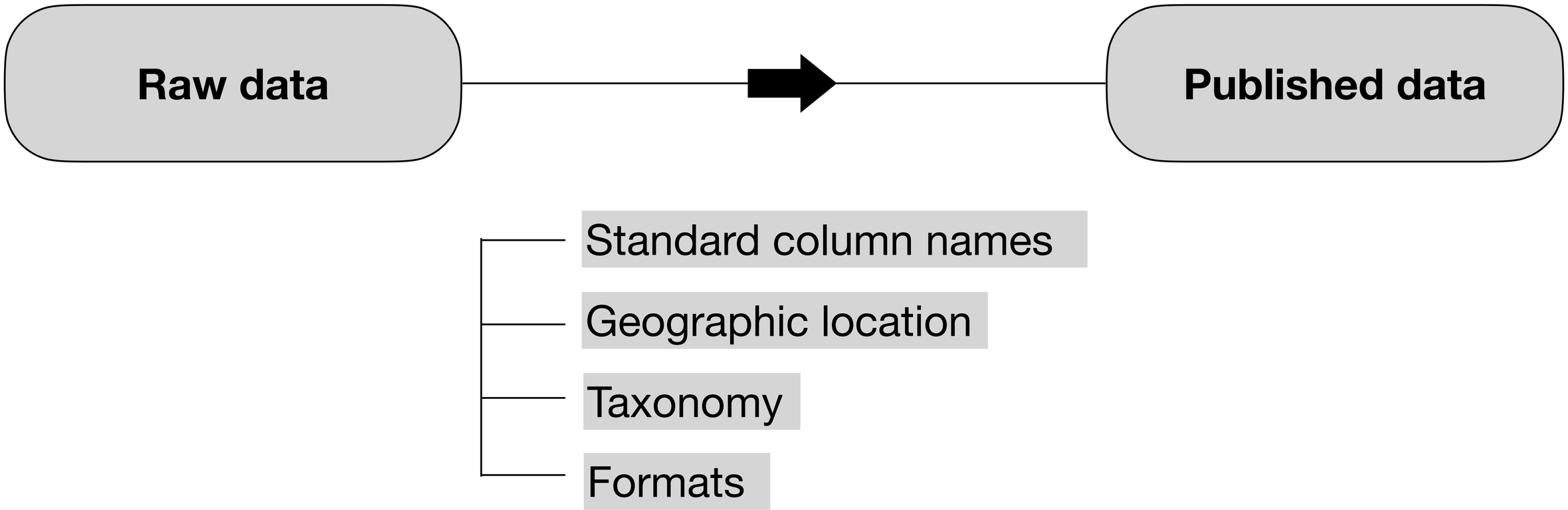


Roadkill mammals

Date	Colector	Latitud	Longitud	Localidad	Departamento	especie	Sexo
11/2/2022	Flo Grattarola	32°08'19.7"S	53°44'38.2"W	Paso Centurión	Cerro Largo	Cerdocyon thous	NA
13/2/2022	Flo Grattarola	32°08’12.5”S	53°44'16.0"W	Paso Centurión	Cerro Largo	Lontra longicaudsi	Fem
15/2/2022	Flo Grattarola	32°08'28.8"S	53°43'56.0"W	Paso Centurión	Cerro Largo	Tamandua tetradactila	M

Tools to come on board

Data sharing: standards



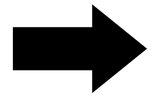
Tools to come on board

Geographic location

Data sharing: standards

- Values in other than decimal degrees
- Latitude and/or longitude = 0
- Latitude and/or longitude with a change in the sign
- Lack of Datum, precision and uncertainty terms
- No political-administrative levels documented

original
Latitud
Longitud
Localidad
Departamento



standard
decimalLatitude
decimalLongitude
coordinateUncertaintyInMeters
coordinatePrecision
geodeticDatum
georeferencedBy
georeferenceProtocol
locality
stateProvince
country
countryCode
continent

