Data Usage Licenses

Amelia Harrison

FAIR Data Practices for Omics Analysis
Presented: April 22, 2022

Disclaimer

We are not lawyers. The content of this presentation is not legal advice. The purpose of this presentation is to provide you with information and resources that will help you understand the goals and importance of releasing your data with a usage license.

If you have are ever unsure about usage licenses for your data or have other legal questions regarding your research, please contact UD's Office of Research and Regulatory Affairs: https://research.udel.edu/regulatory-affairs/

About data use licenses

What is a data use license?

- “A license is a legal document that outlines what people are able to do with a work, project, or collection of information.”

Why do I need a license?

- By default in the US, you own and retain all rights to your work
  - “Your work is under copyright protection the moment it is created and fixed in a tangible form that it is perceptible either directly or with the aid of a machine or device.”
- Releasing data without a license prohibits use of your data

Terminology

Commercial

- Means there is an intent to sell/make a profit from materials

Proprietary

- Relating to property or ownership

Public domain

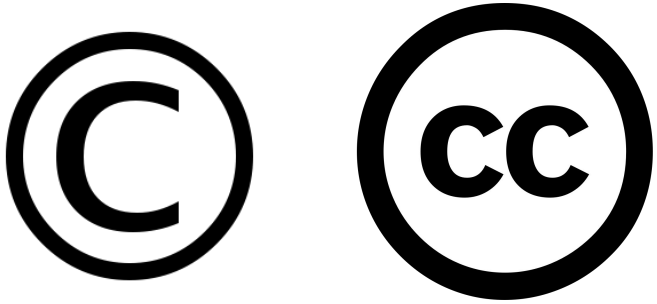
- Creative materials not protected by intellectual property laws

Derivative work

- A work based on one or more pre-existing works
- This one seems straightforward, but gets really tricky...

License Types

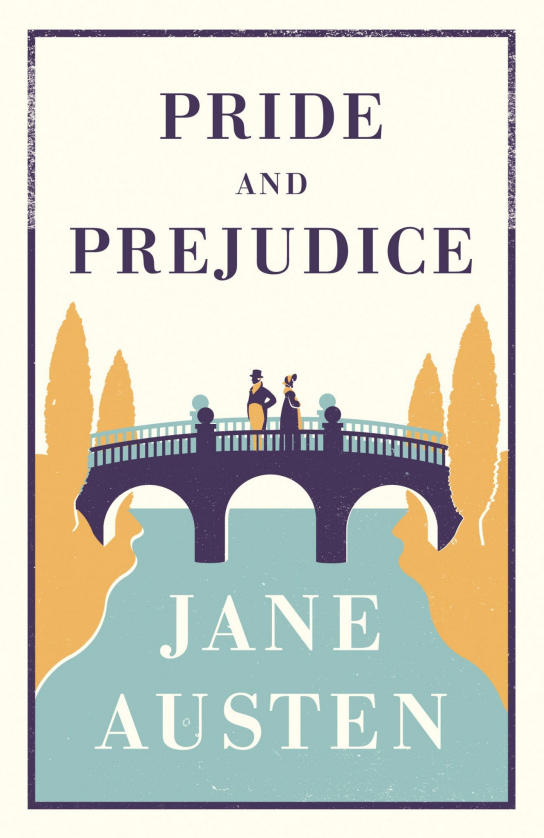
(In order of permissiveness)

