



Midterm update to your DMP



Korbinian Bösl
Data management coordinator
Centre for Digital Life & ELIXIR Norway
27 November 2025

Objectives

? Questions

- How do you backup up you data?
- How can I make sure data is unaltered by storage or transfer?
- How can I document my data?

! Objectives

- Check in and update your DMP
- Identify potential gaps in your RDM strategy

NFDI4Chem Knowledge Base

A place for all knowledge regarding Research Data Management (RDM) in Chemistry

[Get started](#)


Domains



Roles



Handling Data



Electronic Lab Notebooks



Data Publishing

knowledgebase.nfdi4chem.de/knowledge_base


[Data management](#)
[About](#)
[Contribute](#)
[GitHub](#)

The Research Data Management toolkit for Life Sciences

Best practices and guidelines to help you make your data FAIR (Findable, Accessible, Interoperable and Reusable)

What can we help you find?

Browse all topics by



Data life cycle

Start here to get an overview of research data management based on stages in the data life cycle.



Your role

Identify your role in research data management, find data management resources relevant for you, and information to help you progress in your career path.



Your domain

Learn about data management tasks that affect your domain or research community, and the solutions adopted to address them.

rdmkit.elixir-europe.org

Current phase

Before Submitting the DMP

I. Administrative information

✓ I.1 Contributors

Horizon Europe DMP

Each person contributing to creating or executing the data management plan should be added as a contributor. A project probably should have a Contact Person, and a Data Curator.

Chapters

I. Administrative information ✓

▼ Contributors

▼ Jiangli Gui

- 🗨 Name
- 🗨 E-mail address
- 🗨 ORCID Identifier
- 🗨 Affiliation
- 🗨 Role

▼ Marco De Vivo

- 🗨 Name
- 🗨 E-mail address
- 🗨 ORCID Identifier
- 🗨 Affiliation
- 🗨 Role

▼ Research Project(s)

▼ Design compounds ta

- 🗨 Project name
- 🗨 Project acronym
- 🗨 Project number
- 🗨 Project abstract
- 🗨 Starting date of the project
- 🗨 Ending date of the project

▼ Funding

▼ HORIZON EUROPE Mar...

- 🗨 Funder

Select phase



Before Submitting the Proposal

Some data stewardship issues already need to be addressed during the creation of a proposal for a research project. For example these could be questions that can have a significant impact on the budget or timeline of a project. Only such questions are labeled as "desirable" in this phase.

Before Submitting the DMP

In many research projects, the finished data management plan is one of the early deliverables. Questions that should be addressed in those early months of the project are highlighted as desirable when this phase is selected.

Before Finishing the Project

When a project rounds up, this means that all data has found its final spots. In this phase, questions that document the final state of affairs will be labeled as desirable.

After Finishing the Project

After a project is finished, the team conducts a comprehensive project review to evaluate its overall success and identify areas for future improvement and followup. Relevant feedback is gathered from stakeholders, and the team is finalizing reports and archiving project documentation. Finally, lessons learned are being documented for future reference. Questions relevant for these processes will be labeled as desirable in this phase.

🔄 Clear answer

Answered 6 months ago by Jiangli Gui.

✓ I.1.a.3 ORCID Identifier

Backup

✓ IV.3.a.5 Have project members been instructed about the risks (generic and specific to the project)?



Horizon Europe DMP

Project members may need to know about passwords (not sharing accounts, using different passwords for each service, and two factor authentication), about security for data they carry (encryption, backups), data stored in their own labs and in personal cloud accounts, and about the use of open WiFi and HTTPS.

☒ Desirable: *Before Submitting the DMP*

Backup

Avoid losing data!

Use (institutionally) managed storage with automatic backup (and snapshots?)

OneDrive/Sharepoint @ UiB

Avoid: data solely saved to laptops, stand-alone HDDs, thumb drives, ...

In general

3 - 2 -1 rule

3 copies

2 different media

1 off-site backup

5min to check the backup status of your data

Checksums

Checksum your data to detect corruption during storage or transfer

MD5sum is built in every OS

windows powershell: `$ certutil -hashfile [filename] MD5`

Mac OS/nix: `$ md5sum [filename(s)]* > checksums.md5`
`$ md5sum -c checksums.md5`

For security critical or legally relevant: sha256

Demo + 5 min for you to checksum one of your data folders

✓ II.1.b.1.a.3.b.4 Is the reference database or dataset versioned?



Horizon Europe DMP

Many reference datasets and databases evolve over time. If the reference data changes, this may affect your results. If different versions of a reference data exist, you need to establish your "version policy".

☒ Desirable: *Before Submitting the DMP*

 Data Stewardship for Open Science: *Is the data resource versioned?*

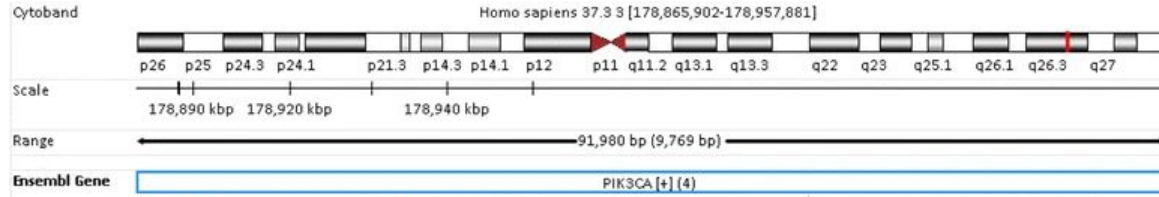
☒ a. No

☐ b. Yes 

 Clear answer

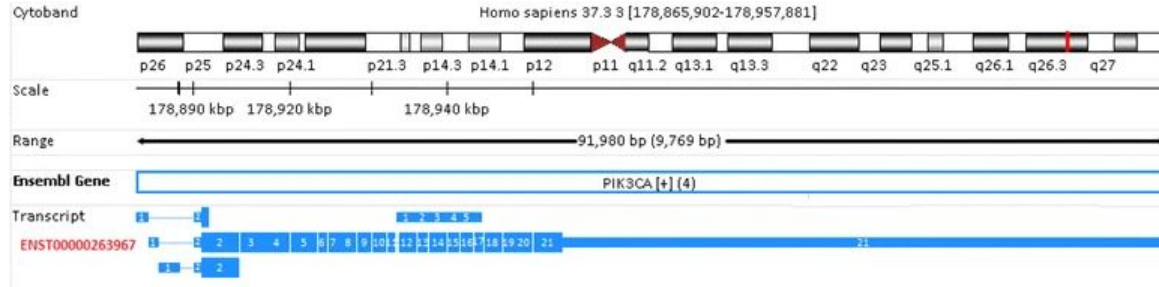
Sequence identifiers:

Gene: PIK3CA



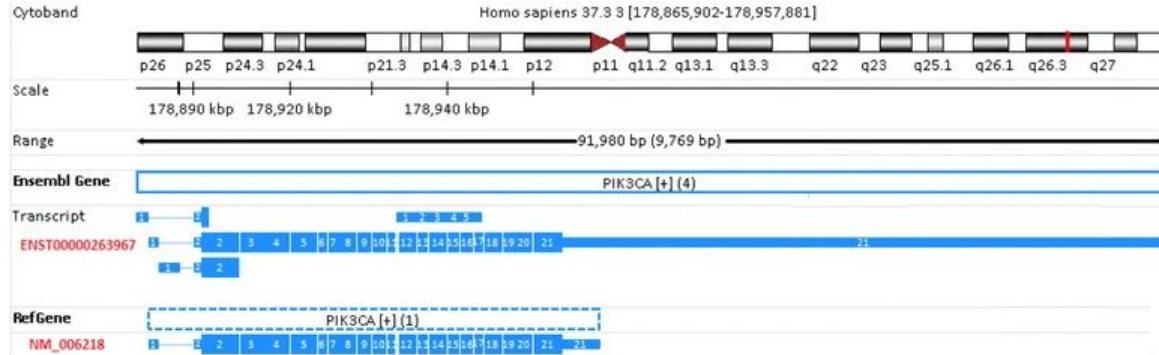
Sequence identifiers:

Gene: PIK3CA



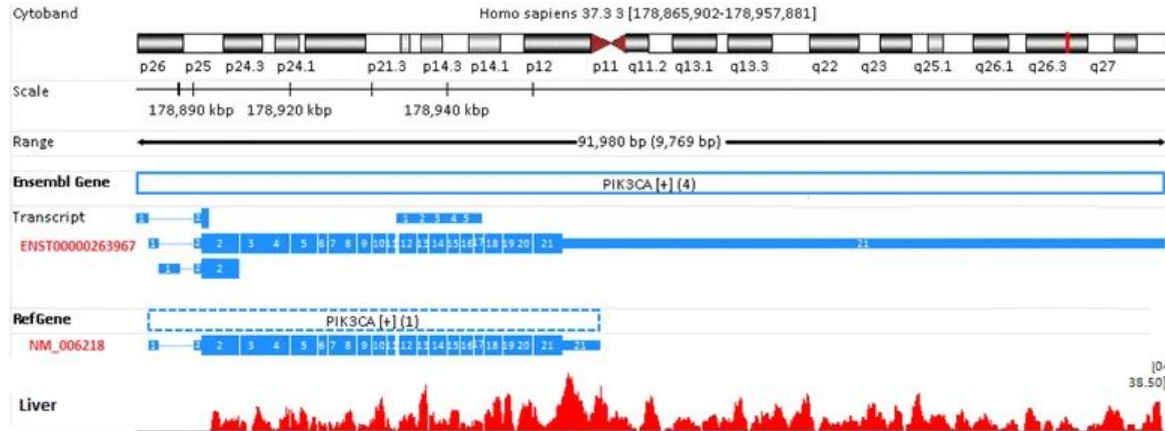
Sequence identifiers:

Gene: PIK3CA



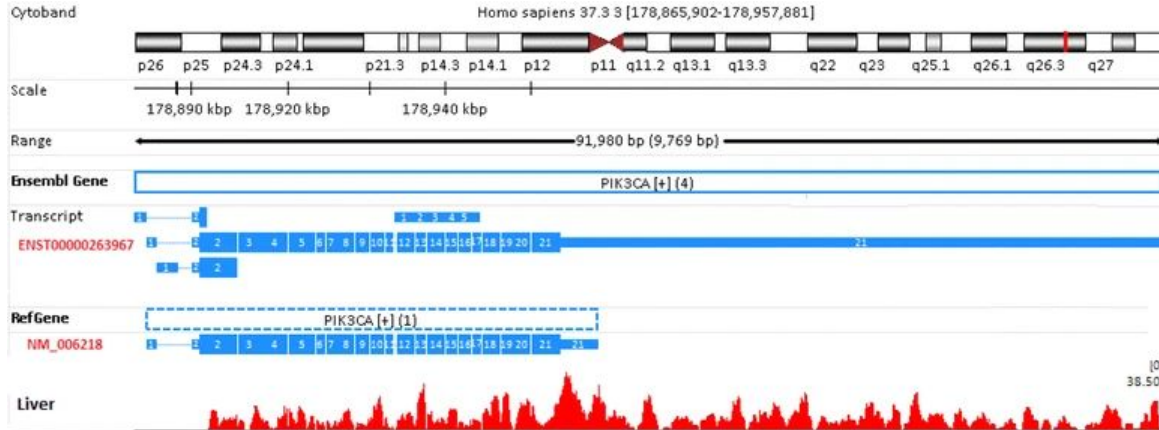
Sequence identifiers:

Gene: PIK3CA



Sequence identifiers:

Gene: PIK3CA



ENST00000263967.