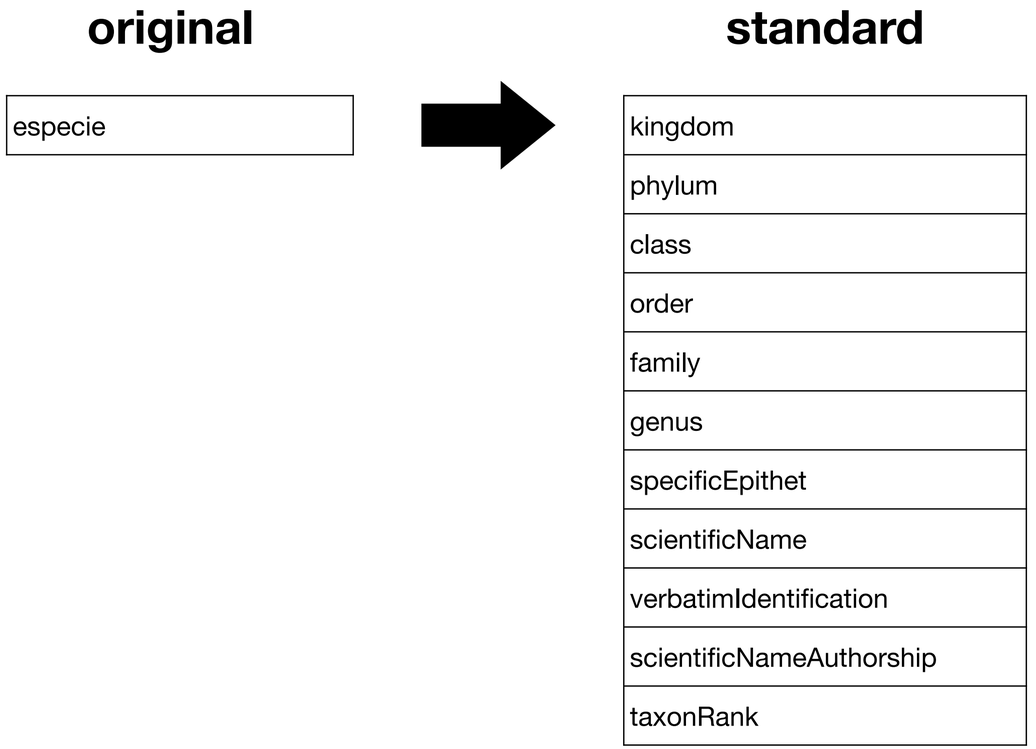
https://docs.ropensci.org/CoordinateCleaner/

