**Licenses for Research Data**

**What License Should Be Applied to the Research Data?**

It depends on what rights protect your research data. Based on the guide "How do I know if my research data is protected?":

- For Research Data Qualifying as a Work (e.g., Literary Works like Journal Articles or Software):

  - Recommended License: CC BY 4.0

  - Compatible Options: Share Alike (SA) - Compatible with Open Access and Plan S licensing guidance.

  - Not Recommended:

    - Non-commercial - Not Open Access compliant.

    - Non-derivative - Should be avoided unless necessary, as it may complicate Open Access.

- For Research Data as a Database or Dataset (Unstructured Data Not Meeting Database Definition):

  - Recommended License: CC0 - Waives all rights in the database.

- Important Considerations:

  - CC licenses address copyright and related matters but do not cover personal data, which must be handled separately.

What is a Creative Commons License?

- Overview: Creative Commons is a global non-profit organization offering legal tools to promote sharing and reuse of works of authorship.

- License Levels: Creative Commons licenses are readable at three levels:

  - Legal

  - Machine (metadata)

  - Human (non-legal descriptions)

- License Tool: Creative Commons provides a tool to help determine the best license for your needs.

- Restrictions: More restrictive CC licenses may not meet Open Access requirements, particularly those limiting commercial use.

Creative Commons (CC) Website

- Website: [creativecommons.org](https://creativecommons.org)

- Resources: Detailed information about different Creative Commons licenses (e.g., CC BY, CC0) and their application to research content.

Open Access Journal Platforms

- PubMed Central: Open access research articles, many under CC licenses.

- arXiv: Preprints in science, mathematics, and computer science, with clear usage rights.

- DOAJ (Directory of Open Access Journals): Lists open-access journals with permissive licenses.

Public Domain Resources

- Project Gutenberg: Texts that have entered the public domain.

- HathiTrust Digital Library: Public domain content and academic resources.

Institutional Repositories

- Harvard DASH: Harvard University’s open-access repository.

- MIT DSpace: MIT’s institutional repository.

Data and Research Communities

- Zenodo: Open-access repository developed by CERN, offering research data with clear licensing.

- Kaggle Datasets: Public datasets with specified licenses for reuse.

----------------------------------------------------------------------------------------------------

Key Open-Source Licenses

1. Creative Commons Attribution-ShareAlike (CC BY-SA)

- Description: Allows distribution, remixing, adaptation, and commercial use with credit to the original creator and identical licensing of new works.

- Key Requirements:

  - Must give appropriate credit.

  - Derivatives must be licensed under CC BY-SA.

- Common Use: Used by Wikipedia and similar projects to keep derivative works free and open.

2. GNU General Public License (GPL)

- Description: Copyleft license for software, permitting free use, modification, and distribution, with the requirement to release modified versions under GPL.

- Key Requirements:

  - Derivative software must also be licensed under GPL.

  - Must provide access to the source code.

- Common Use: Widely used in open-source software (e.g., Linux, WordPress).

3. GNU Affero General Public License (AGPL)

- Description: Variant of GPL with an additional requirement for network use, requiring source code availability to users interacting with the software.

- Key Requirements:

  - Software distributed or offered as a service must be released under AGPL.

  - Source code must be accessible to users interacting with the software.

- Common Use: Used in web-based software, particularly for SaaS applications.

4. Creative Commons Attribution-NonCommercial-ShareAlike (CC BY-NC-SA)

- Description: Allows remixing, adaptation, and non-commercial use with credit and identical licensing of new works.

- Key Requirements:

  - Derivatives must be non-commercial.

  - Attribution and ShareAlike clause must be maintained.

- Common Use: For creators seeking to keep their work non-commercial and open.

5. Mozilla Public License (MPL)

- Description: Middle-ground license requiring that modified files be shared under MPL terms, but allows other parts of the project to be under different licenses.

- Key Requirements:

  - Modifications to original files must be licensed under MPL.

  - Other parts of the project can be under a different license.

- Common Use: Typically used in software where flexibility and openness of modified files are important.

6. Open Database License (ODbL)

- Description: Applies to databases, allowing free sharing, modification, and use, with a requirement that derivative databases be shared under the same terms.

- Key Requirements:

  - Share-Alike requirement for modified databases.

- Common Use: Used in open data projects like OpenStreetMap.

----------------------------------------------------------------------------------------------------

Additional Licenses

1. MIT License

- Description: A permissive license allowing use, modification, and distribution, including for commercial purposes, as long as the original license and copyright notice are included.

2. BSD License (Berkeley Software Distribution)

- Description: Allows use, modification, and distribution with minimal restrictions. Requires attribution but imposes no further conditions on usage.

3. Open Data Commons Licenses

- ODC Public Domain Dedication and License (PDDL): Allows unrestricted use of data.

- ODC Attribution License (ODC-By): Permits use, modification, and distribution of data with attribution to the original source.

- ODC Open Database License (ODbL): Requires that derivative databases remain open and available under the same terms.

4. EUPL (European Union Public License)

- Description: Allows use, modification, and redistribution with provisions ensuring compatibility with other open-source licenses.

5. Proprietary Licenses with Explicit AI Usage Terms

- Description: Some publishers or authors may offer proprietary licenses explicitly permitting AI training or data mining. Always check terms for AI model training usage.

6. Dataset-Specific Licenses

- Description: Some research datasets come with custom licenses explicitly allowing or restricting AI-related uses (e.g., Common Crawl license, Hugging Face’s dataset licenses).

7. CC BY-NC (Non-Commercial)

- Description: Allows non-commercial use with proper attribution, but restricts commercial use.

8. Fair Use (U.S.)

- Description: Under certain conditions, research papers can be used for AI model training under fair use. This is legally complex and should be approached with caution.

9. MPL (Mozilla Public License)

- Description: Balances permissive and copyleft licenses, allowing free use and modification, but requiring modified files to be distributed under the same license.

----------------------------------------------------------------------------------------------------

Important Considerations

- Always verify the terms of each license, especially regarding AI training and commercial applications.