2
3
4

Compound identifiers:

Assigned by authority:

CAS Registry Number

ChEMBL

ChEBI

PubChem

...

Structure (based):

SMILES

Non-Unique

Limited stereochemical & aromatic coverage

Limited geometric representation

InChi(Key)

...



ID mapping service

ebi.ac.uk/unichem/

**2 min: Find an example molecule
on e.g. PubChem/Wikipedia and
run the search on Unichem**

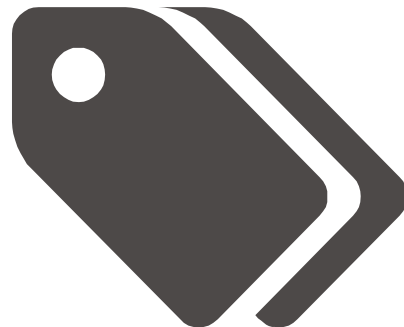
README files

For information not in your structured metadata

Description of files, folder structure, ...

Information from your (paper) labbook

Processing steps



README files

Study/project level README

This README file was generated on [YYYY-MM-DD] by [NAME]

GENERAL INFORMATION

- Study/project title:
- Description: <provide a short description of the study/project>
- Principle Investigator:
- Link to Data management plan

ORGANIZATION

- Folder structure: similar to folder structure example above (below)
- File naming conventions (with examples) <unless your project is big and you have README files in every subfolder with this information provided there>
- File formats: <Provide a list of all file formats present in this study/project>

RDMkit -> your task -> data documentation

Data level README

GENERAL INFORMATION

- Dataset title:
- Description: <provide description of the dataset origin, steps used in its generation, content and its purpose>

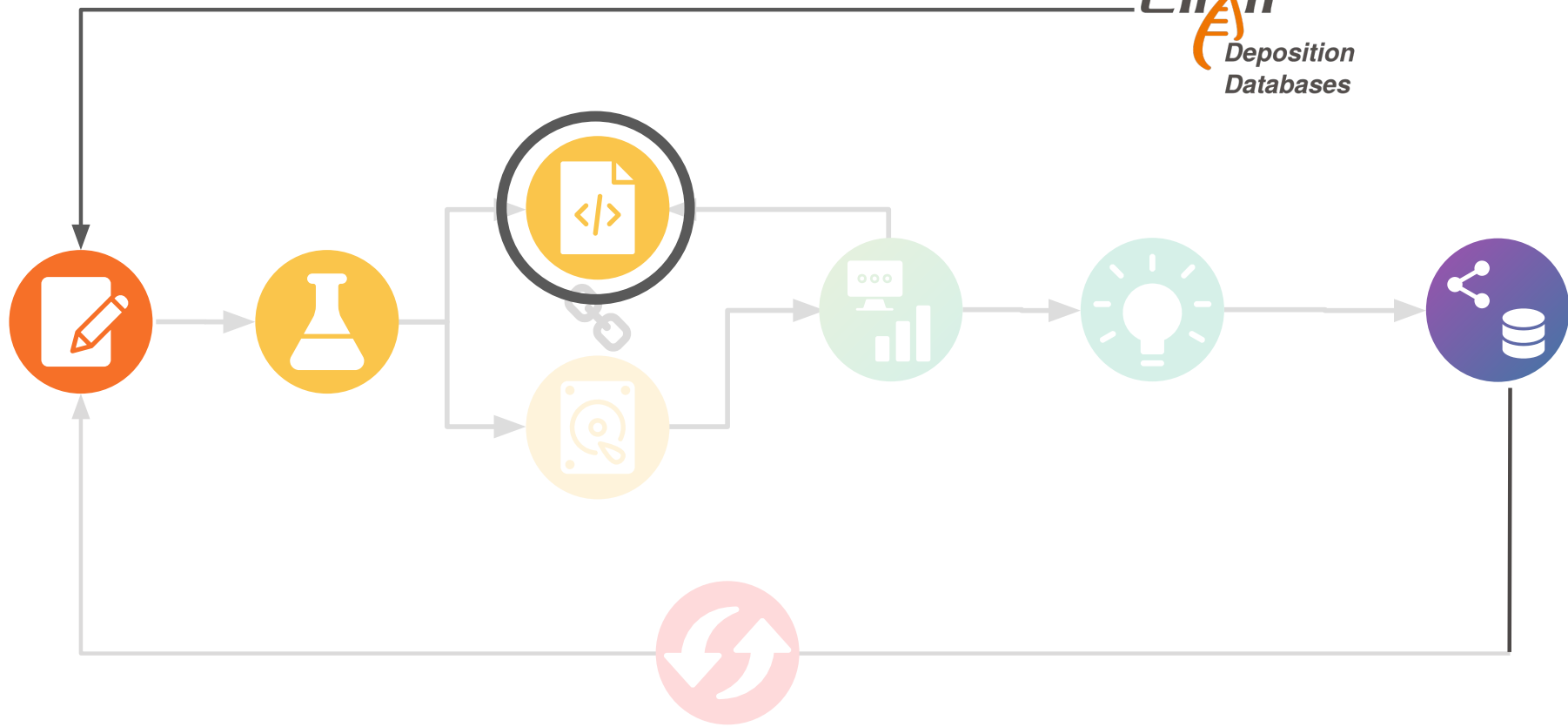
ORGANIZATION

- Folder structure: similar to folder structure example above (below)
- File naming conventions: <provide explanation of the elements used, allowed values and examples>
- File formats: <Provide a list of all file formats present in this dataset>

[...]



15 min for you to create a README for one of your data folders



Minimum Information about a high-throughput Nucleotide SeQuencing Experiment (MINSEQE)



<https://www.fged.org/projects/minseqe/>

Reporting standards for e.g. **RNA-Seq**

Required for deposition of sequencing data on ArrayExpress



Information regarding submission on ArrayExpress

<https://www.ebi.ac.uk/arrayexpress/submit/overview.html>



MINSEQE entry on FAIRharing:

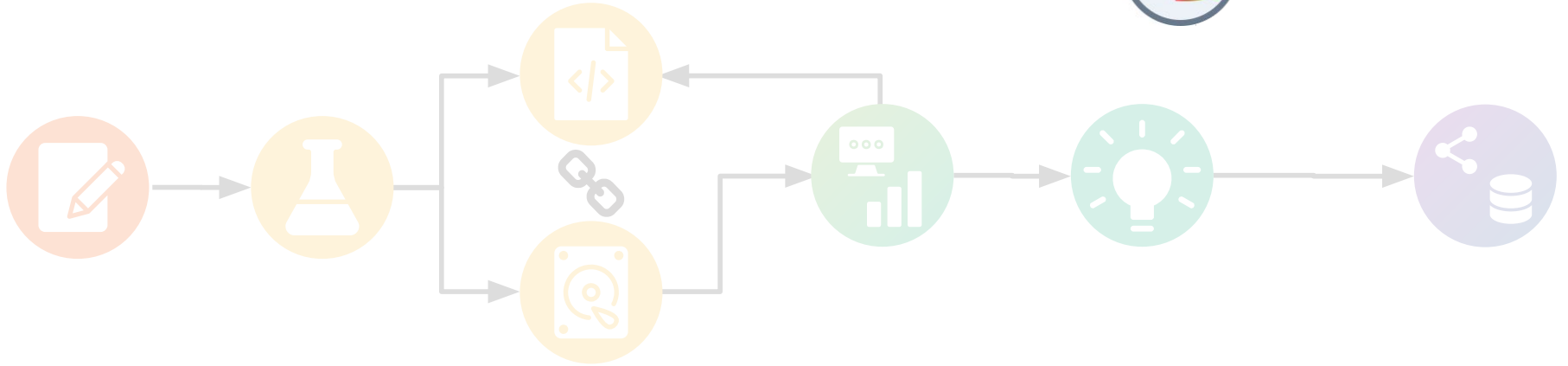
<https://fairsharing.org/bsg-s000174/>

Example

MINSEQE

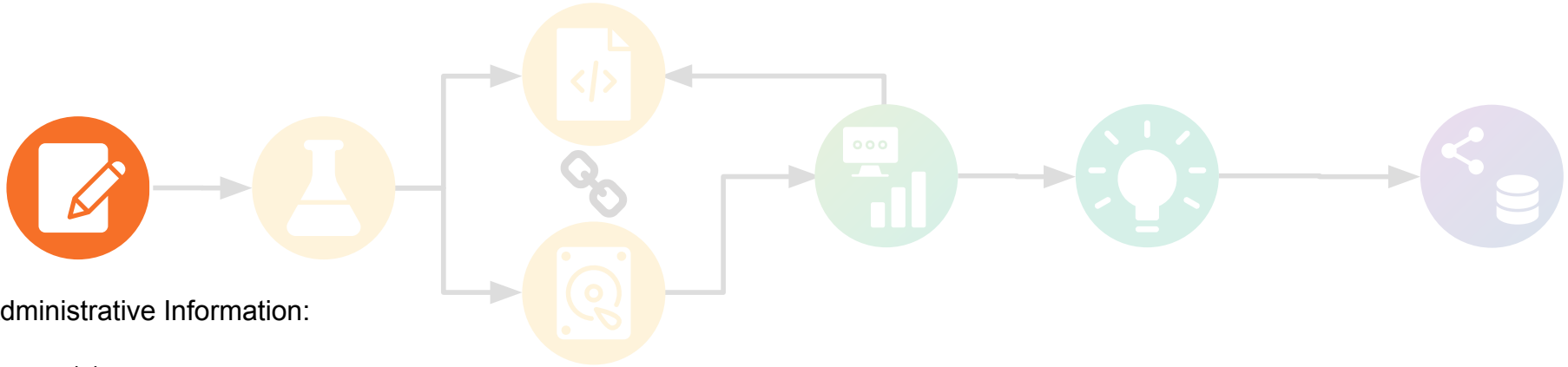


ArrayExpress



Type of metadata

MINSEQE



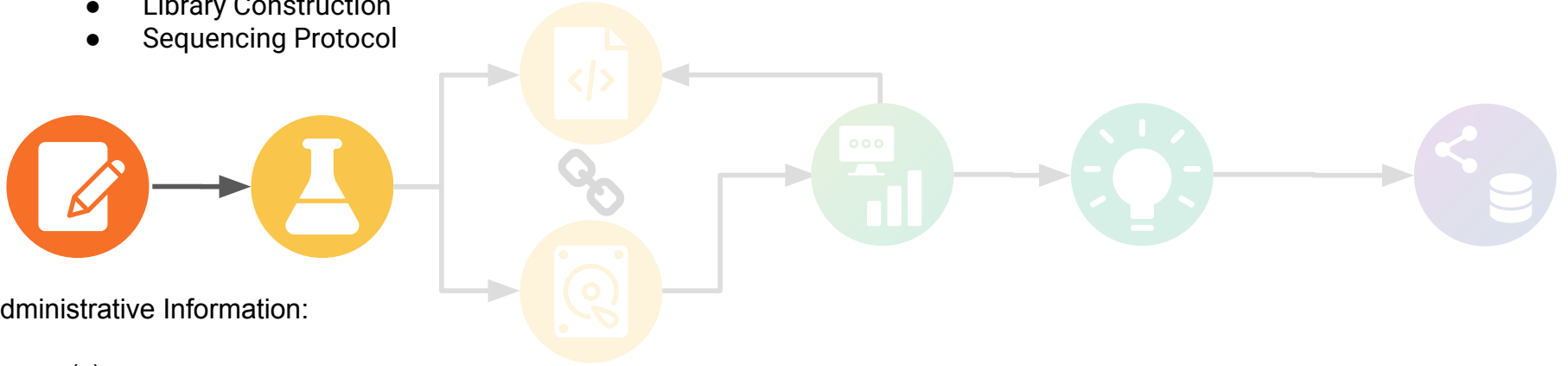
Administrative Information:

Person(s)
Organisations
Publications

Experimental Conditions/Design

Protocols:

- Treatment
- Sample Collection
- Growth
- Nucleic Acid Extraction
- Library Construction
- Sequencing Protocol



Administrative Information:

Person(s)

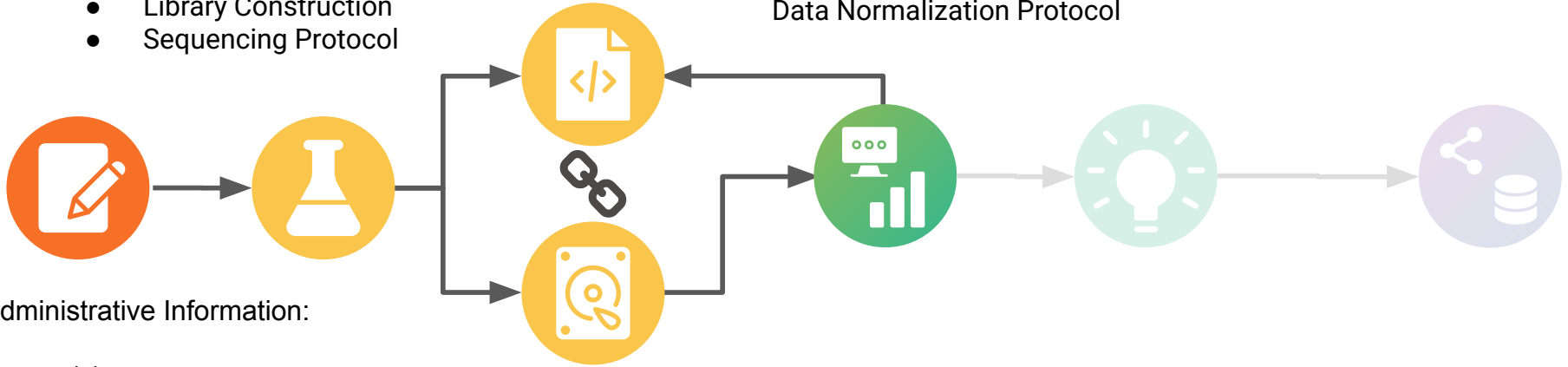
Organisations

Publications

Experimental Conditions/Design
Protocols:

- Treatment
- Sample Collection
- Growth
- Nucleic Acid Extraction
- Library Construction
- Sequencing Protocol

High-throughput Sequence Alignment Protocol
Data Normalization Protocol



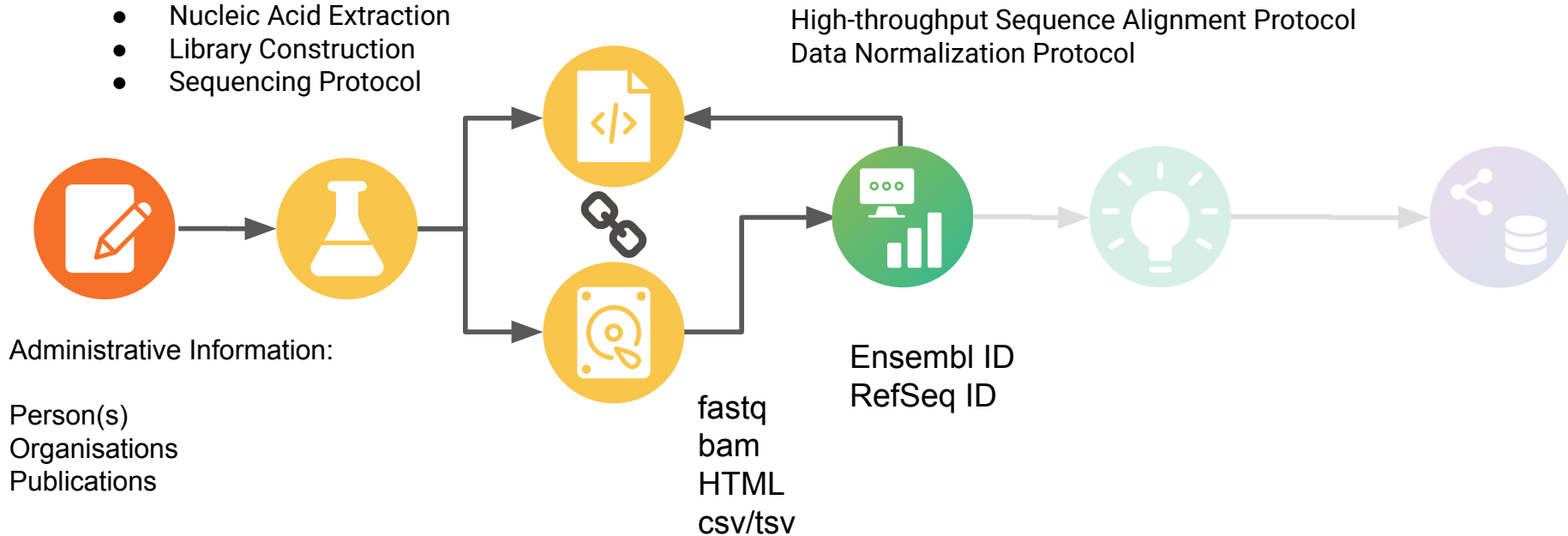
Administrative Information:

Person(s)
Organisations
Publications

Experimental Conditions/Design

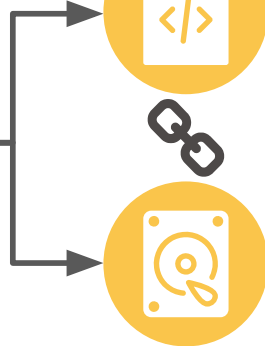
Protocols:

- Treatment
- Sample Collection
- Growth
- Nucleic Acid Extraction
- Library Construction
- Sequencing Protocol



Experimental Conditions/Design
Protocols

Taxon



High-throughput Sequence Alignment Protocol
Data Normalization Protocol

Ensembl ID
RefSeq ID

Administrative Information:

Person(s)
Organisations
Publications

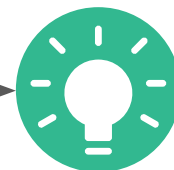
Experimental Conditions/Design
Protocols

Taxon



ArrayExpress

High-throughput Sequence Alignment Protocol
Data Normalization Protocol



Administrative Information:

Person(s)
Organisations
Publications

fastq
bam
csv/tsv

Ensembl ID
RefSeq ID



FAIRsharing Demo

+ 5 min for you to find an archive and its metadata standard



Licensing of research outputs



Korbinian Bösl
Data management coordinator
Centre for Digital Life & ELIXIR Norway
27 November 2025

Objectives

? Questions

- What is intellectual property/copyright/derivative work?
- What is free software/data?
- What types of licenses exist?

