Master's thesis

Master's Programme in Computer Science

Public Copyright Licenses in Software Engineering: A Multivocal Systematic Literature Review

Akira Taguchi

February 19, 2024

Faculty of Science

University of Helsinki

Contact information

P. O. Box 68 (Pietari Kalmin katu 5)
00014 University of Helsinki,Finland

Email address: info@cs.helsinki.fi
URL: http://www.cs.helsinki.fi/

HELSINGIN YLIOPISTO – HELSINGFORS UNIVERSITET – UNIVERSITY OF HELSINKI

Tiedekunta — Fakultet — Faculty	Koulutusohjelma — Utbildningsprogram — Study programme
Faculty of Science	Master's Programme in Computer Science

Tekijä — Författare — Author

Akira Taguchi

Työn nimi — Arbetets titel — Title

Public Copyright Licenses in Software Engineering: A Multivocal Systematic Literature Review

Ohjaajat — Handledare — Supervisors

Prof. Tomi Männistö

Työn laji — Arbetets art — Level	Aika — Datum — Month and year	Sivumäärä — Sidoantal — Number of pages
Master's thesis	February 19, 2024	8 pages, 3 appendix pages

Tiivistelmä — Referat — Abstract

Public copyright licenses play a major part in software engineering. For example in open source there must be an appropriate public copyright license attached to the source code in order for open-source software to be freely available for possible modification and redistribution. Understanding public copyright licenses can be difficult. This could stem from the legal nature of the license texts and the large number of already-existing public copyright licenses. sub-problem here. thesis' contribution to solution here.

Tell about the research method here.

Tell about the results here.

Tell about the discussion here.

ACM Computing Classification System (CCS)
Social and professional topics → Computing / technology policy → Intellectual property → Licensing

Avainsanat — Nyckelord — Keywords

open source, free / libre software, copyright, proprietary software, copyleft, license

Säilytyspaikka — Förvaringsställe — Where deposited

Helsinki University Library

Muita tietoja — övriga uppgifter — Additional information

Software study track

Acknowledgements

Contents

1 Introduction

Public copyright licenses play a major part in software engineering. For example in open source there must be an appropriate public copyright license attached to the source code in order for open-source software to be freely available for possible modification and redistribution. Because open source is central to software engineering the licenses enabling open source must also be considered important in the same context.

Public copyright license is defined by Wikipedia with the following words (Category:Public copyright licenses, 2012):

> "A public copyright license is a copyright license where the licensees are not limited. Examples include free content, open content, Creative Commons, free software and open source licences."

Understanding public copyright licenses can be difficult. This could stem from the legal nature of the license texts and the large number of already-existing public copyright licenses. The license texts usually favors correctedness over the readability for the developer. This is because the license text has to act as a valid legal instrument otherwise it cannot be endorsed (Ferguson, 2006). The lack of understanding of public copyright licenses leaves too much room for interpretation. In June 21, 2023 IBM's Red Hat seemingly violated a public copyright license, GNU General Public License version 2 or GPLv2 for short (Kuhn, 2023) (McGrath, 2023). This was an unpleasant surprise to the public since the project behind GPL, GNU initially attempted to ensure the users via the GPL have to the following three freedoms (GNU, 1996):

- Freedom 1: The freedom to study how the program works, and change it so it does your computing as you wish. Access to the source code is a precondition for this.
- Freedom2: The freedom to redistribute copies so you can help others
- Freedom 3: The freedom to distribute copies of your modified versions to others. By doing this you can give the whole community a chance to benefit from your changes. Access to the source code is a precondition for this.

Regardless, IBM's Red Hat essentially rendered the previously public Red Hat Enterprise

Linux, or RHEL for short, proprietary software. If the licenses would be more easily understood the proprietarization of RHEL would have been less of a surprise to the users.

On top of public copyright license details, software engineers in general have a tough time understanding the basic goals of public copyright licenses used in software engineering. In the instance of the RHEL incident it would not have been a big surprise to software engineers if they would have known about other licenses and what they try to achieve or how old is GPLv2 and why it has been succeeded by GPLv3.

This thesis' goal is to contribute into the solving these problems in a structured manner. First we state definitions and terminology used in the scope of this thesis. We go over the reasons why there does not exist consistent terminology in this area and why the conversely the definitions are the most stabile ones in this area. Second we take a deep dive into the multivocal literature through a systematic literature review. To make more information available, a mapping study connected to the terminology scope defined in the first step is needed. Third step will be to offer our own suggestions and basic knowledge for professionals and academics in the industry to enhance the understanding of public copyright licenses in software engineering. This is also the basis for future research and contributes to stablizing the terminology in the academic field.

1.1 Research goal, questions and contributions

The primary objective of this research is to conduct a systematic literature review of the current state of the public copyright licenses in software engineering, the evaluation of the them and the evidence level of the research. The research aims to provide a broad and scientifically interdisciplinary. perspective on relevant licenses and to extract key findings through a rigorous literature review process. The research questions of the review are:

- RQ1: How often do public copyright licenses in software engineering change?
- RQ2: How have the public copyright licenses in software engineering changed?
- RQ3: How long is the average public copyrigth license in software engineering?
- RQ4: What are the common reasons for version changes to public copyright licenses in software engineering?

1.2 Background and terminology of public copyright licenses

literature review here

1.3 Thesis structure

Write how this thesis is going to contribute to the aforementioned problems.

2 Methods

2.1 Research questions

2.2 Search stragey

2.2.1 Search method

2.2.2 Search scope and terms

2.3 Search process

2.4 Inclusion and exclusion criteria

2.5 Quality and evidence criteria

2.6 Data collection and data analysis

3 Results

3.1 Placeholder question (RQ1)

3.2 Placeholder question (RQ2)

3.3 Placeholder qiestion (RQ3)

4 Discussion

4.1 Implications for research

4.2 Implications for software engineering professionals

4.3 Limitations and threats to validity

4.3.1 Limitations of license selection for review

4.3.2 Limitations in data extraction

5 Conclusions

5.1 Future research

Bibliography

Category:Public copyright licenses (2012). *Category:Public copyright licenses — Wikipedia, The Free Encyclopedia*. https://web.archive.org/web/20240131085240/https://en.wikipedia.org/wiki/Category:Public_copyright_licenses. Accessed: 2024 Jan 31.

Ferguson, D. (2006). "Syntar Errors: Why Version 3 of the GNU General Public License Needs Debugging". In: *North Carolina Journal of Law & Technology* 7.2, pp. 397–420. URL: https://heinonline.org/HOL/Page?handle=hein.journals/ncjl7&id=403.

GNU (1996). *What is Free Software?* https://web.archive.org/web/19980126185518/https://www.gnu.org/philosophy/free-sw.html. Accessed: 2024 Feb 8.

Kuhn, B. M. (2023). *A Comprehensive Analysis of the GPL Issues With the Red Hat Enterprise Linux RHEL Business Model*. https://web.archive.org/web/20240205080551/https://sfconservancy.org/blog/2023/jun/23/rhel-gpl-analysis/. Accessed: 2024 Feb 5.

McGrath, M. (2023). *Furthering the evolution of CentOS Stream*. https://web.archive.org/save/https://www.redhat.com/en/blog/furthering-evolution-centos-stream. Accessed: 2024 Feb 5.

Appendix A Primary studies identified in the search process and their inclusion/exclusion criteria

Appendix B Primary studies reviewed in the quality criteria step and manual exclusions reasons

Appendix C Primary studies reviewed, read in full and data extracted