Tools to come on board

Taxonomy

Data sharing: standards

- Synonym
- Misspelling
- Conceptual error
- Format error



Zermoglio et al. 2016. https://doi.org/10.1371/journal.pone.0146894

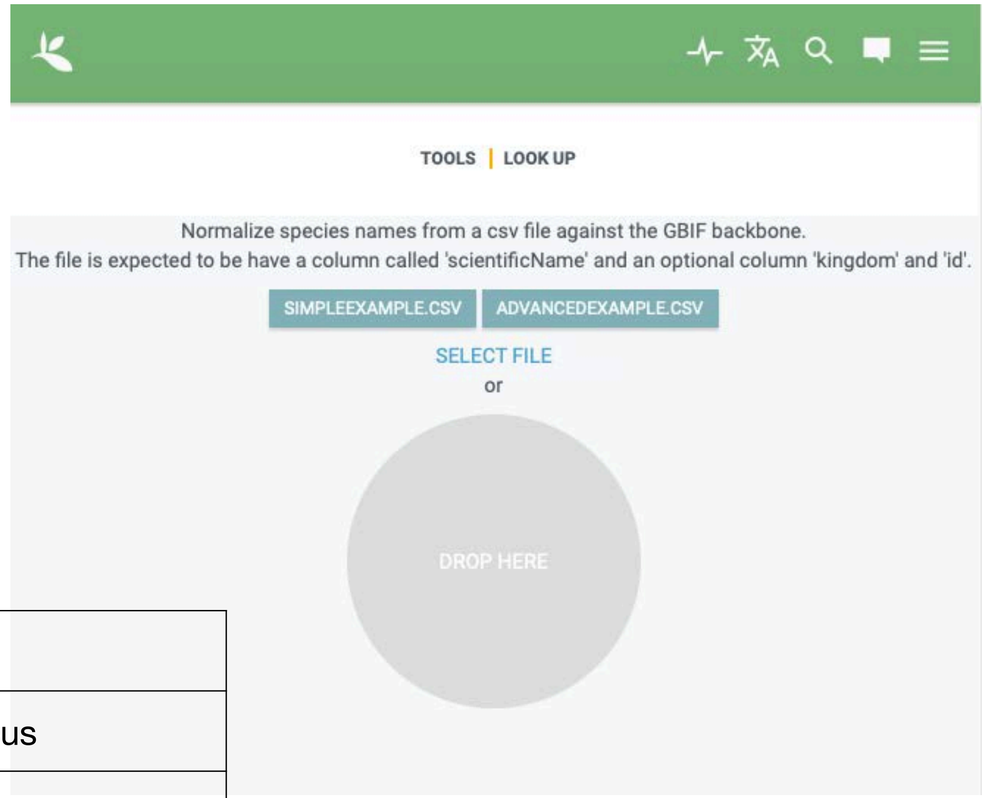
Tools to come on board

Taxonomy

Data sharing: standards

- Check species names

GBIF https://www.gbif.org/tools/species-lookup



especie
Cerdocyon thous
Lontra longicaudsi
Tamandua tetradactila

Tools to come on board

Data sharing: standards

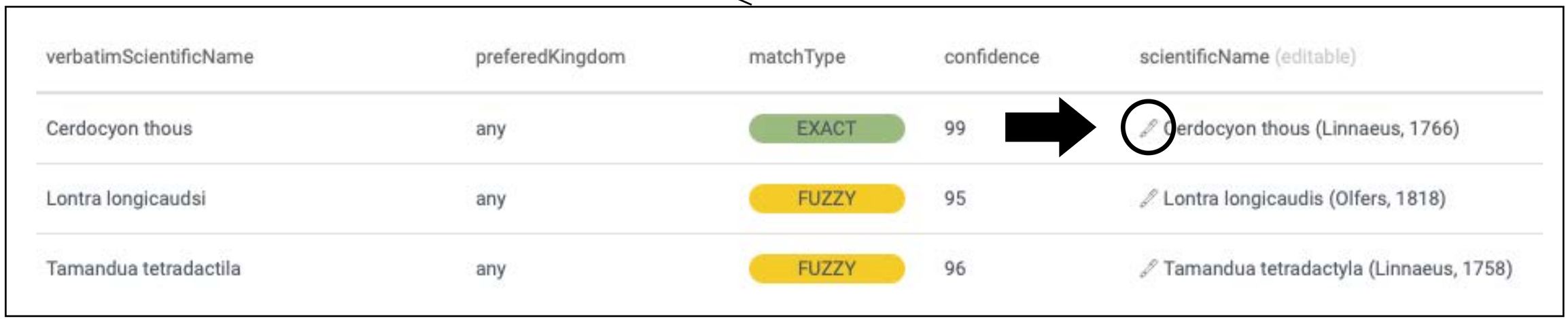
Taxonomy

GBIF

https://www.gbif.org/tools/species-lookup

TOOLS | LOOK UP

verbatimScientificName	preferedKingdom	matchType	confidence	scientificName (editable)	status	rank	kingdom	phylum	class	order	family	genus	species
Cerdocyon thous	any	EXACT	99	Cerdocyon thous (Linnaeus, 1766)	ACCEPTED	Species	Animalia	Chordata	Mammalia	Carnivora	Canidae	Cerdocyon	Cerdocyon thous
Lontra longicaudsi	any	FUZZY	95	Lontra longicaudis (Olfers, 1818)	ACCEPTED	Species	Animalia	Chordata	Mammalia	Carnivora	Mustelidae	Lontra	Lontra longicaudis
Tamandua tetradactila	any	FUZZY	96	Tamandua tetradactyla (Linnaeus, 1758)	ACCEPTED	Species	Animalia	Chordata	Mammalia	Pilosa	Myrmecophagidae	Tamandua	Tamandua tetradactyla

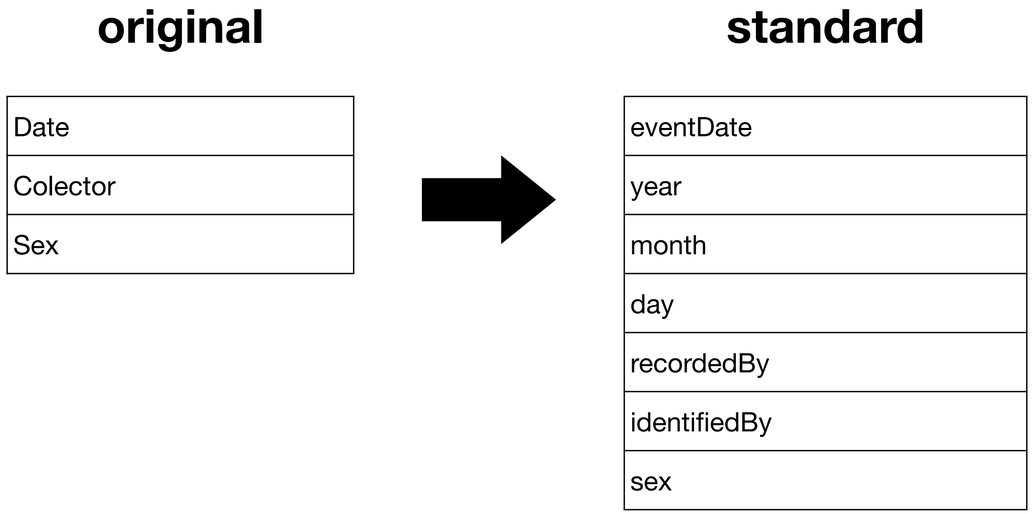


Tools to come on board

Formats