! Objectives

- Get familiar with terminology around licensing
- Discuss what is and is not derivative work

Intellectual property rights (IPRs)

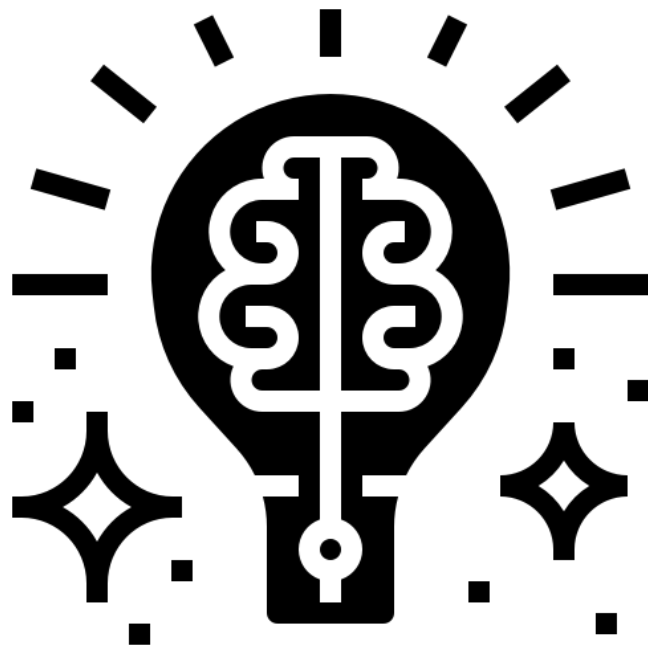
Patent: Protects novel, non-obvious, inventions

Copyright: creative products: software, writing, figures, photos, some datasets, this presentation

Database directive

Trademark: Protects a name/brand

Trade Secrets



CC-BY 3.0 Wichai.Wi

Intellectual property rights (IPRs)

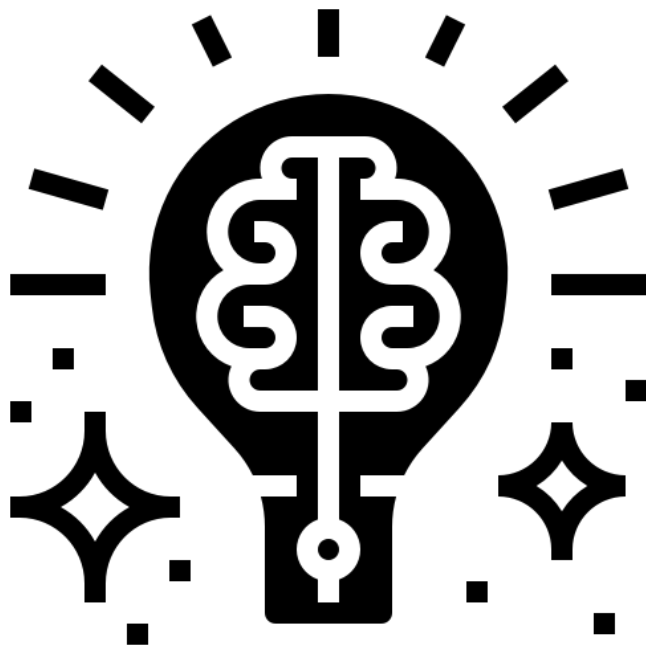
Patent: Protects novel, non-obvious, inventions

Copyright: creative products: software, writing, figures, photos, some datasets, this presentation

Database directive

Trademark: Protects a name/brand

Trade Secrets



CC-BY 3.0 Wichai.Wi

IPR often regulated in work contract

Software patents in Europe

(it's complicated)

The European Patent Convention states that software is not patentable.
in contrast with e.g. the US

The European Patents Office (EPO) grants software patents by declaring them as "computer implemented inventions".

Erosions: VICOM 1986, COMVIK 2002, ML guidelines 2018, Bentley 2021

Patents in Europe

Active process

No prior disclosure

Apply only in limited geographical scopes

Fees (varying by business size and scope)

Apply for max 20 years after application

Copyright in Europe

Creative products: software, writing, figures, photos,...

Details for software: Directive 2009/24/EC

Applies by default

Does not apply for facts

Economic rights vs. moral rights

Author lifetime + 70 years

[Commons:Copyright rules by territory/Norway](#)

<https://lovdata.no/pro/#document/NL/lov/2018-06-15-40>

<https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:32009L0024>

<https://lovdata.no/pro/#document/NLX2/avtale/avt-1992-05-02-1-v17>

Example WIPO: Genetic data

Raw sequence does not have copyright

Processed data might have

Patenting unaffected

Special consideration regarding traditional knowledge (Nagoya protocol)

EU Database directive

Defined additional (*sui generis*) IPR for:

“collection[s] of independent works, data or other materials [that does not fall under copyright] arranged in a systematic or methodical way and individually accessible by electronic or other means”

15 years by default

<https://eur-lex.europa.eu/eli/dir/1996/9/oj>

<https://lovdata.no/pro/#document/NLX2/avtale/avt-1992-05-02-1-v17>

Licensing

A licensor may grant a license under intellectual property laws to authorize a use (such as copying software or using a patented invention) to a licensee, sparing the licensee from a claim of infringement brought by the licensor.

Which of these is derivative work?

Download some code and use one of the functions in your code

Linking to libraries (static or dynamic), plug-ins, and drivers

Clean room design (you never see the code)

Extending code you got from somewhere

Rewriting code to a different programming language

Which of these is derivative work?

- ✓ Download some code and use one of the functions in your code
- ✗ Linking to libraries (static or dynamic), plug-ins, and drivers
- ✗ Clean room design (you never see the code)
- ✓ Extending code you got from somewhere
- ✓ Rewriting code to a different programming language

Why should I licence my research outputs?

Why should I licence my research outputs?



Legal security for users (Accessibility) + Funder requirement

Why should I licence my research outputs?



Legal security for users (Accessibility) + Funder requirement



Increase of willingness to reuse outputs (Reusability)

Why should I licence my research outputs?



Legal security for users (Accessibility) + Funder requirement



Increase of willingness to reuse outputs (Reusability)



Allows deposition/mirroring in 2nd databases (Findability)

Concepts in open licenses



Waive all your interests that may exist in your work

Copy left:

Concepts in open licenses



Waive all your interests that may exist in your work

Copy left:



Credit for the original creation

Concepts in open licenses



Waive all your interests that may exist in your work

Copy left:



Credit for the original creation



License new creations under identical terms

Concepts in open licenses



Waive all your interests that may exist in your work

Copy left:



Credit for the original creation



License new creations under identical terms



Non-commercial

Concepts in open licenses



Waive all your interests that may exist in your work

Copy left:



Credit for the original creation



License new creations under identical terms



Non-commercial



Cannot be shared with others in adapted form



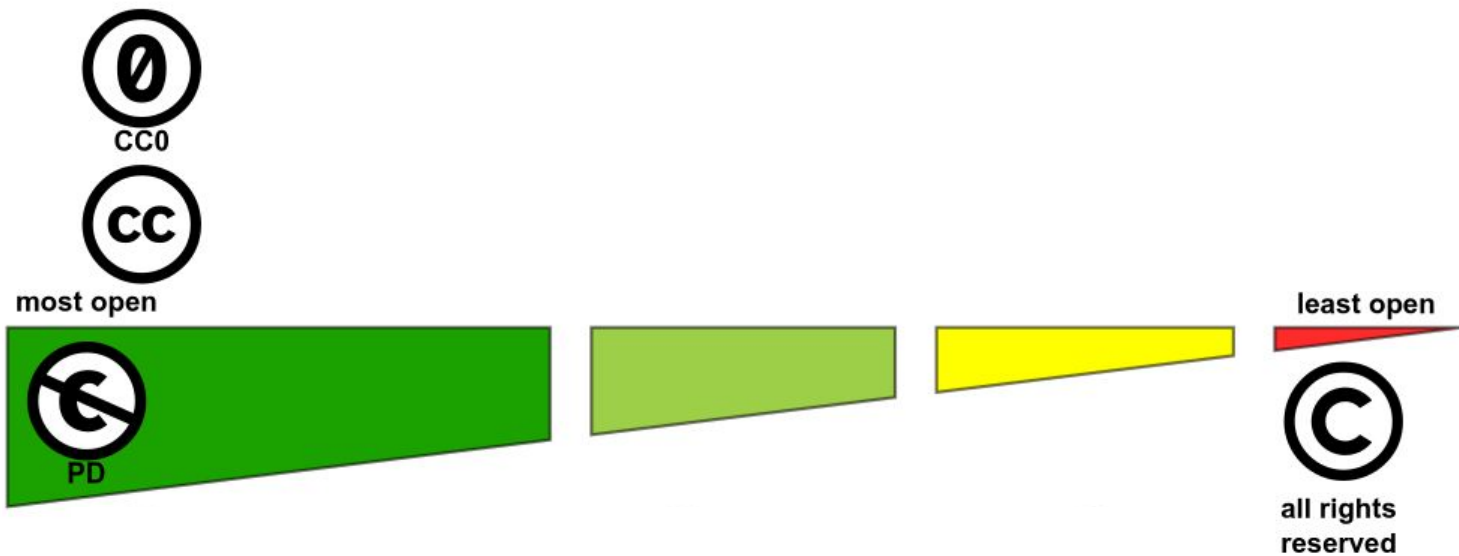
most open

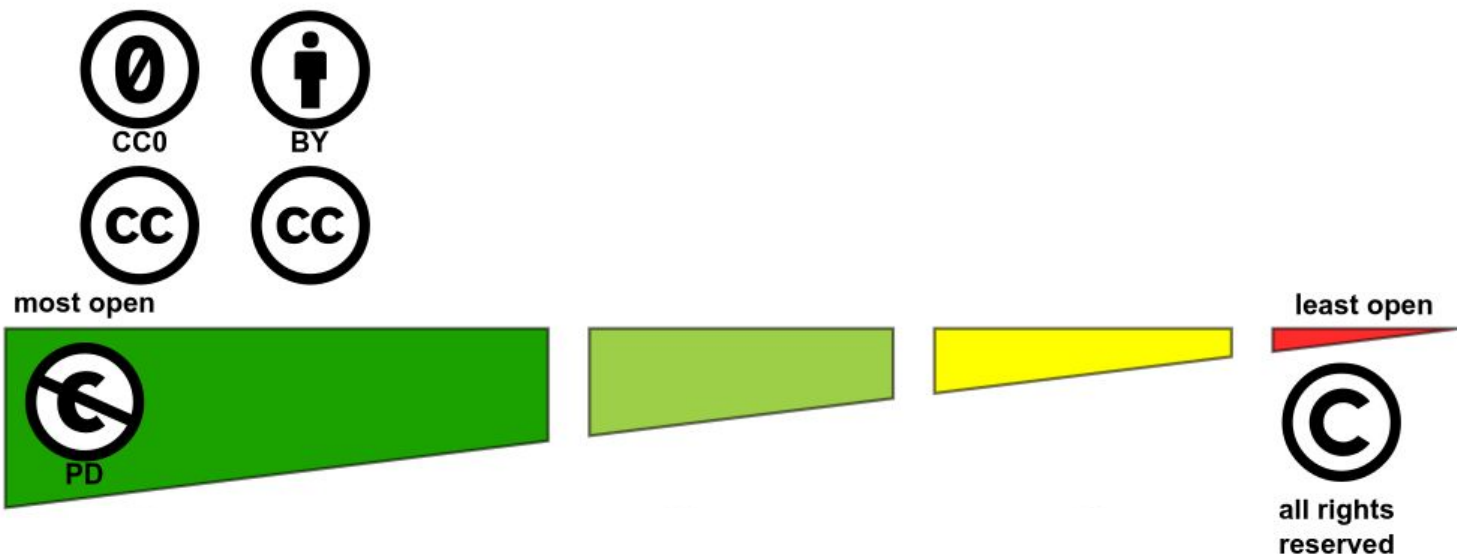


least open

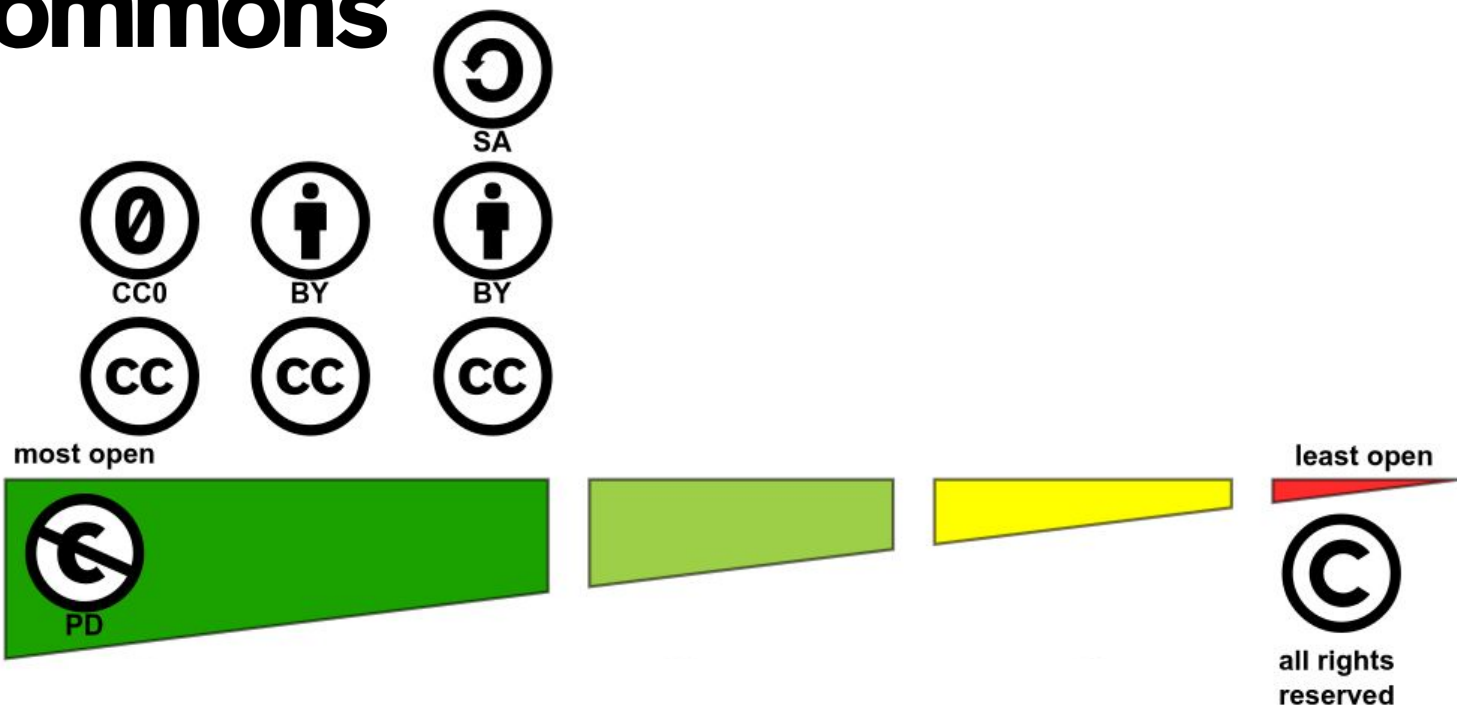


all rights
reserved

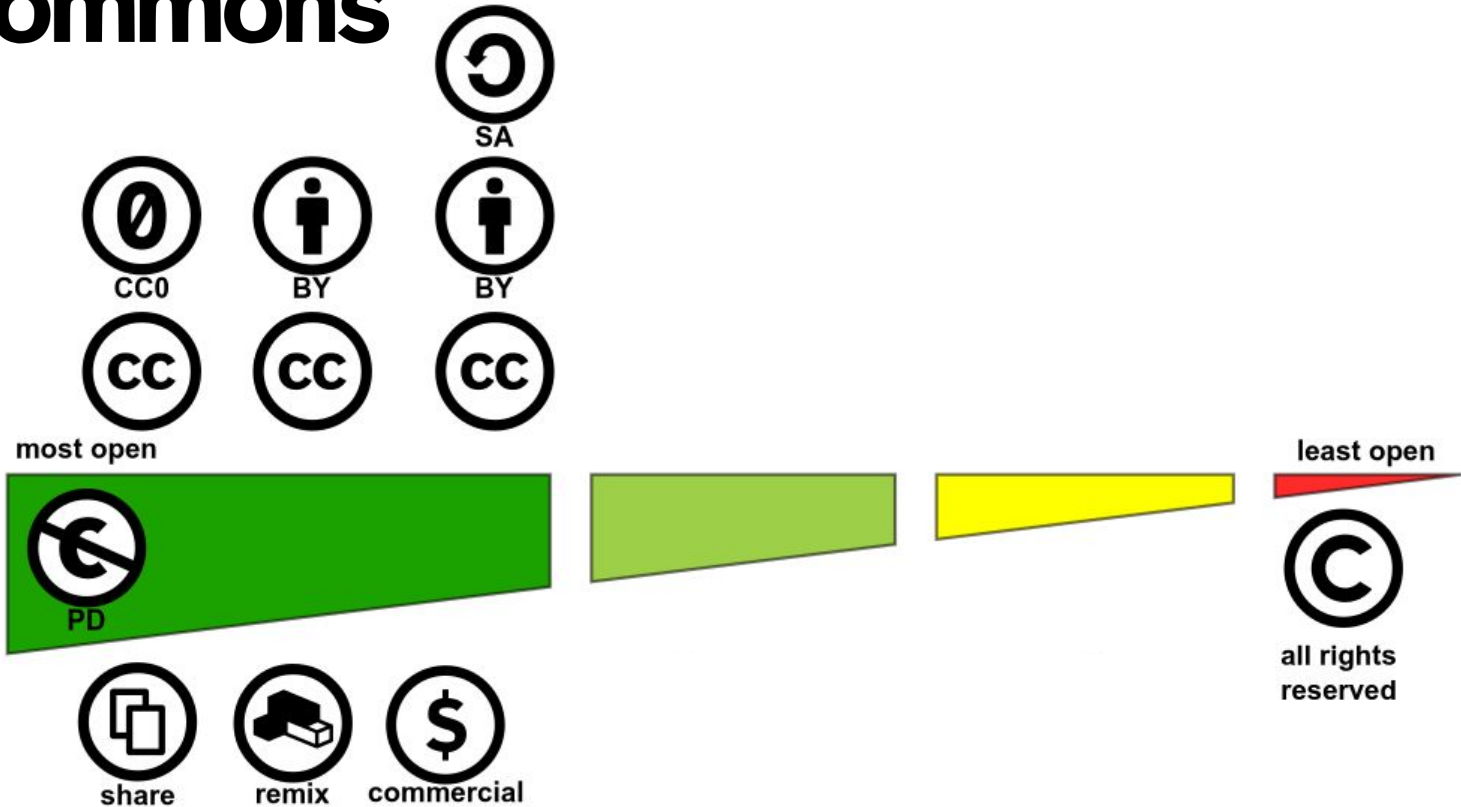


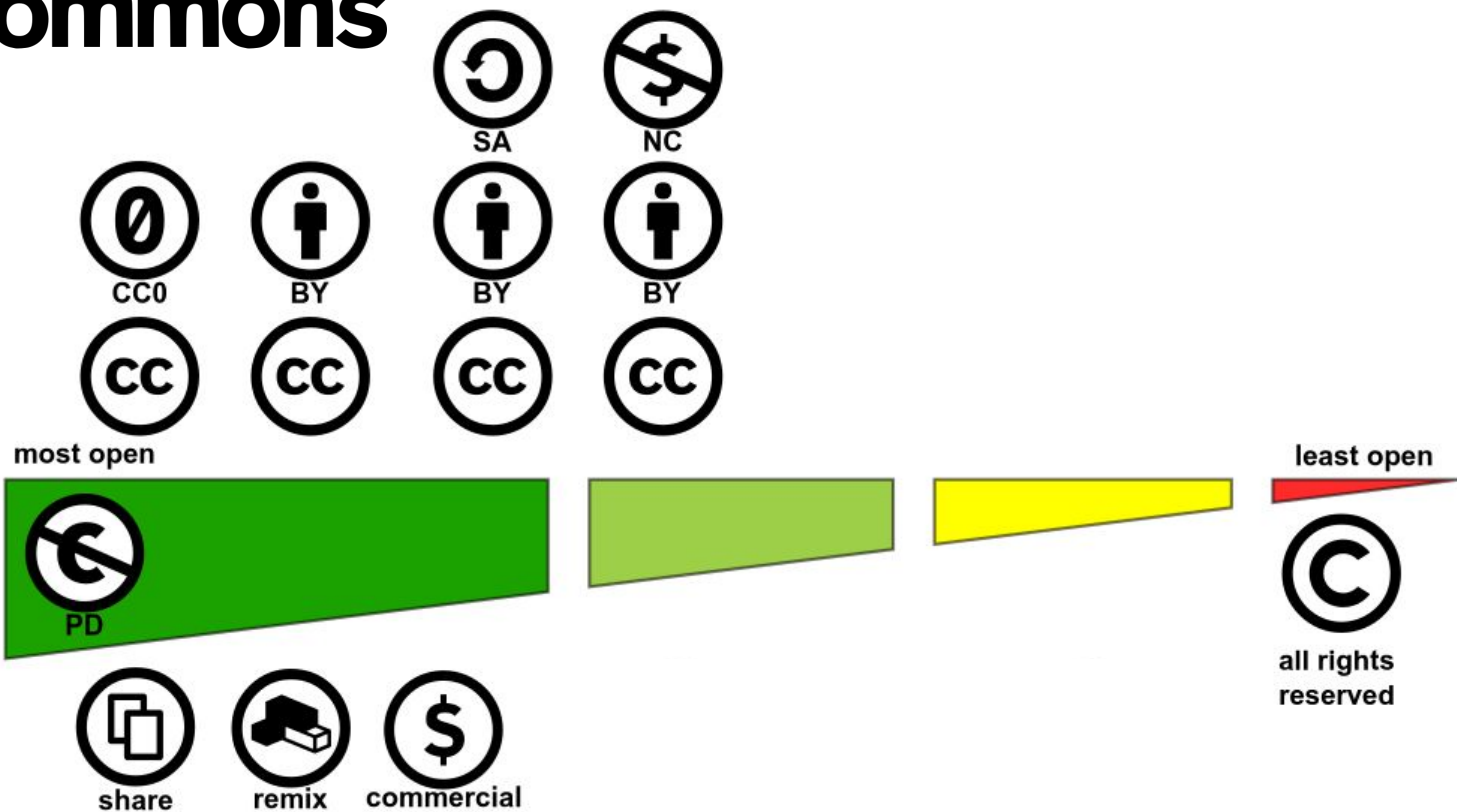


creative commons

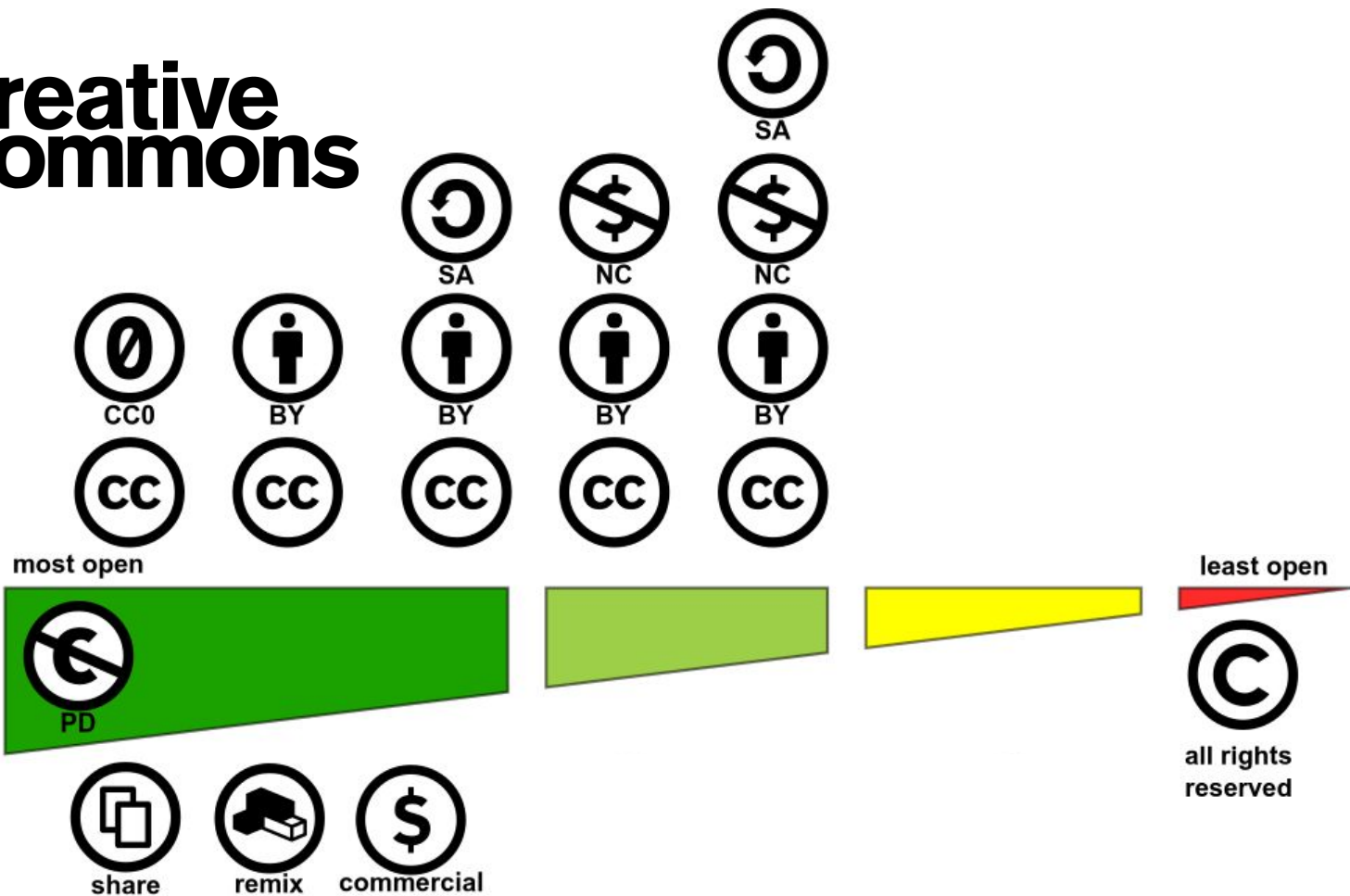


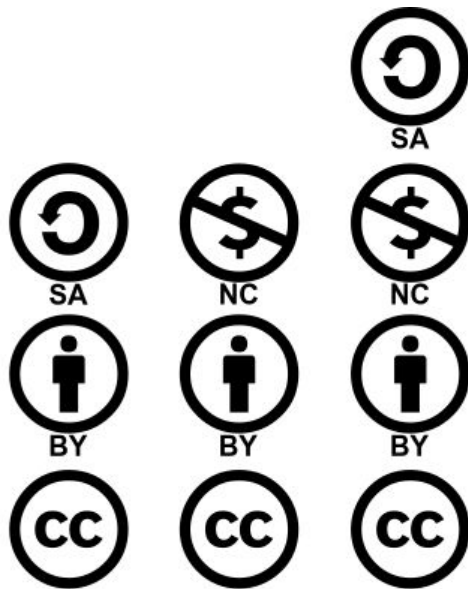
creative commons





creative commons

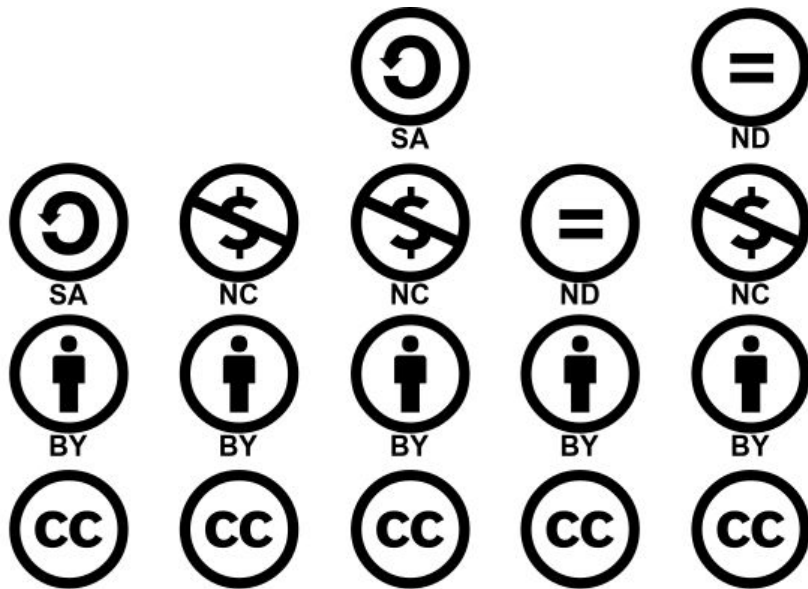




©

all rights
reserved



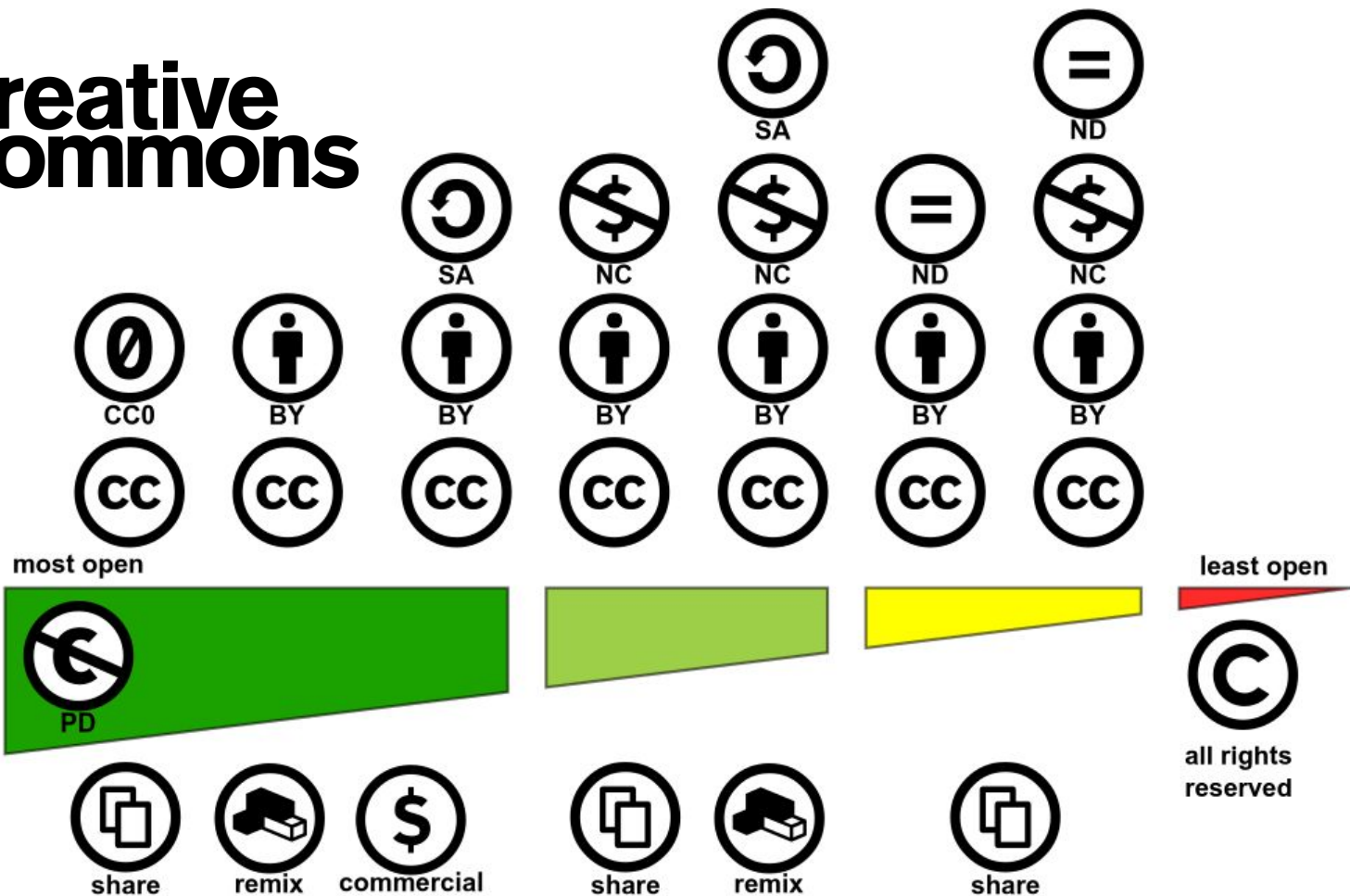


least open



all rights
reserved

creative commons

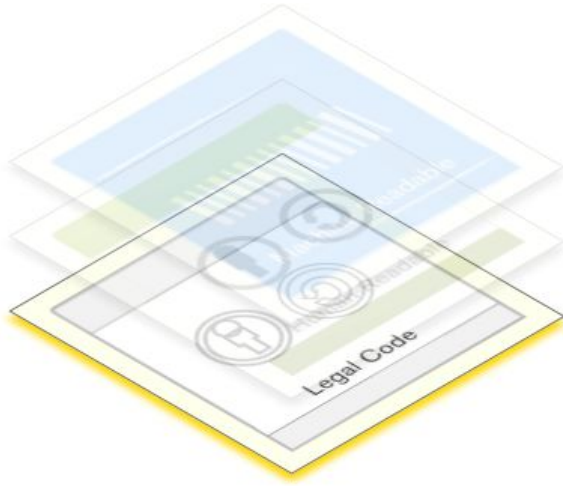




“open”

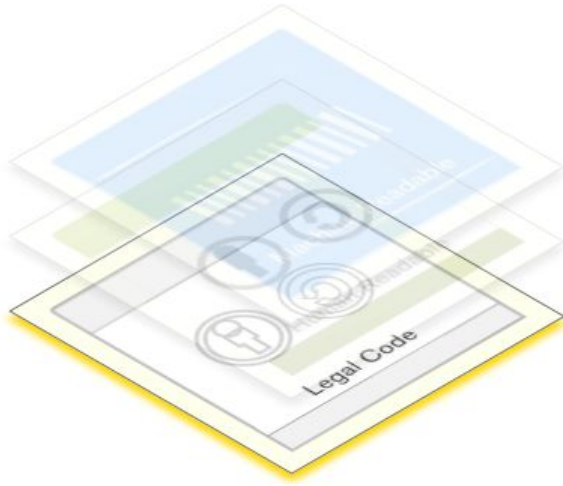


Legal code, harmonized for
national/international law





Legal code, harmonized for
national/international law





Legal code, harmonized for
national/international law

Human readable, understandable text



Legal code, harmonized for
national/international law

Human readable, understandable text

Machine readable html tag attachable
to metadata



Who to use this in practise?



Creative Commons - **What is licensed?**

Nathan Yergler, Alex Roberts - **Who is to be attributed?**

Licensed to the public under [CC BY 3.0 Unported](#) - **Which license?**



Challenges



Challenges



Multiple Attributions for several sources (license stacking)



Challenges



Multiple Attributions for several sources (license stacking)



Multiple incompatible source licenses



Multiple Attributions for several sources (license stacking)



Multiple incompatible source licenses



Legal commercial definition (e.g. use by journals)