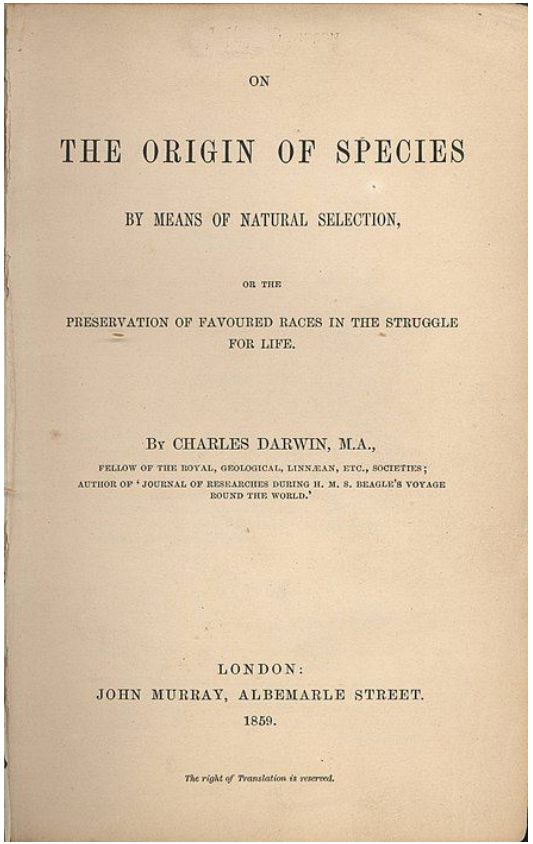


Permissive

- Permissive licenses permit reuse, transformation, and redistribution, allowing for attribution
- Allows commercial use
- Examples:
  - Creative Commons Attribution 4.0 International license (CC-BY 4.0)
  - MIT License (MIT) – For software
  - Public domain

Public domain




ON

THE ORIGIN OF SPECIES

BY MEANS OF NATURAL SELECTION,

OR THE

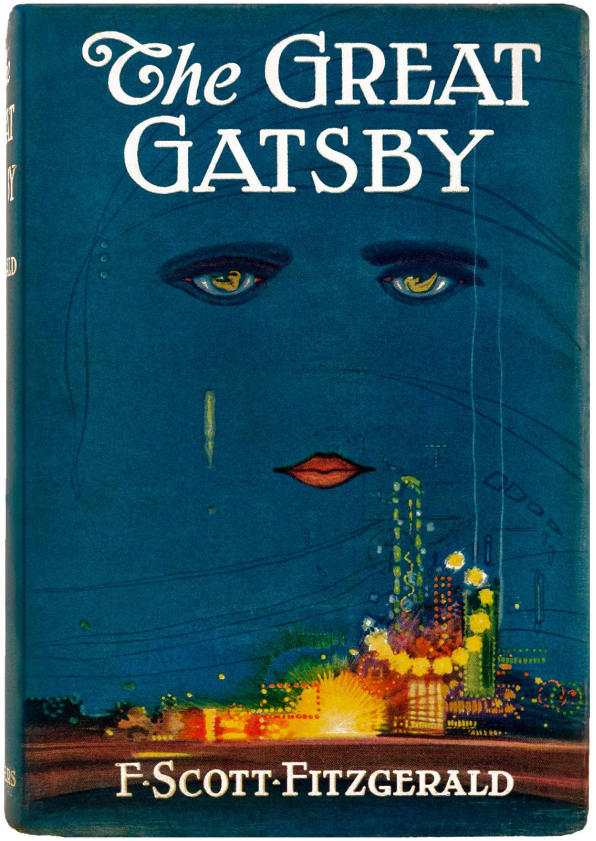
PRESERVATION OF FAVOURED RACES IN THE STRUGGLE FOR LIFE.

By CHARLES DARWIN, M.A.,

FELLOW OF THE ROYAL, GEOLOGICAL, LINNÆAN, ETC., SOCIETIES;
AUTHOR OF 'JOURNAL OF RESEARCHES DURING H. M. S. BEAGLE'S VOYAGE ROUND THE WORLD.'

LONDON:
JOHN MURRAY, ALBEMARLE STREET.
1859.

The right of Translation is reserved.



ShareAlike/Copyleft

- Allow for reuse, transformation, and redistribution. However, new contributions derived from the original data resource must be distributed under the same license.
- Allows commercial use
- Examples:
  - Creative Commons Attribution-ShareAlike 4.0 International license (CC BY-SA 4.0)
  - GNU General Public License v3.0 (GNU GPL 3.0) – for software

Restrictive

- Restrictive licenses provide more permissions compared to data resources wherein all copyrights have been reserved by the provider, but still include terms that may hinder data integration and reuse.
- Usually do not allow commercial use
- Usually do not allow redistribution after remixing/transforming/building upon a work
    - Therefore, they are bad licenses for scientific data
- Examples:
    - Creative Commons Attribution-NoDerivatives 4.0 International (CC BY-ND 4.0)

Private Pool

- A "private pool" license is one where the resource requires data users to add their own data to the pool, or limits the accessibility of derivative data to others that have also joined the pool.
- Conceptually similar to some copyleft licenses, but without the public "open" component.

Copyright

- “All rights reserved”
- Asserts a resource provider’s exclusive copyrights
- This is the default for anything you produce (including data)
    - Under current US copyright law, creators do not have to explicitly register or copymark their creations to claim their exclusive rights

How do I make my data most accessible?