Data sharing: standards

- Dates
- Fields for which restricted values are recommended
- Controlled vocabulary fields



Tools to come on board

Data sharing: standards

- There are plenty of tools online.

OpenRefine

https://youtu.be/_YFw_bfwc3Y?feature=shared

GitHub

BIODIVERSIDATA
CONSORCIO DE DATOS DE BIODIVERSIDAD DEL URUGUAY

rBiodiversidata

These are useful scripts for biodiversity data cleaning, processing and quality controlling.

Tetrapod vertebrates

1. Check species names. (DwC term: **scientificName**).
2. Get taxonomic information for a species. (DwC terms: **kingdom**, **phylum**, **class**, **order**, **familiy**).
3. Get scientific name authorship for a species. (DwC term: **scientificNameAuthorship**).
4. Get conservation status and population trend (IUCN).

Plants

5. Check species names and get taxonomic information for a species. (DwC term: **scientificName**, **genus**, **specificEpithet**, **infraspecificEpithet**, **scientificNameAuthorship**, **taxonRank**, **taxonID**).
6. Get higher rank taxonomic information for a species (DwC terms: **kingdom**, **phylum**, **class**, **order**).
7. Get the state or province of the geographic location of a record (DwC term: **stateProvince**).
8. Update de event date of a record (DwC term: **eventDate**)