Multiple Attributions for several sources (license stacking)



Multiple incompatible source licenses



Legal commercial definition (e.g. use by journals)



Unintentional restrictive

Open Data Commons

Databases are different to simple data (e.g. EU-copyright)

Open Data Commons

Databases are different to simple data (e.g. EU-copyright)



Open Data Commons Public Domain Dedication and License

(PDDL)

Open Data Commons

Databases are different to simple data (e.g. EU-copyright)



Open Data Commons Public Domain Dedication and License

(PDDL)



Open Data Commons Attribution License (ODC-BY)

Open Data Commons

Databases are different to simple data (e.g. EU-copyright)



Open Data Commons Public Domain Dedication and License
(PDDL)



Open Data Commons Attribution License (ODC-BY)



Open Data Commons Open Database License (ODbL)

Norwegian Licence for Open Government Data (NLOD) 2.0



Digitaliseringsdirektoratet
data.norge.no

Norwegian Licence for Open Government Data (NLOD) 2.0



Digitaliseringsdirektoratet
data.norge.no

A licence compatible by contract shall mean the following licences:

for all information: Open Government Licence (version 1.0, 2.0 and 3.0), **Creative Commons Attribution Licence (international version 4.0 and norwegian version 4.0)**

for those parts of the information which do not constitute databases: **Creative Commons Attribution Licence (generic version 1.0, 2.0, 2.5 and unported version 3.0) and Creative Commons Navngivelse 3.0 Norge**

for those parts of the information which constitute databases: **Open Data Commons Attribution License (version 1.0).**

Norwegian Licence for Open Government Data (NLOD) 2.0



Digitaliseringsdirektoratet
data.norge.no



A licence compatible by contract shall mean the following licences:

for all information: Open Government Licence (version 1.0, 2.0 and 3.0), **Creative Commons Attribution Licence (international version 4.0 and norwegian version 4.0)**

for those parts of the information which do not constitute databases: **Creative Commons Attribution Licence (generic version 1.0, 2.0, 2.5 and unported version 3.0) and Creative Commons Navngivelse 3.0 Norge**

for those parts of the information which constitute databases: **Open Data Commons Attribution License (version 1.0).**

Open Source Software licenses

Special considerations for Software

Liability

Warranty

Modifications

Network use = Distribution?



Open Source Software licenses





Open Source Software licenses



MIT license

GNU
AGPLv3





Open Source Software licenses

← BSD licenses →



MIT license

Apache 2.0

GNU LGPLv3

GNU GPLv3

GNU
AGPLv3





Open Source Software licenses

← BSD licenses →



MIT license

GNU LGPLv3

Apache 2.0

GNU

AGPLv3

GNU GPLv3



<https://opensource.org/licenses>

<https://choosealicense.com/>



GitHub

Icons [CC-BY 4.0](https://creativecommons.org/licenses/by/4.0/)
<https://creativecommons.org/>

Indices & Controlled vocabularies



License List - machine readable IDs

on file level

SPDX-License-Identifier: MIT



**Approved Licenses - also EC definition
of open source license**

<https://spdx.org/licenses/>
<https://opensource.org/licenses>

Permissive

MIT
BSD

Permissions

- Commercial use
- Distribution
- Modification
- Private use

Conditions

- License and copyright notice

Limitations

- Liability
- Warranty

weak

Copyleft

LGPL
MPL

Permissions

- Commercial use
- Distribution
- Modification
- Patent use
- Private use

Conditions

- Disclose source
- License and copyright notice
- Same license (library)
- State changes

