Open Source, Sustainability, and CC0

Jacob Barhak

Sole Proprietor
Software Developer &
Computational Disease Modeler

August Penguin
2023 March 3
Tel-Aviv, Israel

Conflict of Interest

- Payment/services info: Dr. Barhak reports non-financial support and other from Rescale, and MIDAS Network, other from Amazon AWS, Microsoft Azure, MIDAS network, other from The COVID tracking project at the Atlantic, other from John Rice and Jered Hodges,

- Financial relationships: Jacob Barhak declare(s) employment from MacroFab, United Solutions, B. Well Connected health. The author had a contract with U.S. Bank / Apexon, MacroFab, United Solutions, and B. Well during the work. However, none of these companies had influence on the work reported here.

- Jacob Barhak declare(s) employment and Technical Support from Anaconda. The author contracted with Anaconda in the past and uses their free open source software tools. Also the Author received free support from Anaconda Holoviz team and Dask teams.

- Intellectual property info: Dr. Barhak has a patent US Patent 9,858,390 - Reference model for disease progression issued to Jacob Barhak, and a patent US patent Utility application #15466535 - Analysis and Verification of Models Derived from Clinical Trials Data Extracted from a Database.

- Other relationships: personal fees from United Solutions, personal fees from B. Well Connected health, personal fees and non-financial support from Anaconda.

- Also, from 2006-2012 Jacob Barhak was paid by the University of Michigan to develop software licensed under GPL license.

- **However, despite all support, Dr. Barhak is solely responsible for contents of this publication.**

I Benefited From Open Source

- I left an institution and continued develop open source after I left
- I developed:
  - The MIcro Simulation Tool (MIST) as a modeling engine **under GPL**
    https://simtk.org/projects/mist
  - The Reference Model for Disease Progression as a **proprietary** model using MIST
    https://simtk.org/projects/therefmodel
- Some achievements:
  - The Reference Model is the most validated diabetes model known worldwide
  - The Reference Model is the first multi-scale ensemble model for COVID-19
  - The work connected with SBML towards model sharing: http://dx.doi.org/10.1177/0037549718793214
  - Awarded United States Patent 9,858,390, January 2, 2018
  - Awarded United States Patent 10,923,234, February 16, 2021

So Why am I here?

- Investing in open source is not sustainable!
  - Funding prefers larger organizations
  - Some potential users prefer
    - Commercial solutions
    - Develop their own
  - Eventually control issues appear

What Control Issues?

- Opens source code is mostly governed by Copyright law

- The original owner of the code has some control over the code and can release or deny the code under a different license

- Modified work therefore depends on agreement between contributors

- What happens in case of a disagreement?

Issue #1: Over Interpretation

- In 2017 I asked the institution where the original MIST GPL code was developed to:
  - Release other code that started from some documentation examples used by MIST under a different license
    - There was no code overlap of the new code
- The request was denied
  - There was no good reason for denial since there was no code overlap or reuse of GPL code.
- I reached an agreement to use LGPL

Issue #2: Abandoned Code

- In 2019 I found out that GPL code was abandoned
  - According to a 3rd party, the institution chose to use a different engine
  - In fact, abandonment may have happened long before:
    - to the best of my knowledge, the institution did not release an update for 7 years
  - I was the only maintainer for a many years
- In 2020 when started work on COVID-19:
  - I asked for ownership transfer
    - I was denied - One argument was "that is a large ask"
  - I asked for a different license like CC0
    - I was denied
- Therefore, one contributor for the code can:
  - Deny incentive from any other by abandoning code and not transferring ownership
  - This remove incentive to improve the code or even maintain it

Some Say Patents are Evil
What is More Evil?

	Patents	GPL
Duration	Up to 20 years / ~ 1 generation	~70+ years / ~ lifetime
Restrictions	Per Country - Owner responsible to abide law	Copy Left / viral, Requires abiding to license - not necessarily local law
Maintenance	Fee at application, 3/7/11 years	None
Screening	Patent examiner scrutiny takes years	None
Ownership	Private, public upon expiration	Communal? Contributors?
Incentives	Owner can trade / sell license	Anyone can sell, yet sale looses value
Registration	Government / Archived	FSF / Distributed
Summary	Hard to get, private ownership and short term, provides incentives	Easy to get, communal ownership with long term restrictions, little incentive

(c) Jacob Barhak

Is This Why GPL is Used Less?

- GPL usage has dropped compared to other licenses:
  - https://opensource.com/article/17/2/decline-gpl

- Is the public becoming aware of the long term stagnation effects of GPL?
  - Software generation is shorter than human generation - More compatible with patent restriction time
  - Abandoned code cannot be re-purposed efficiently

- Getting communal support to change a license once released as GPL becomes difficult
  - All contributors to a version must agree?
  - The more contributors the harder it gets.
  - When in the last time you saw a parliament vote unanimously?