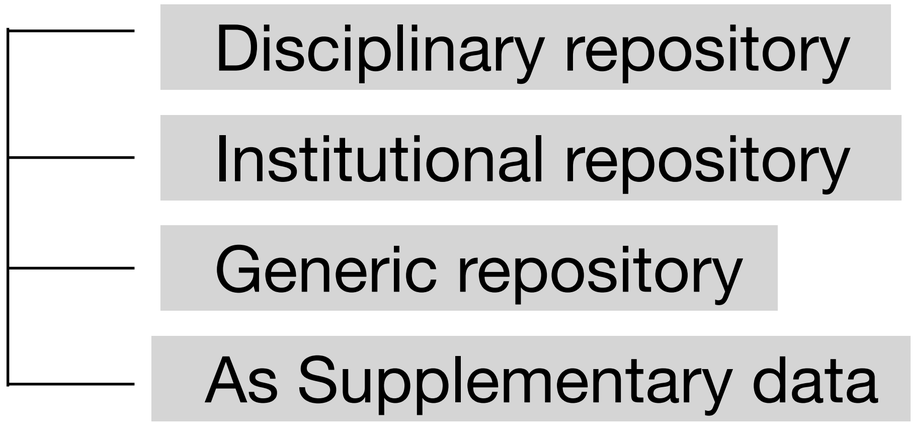
https://biodiversidata.org/recursos/codigo/

https://flograttarola.com/talk/flujos-de-trabajo-de-calidad-de-datos/Aplicacion_de_Flujos_de_Trabajo_de_Calidad_de_Datos.pdf

Tools to come on board

Data sharing: repositories

- A big question regarding data sharing is, **where** should I deposit my data?
- There are many repository options! Choose wisely :)



Tools to come on board

Disciplinary repository

Data sharing: repositories

- You can use a **disciplinary repository** to preserve your data according to recognised **standards** in the discipline.
- **GBIF** is the main data repository for primary biodiversity data. It's also a data aggregator, which means that data in other databases end up in GBIF.

GBIF

NCBI

Genbank

Tools to come on board

Institutional repository

Data sharing: repositories

- You can submit your data to an **institutional repository** if your institution/university/funding agency has one.
- These are usually **generic** repositories (not discipline-specific).

Tools to come on board

Institutional repository

Data sharing: repositories

- The Dataverse project provides infrastructure for institutional repositories.

https://dataverse.org/

The Dataverse® Project

Tools to come on board

Data sharing: repositories

Generic repository

- You can also preserve your data in a public **generic repository**.
- They can usually preserve a lot of different types of data types derived from diverse disciplines.
- A disadvantage is that the quality of the data and the **metadata** are not usually controlled.