Limitations

- Liability
- Warranty

strong

Copyleft

GPL
AGPL

Permissions

- Commercial use
- Distribution
- Modification
- Patent use
- Private use

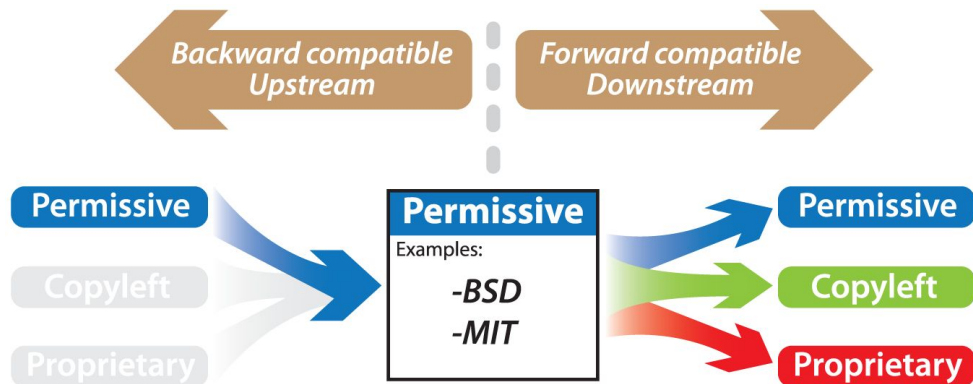
Conditions

- Disclose source
- License and copyright notice
- Same license
- State changes

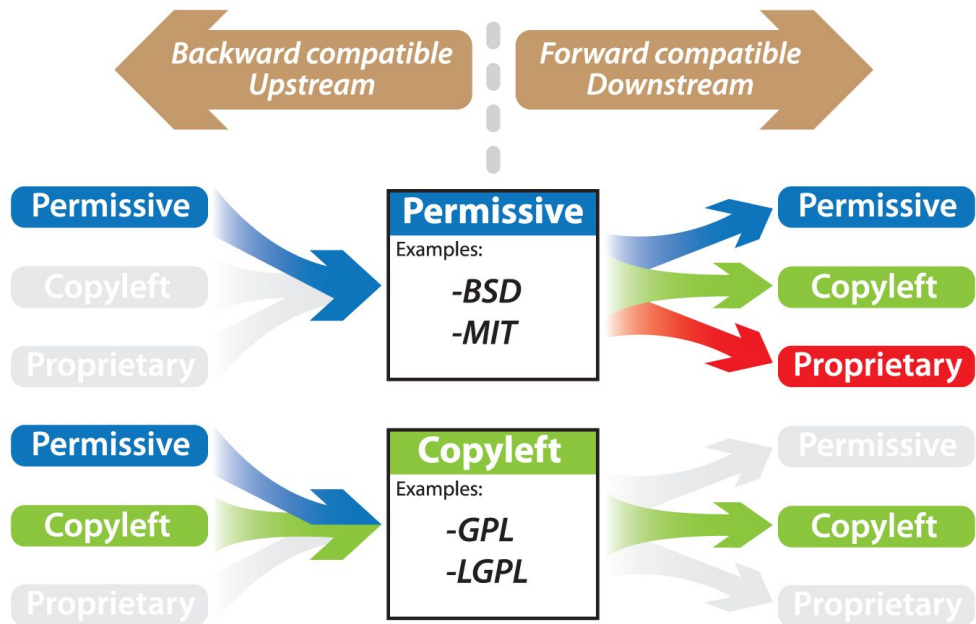
Limitations

- Liability
- Warranty

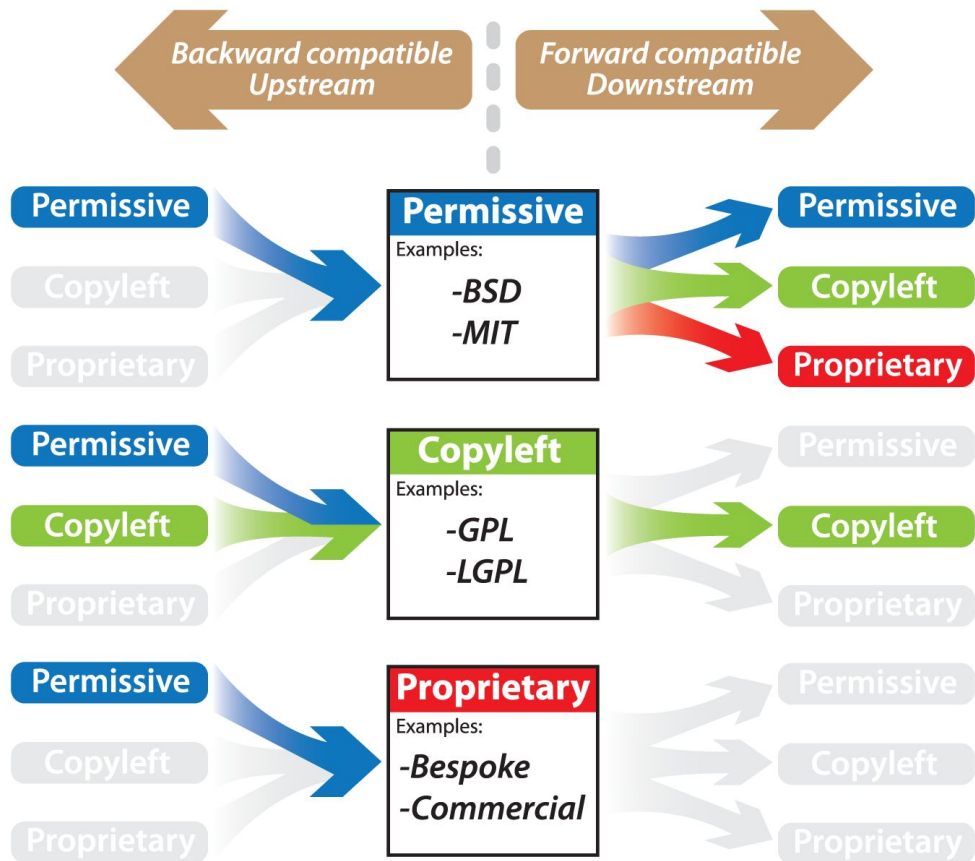
Open Source Software licenses



Open Source Software licenses



Open Source Software licenses



Does it hold in court?

Yes!



Busybox GPL lawsuits 2007-2009



Free Software Foundation, Inc. v. Cisco Systems, Inc., 2009

...

Does it hold in court?

Yes!

CREATIVE COMMONS ATTRIBUTION- SHAREALIKE LICENSE ENFORCED IN GERMANY

by [mike](#)

The Creative Commons Attribution-ShareAlike 3.0 Unported license (BY-SA) has been enforced by a judicial injunction in Germany. Legal analysis will be added to our [case law database](#) in the coming days. Till Jaeger [reported the case](#) (in German; [English machine](#)

Unintended consequences



klmr commented on 10 Jul 2018 • edited



The GPLv3 license [creates some nontrivial problems](#) for use of this code in scientific software. This might be reason enough to **reconsider whether GPL is the best fit for this project**, and whether its restrictions are intended by the authors, or merely accidental.

For reference, [Titus C Brown](#) has argued forcefully and influentially that copyleft licenses are non-open, and therefore impede Open Science.

In fact, Lior Pachter, one of the major proponents of non-free software in bioinformatics has subsequently conceded that [he was wrong](#) about the non-free licensing of the Kallisto software.

(This is related to, but distinct from [#478](#).)



10

Unintended consequences



Replying to @karinlag

The proper license for Nextflow pipelines is GPL because your program (the pipeline) is dynamically linked with the Nextflow runtime that's distributed as GPL. gnu.org/licenses/gpl-f...

9:42 AM · Jul 10, 2018 · Twitter Web Client

...

However, this was not our original intention. We don't consider workflow applications to be subject to the GPL copyleft obligations of the GPL even though they may link dynamically to Nextflow functionality through normal calls and we are not interested to enforce the license requirement to third party workflow developers and organizations.

Unintended consequences

Goodbye zero, Hello Apache!

 Paolo Di Tommaso  24 October 2018

Today marks an important milestone in the Nextflow project. We are thrilled to announce three important changes to better meet users' needs and ground the project on a solid foundation upon which to build a vibrant ecosystem of tools and data analysis applications for genomic research and beyond.

Apache license

Nextflow was originally licensed as GPLv3 open source software more than five years ago. GPL is designed to promote the adoption and spread of open source software and culture. On the other hand it has also some controversial side-effects, such as the one on [derivate works](#) and [legal implications](#) which make the use of GPL released software a headache in many organisations. We have previously discussed these concerns in [this blog post](#) and, after community feedback, have opted to change the project license to Apache 2.0.

This is a popular permissive free software license written by the [Apache Software Foundation](#) (ASF). Software distributed with this license requires the preservation of the copyright notice and disclaimer. It allows the freedom to use the software for any purpose, to distribute it, to modify it, and to distribute modified versions of the software without dictating the licence terms of the resulting applications and derivative works. We are sure this licensing model addresses the concerns raised by the Nextflow community and will boost further project developments.

<https://www.nextflow.io/blog/2018/goodbye-zero-hello-apache.html>

1. What is the StackOverflow license for code you copy and paste?
2. A journal requests that you release your software during publication. You have copied a portion of the code from another package, which you have forgotten. Can you satisfy the journal's request?
3. You want to fix a bug in a project someone else has released, but there is no license. What risks are there?
4. How would you ask someone to add a license?
5. You incorporate MIT, GPL, and BSD3 licensed code into your project. What possible licenses can you pick for your project?
6. You do the same as above but add in another license that looks viral. What possible licenses can you use now?
7. Do licenses apply if you don't distribute your code? Why or why not?
8. Which licenses are most/least attractive for companies with proprietary software?

DO

Always add a (standard) license

Ideally: File level license indication

Be aware of the IPR situation of the code you (re-)use

DON'T

Invent your own license

Compliance monitoring

Costs of data management

Data analysis

Data brokering

Data management coordination

Data management plan

Data organisation

Data security

Data sensitivity

Data provenance

Data publication

Data quality

Data storage

Data transfer

Documentation and metadata

Ethical aspects

Existing data

GDPR compliance

Identifiers

Licensing

Machine actionability



Link to RDMkit:
<https://rdmkit.elixir-europe.org/>

Next:

[Digital Life 2021: Know-how workshop on IPR in digital biotech](#)

[Data life cycle: Sharing | RDMkit](#)

[Your tasks: Licensing | RDMkit](#)

[EUDAT license chooser for data and software](#)

[Choose a License](#)

[Licenses & Standards | Open Source Initiative](#)



<https://coderefinery.org/>

1. As indicated [here](#), all publicly accessible user contributions are licensed under [Creative Commons Attribution-ShareAlike](#) license. See Stackoverflow [Terms of service](#) for more detailed information.
2. “Standard” licensing rules apply. So in this case, you would need to remove the portion of code you have copied from another package before being able to release your software.
3. By default you are not authorized to use the content of a repository when there is no license. And derivative work is also not possible by default. Other risks: it may not be clear whether you can use and distribute (publish) the bugfixed code. For the repo owners it may not be clear whether they can use and distributed the bugfixed code. However, the authors may have forgotten to add a license so we suggest you to contact the authors (e.g. make an issue) and ask whether they are willing to add a license.
4. As mentioned in 3., the easiest is to fill an issue and explain the reasons why you would like to use this software (or update it).
5. Combining software with different licenses can be tricky and it is important to understand compatibilities (or lack of compatibilities) of the various licenses. GPL license is the most protective (BSD and MIT are quite permissive) so for the resulting combined software you could use a GPL license. However, re-licensing may not be necessary.
6. Derivative work would need to be shared under this viral license (e.g. AGPL or GPL), unless the components are only plugins or libraries.
7. If you keep your code for yourself, you may think you do not need a license. However, remember that in most companies/universities, your employer is “owning” your work and when you leave you may not be allowed to “distribute your code to your future self”. So the best is always to add a license!
8. The least attractive licenses for companies with proprietary software are licenses where you would need to keep an open license when creating derivative work. For instance GPL and AGPL. The most attractive licenses are permissive licenses where they can reuse, modify and relicense with no conditions. For instance MIT, BSD and Apache License.

Thank you!



elixir-norway.org
digitallifenorway.org



contact@elixir.no
contact@digitallifenorway.org
korbinian.bosl@uib.no



@elixirnorway
@DigitaltLiv
@BoslKorbinian



**Except where otherwise noted, this work is licensed under a
Creative Commons Attribution 4.0 International License**

<https://creativecommons.org/licenses/by/4.0/>

Thank you!



elixir-norway.org
digitallifenorway.org



contact@elixir.no
contact@digitallifenorway.org
korbinian.bosl@uib.no



@elixirnorway
@DigitaltLiv
@BoslKorbinian



**Except where otherwise noted, this work is licensed under a
Creative Commons Attribution 4.0 International License**

<https://creativecommons.org/licenses/by/4.0/>