Creative Commons Zero - CC0 and Public Domain

- In 2010 Creative Commons released the CC0 license:
  - https://creativecommons.org/share-your-work/public-domain/cc0/
  - https://en.wikipedia.org/wiki/Creative_Commons_license

- The basic idea is "No Rights Reserved"
  - In lay terms it attempts to remove copyright and put the work in public domain
  - No copyleft or attribution needed

- The license is honest:
  - The license is aware of restrictions like patents and warns the creator before releasing the work

- At of 26-Feb-2023 there were 421 projects on PyPi with CC0 license:
  - https://pypi.org/search/?c=License+%3A%3A+CC0+1.0+Universal+%28CC0+1.0%29+Public+Domain+Dedication

Recent Issues with Open Source and CC0

- The Journal of Open Source (JOSS) decided against including CC0 in its license sets
  - https://github.com/openjournals/joss/issues/889

- NumFocus has decided against adopting CC licenses in a confusing manner:
  - Discussion links:
    - https://groups.google.com/g/numfocus/c/2PoF-n2OT2Q/m/sG1ZR6AYCAAJ
    - https://groups.google.com/a/numfocus.org/g/licensing/c/naKouplSf5Q
    - https://groups.google.com/a/numfocus.org/g/licensing/c/lM0Kk7bWlUU

- Open Source Initiative (OSI) did not approve CC0:
  - https://opensource.org/faq/#cc-zero

- JOSS related to NumFocus which related to OSI - they followed OSI decision without investigating the issue seriously.

- Should there be an entity that defines what is open that confines other entities?

Recent Developments in Biological Modeling

- BioModels - a repository of many biological models recommends the CC0 license:
  - https://www.ebi.ac.uk/licencing

- The a group of researchers who met at the viral pandemic working group published a paper discussing licenses has published a paper supporting CC0 as a solution.
  - Karr Jonathan, Malik-Sheriff Rahuman S., Osborne James, Gonzalez-Parra Gilberto, Forgoston Eric, Bowness Ruth, Liu Yaling, Thompson Robin, Garira Winston, Barhak Jacob, Rice John, Torres Marcella, Dobrovolny Hana M., Tang Tingting, Waites William, Glazier James A., Faeder James R., Kulesza Alexander. Model Integration in Computational Biology: The Role of Reproducibility, Credibility and Utility. Frontiers in Systems Biology, Vol 2. 2022. https://doi.org/10.3389/fsysb.2022.822606
  - Discussion 1: https://lists.simtk.org/pipermail/vp-integration-subgroup/2021-January/000022.html
  - Discussion 2: https://lists.simtk.org/pipermail/vp-integration-subgroup/2021-May/000210.html

- Important Reasoning:
  - CC0 allows combinations of models
  - Supports reproducibility, integration, traceability, transparency, and even commercialization if needed.

Thank You

Feel free to contact:

Jacob Barhak
jacob.barhak@gmail.com

https://sites.google.com/view/jacob-barhak/home

Issue #3: Restrictions

- Some software may need some regulation:
  - Examples include:
    - some biological modeling that can be potentially dangerous
    - security tools that can be used to break laws to compromise privacy

- If released as open source, the code can be potentially misused

- An open source license may need restrictions to prevent misuse

- However, open source licenses typically discuss only permissive parts
  - This may mislead naive young developers

(c) Jacob Barhak

Specifically for Disease Modeling

- I personally called for publishing modeling code openly in a large forum:
  - My calls were ignored and not documented
  - Different entities look at this topic differently

- A group of Scientists published a call for transparency:

  https://doi.org/10.1126/science.abb8637

- I called for public discussion on ways to make it possible:
  - https://forum.comses.net/t/issues-with-regard-to-call-for-transparency-of-covid-19-models/8433
  - One major argument was that government funded research should benefit the public

Should Government Fund Open Source?

- Here are some previous signs in that direction where government wants reuse of products:

- Fair Access to Science and Technology Research (FASTR) Act:
  https://www.cornyn.senate.gov/content/news/cornyn-wyden-introduce-bill-increase-access-taxpayer-funded-research
  - Focuses on publication
  - Recognizes patents and national security as exceptions

- Previous US administration policy:
  http://blogs.nature.com/news/2013/02/us-white-house-announces-open-access-policy.html
  - Focuses on mostly publication

- NIH Strategic Plan For Data Science https://datascience.nih.gov/sites/default/files/NIH_Strategic_Plan_for_Data_Science_Final_508.pdf
  - Mentions the word open over a dozen times with regards to data and software

- However, to date NIH still allows funded bodies to retain ownership:
  - https://grants.nih.gov/policy/intell-property.htm

Suggested Solution for Funding Bodies and Research

- Entities that fund open source should embrace the following strategy:
  - The definition of open source should include licenses like CC0
  - Ask software products of research to be released as open source license for the duration of funding
  - The work and license should include explicit restrictions
  - When research/development is abandoned, code should be released to public domain
    - For example Creative Commons Zero:
    - This will provide incentive for future development by private / public entities