zenodo

figshare

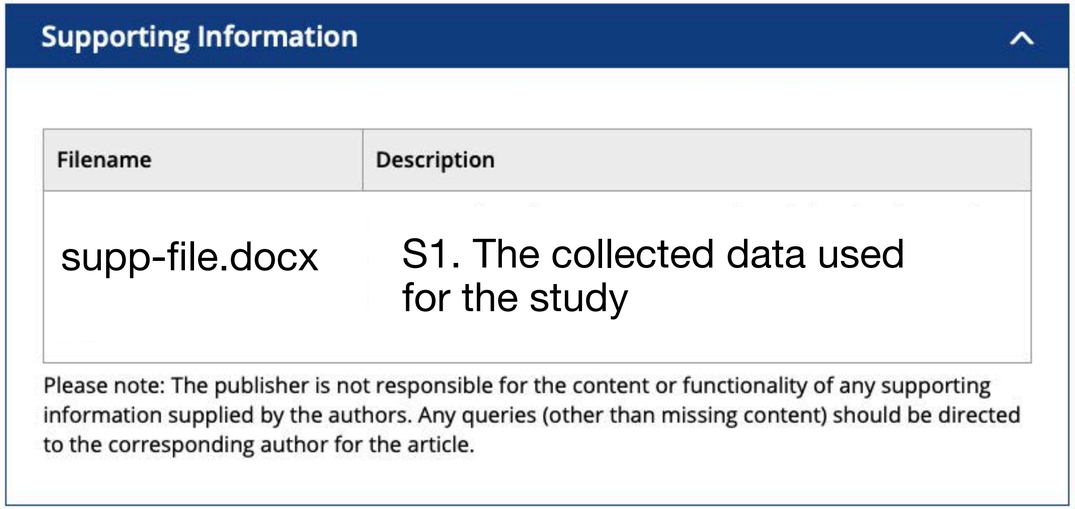
OSF

Tools to come on board

Data sharing: repositories

As Supplementary data

- You can also share your **data as supplementary material** for your research paper.
- A disadvantage is that the data are often **not curated** or stored for the long term. Also, the quality of the data and metadata are not usually controlled.



Tools to come on board

Data sharing: licensing

- For most of your content, you can use Creative Commons licenses.
- There are other options!

https://opensource.org/licenses

https://creativecommons.org/choose/

https://opendatacommons.org/licenses/by/1-0/

License (SPDX IDs)	Domain	By	SA	Comments
Creative Commons CCZero (CC0-1.0)	Content, Data	N	N	Dedicate to the Public Domain (all rights waived)
Open Data Commons Public Domain Dedication and Licence (PDDL-1.0)	Data	N	N	Dedicate to the Public Domain (all rights waived)
Creative Commons Attribution 4.0 (CC-BY-4.0)	Content, Data	Y	N	
Open Data Commons Attribution License (ODC-By-1.0)	Data	Y	N	Attribution for data(bases)
Creative Commons Attribution Share-Alike 4.0 (CC-BY-SA-4.0)	Content, Data	Y	Y	
Open Data Commons Open Database License (ODbL-1.0)	Data	Y	Y	Attribution-ShareAlike for data(bases)

http://opendefinition.org/licenses/

Tools to come on board

Data sharing: licensing

- You can apply a license by:
  1. Choosing a license.
  2. Attaching the license to the metadata of the research data.
  3. Setting up a README file for the data.

creative commons

open source initiative®

GitHub

This work is available under CC-BY license. If you reuse these scripts for your work, please cite our paper:

Grattarola, F., Bowler, D. E., & Keil, P. (2023). Integrating presence-only and presence–absence data to model changes in species geographic ranges: An example in the Neotropics. *Journal of Biogeography*, 00, 1– 15. https://doi.org/10.1111/jbi.14622

https://choosealicense.com/

WHY, WHAT, HOW..

This was a lot of information

I know

First steps to come on board

1. Create an ORCID iD
2. Do research data management
3. Join the community!

Join the community!

OS communities

RDA
RESEARCH DATA ALLIANCE

CoARA

LA Referencia
Red de repositorios de acceso abierto a la ciencia

Biodiversity Information Standards
TDWG

GBIF

SORTEE
Society for Open, Reliable, and Transparent Ecology and Evolutionary Biology

Meta Docencia

Second steps to come on board

- Spread the word!

¡Gracias!

Czech University of Life Sciences Prague

MOBI Lab

These slides can be downloaded and reused. Please credit the authors.

Acknowledgements
To those that have contributed to the commons.

References
Alejandro Gortázar: Por qué pagamos dominio público en Uruguay; Elinor Ostrom: Governing the Commons; DANS data game (dans.knaw.nl); PARSE project (http://parsecproject.org/); Everything is a remix (watch here: https://kirby-ferguson.squarespace.com/everything-is-a-remix-remastered).

Icons
Stupid Fun Science (CC0).

Photos
Nana Smirnova, Marino Linic, Julia Joppien, Fredy Jacob, Finn Hackshaw, Bernard Hermant, Claudio Schwarz, Shane Rounce, MagicPattern on Unsplash

flograttarola.com | GitHub github.com/bienflorencia | @flograttarola | ecoevo.social/@flograttarola