- The most accessible option is to release the data to the public domain
  - But, some jurisdictions don’t allow this
  - Not an attractive option to everyone
- Next most open is CC-BY 4.0
- CC-BY-SA 4.0 is probably the most restrictive you want to use
  - Materials need to be at least this open for Wikipedia
- Remember, a lot of research is funded by public money

General rules to follow

- Always check to see if your funding specifies the use of a particular license
- Do not use CC licenses for software
  - “CC licenses do not contain specific terms about the distribution of source code”
  - Instead, use MIT, Apache, or GPL licenses
- Do not try to write your own license
  - Licenses should be written by legal experts
- If you are unsure about the best license for your needs, talk to a legal professional

Licenses can be revoked or altered

However, the license in place at the time information is acquired is the one that dictates what individuals can do with the licensed material.

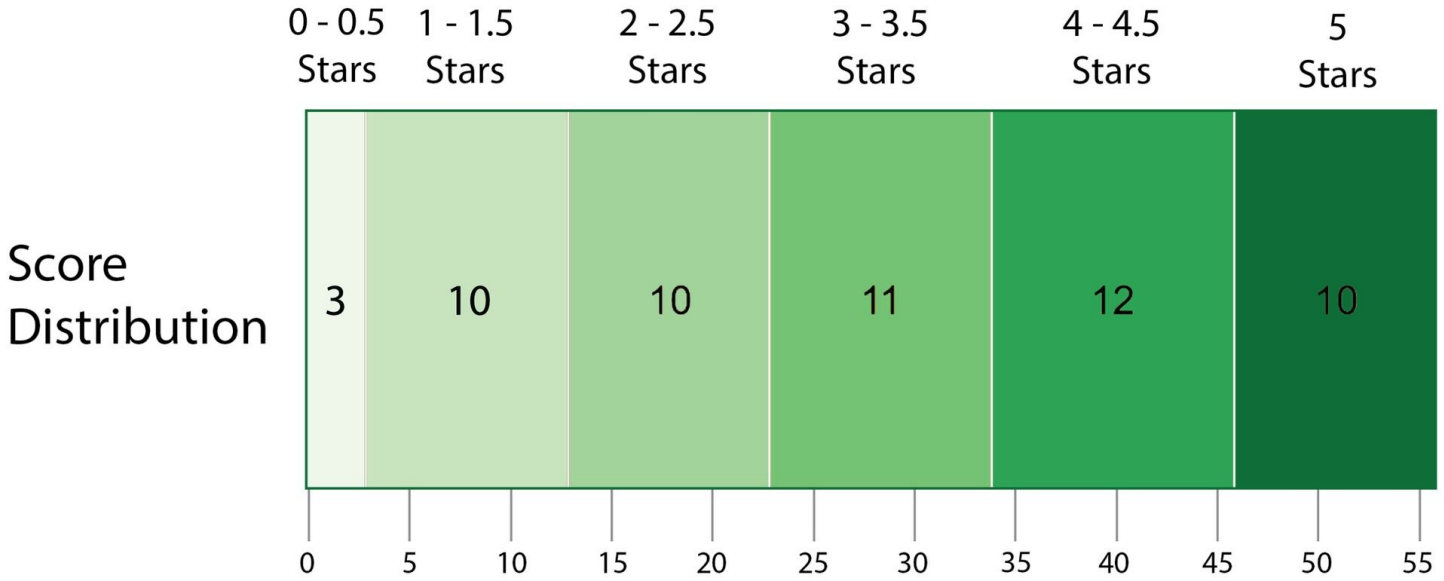
Data licenses and scientists: A case study

An analysis and metric of reusable data licensing practices for biomedical resources (2019)

- Assessed the reusability of datasets from 56 scientific data sources
- Used the (Re)usable Data Project's (RDP) guidelines
    - Star scale from 0-5

41% of resources scored below 3 stars

Indicates significant barriers to reusability



Sources and other resources

- US copyright FAQ: https://www.copyright.gov/help/faq/faq-general.html
- Creative commons: https://creativecommons.org
  - And their FAQ: https://creativecommons.org/faq/
- Reusable data project: http://reusabledata.org/index.html#home
  - And their publication: https://doi.org/10.1371/journal.pone.0213090
- OpenAire-Nexus' guide to licensing research data: https://www.openaire.eu/how-do-i-license-my-research-data
- Figshare license help page: https://help.figshare.com/article/what-is-the-most-appropriate-licence-for-my-data
- Software licensing:
  - https://towardsdatascience.com/a-data-scientists-guide-to-open-source-licensing-c70d5fe42079
  - https://choosealicense.com