www.software.ac.uk

Open Software for (Open) Science

http://dx.doi.org/10.6084/m9.figshare.1434044

e-IRG Workshop, 3 June 2015, Riga

Neil Chue Hong (@npch), Software Sustainability Institute
ORCID: 0000-0002-8876-7606 | N.ChueHong@software.ac.uk

Supported by BBSRC EPSRC Pioneering research E·S·R·C Economic & Social Research Council

Project funding from Jisc NERC Science of the Environment

Overview

- Open source: philosophy and practice
- Open source licenses
- Other types of licenses
- Pros/Cons of licenses
- Case studies

DOI: *10.6084/m9.figshare.1434044*

I Am Not A Lawyer

For legal insight, I recommend: http://ifosslawbook.org/ and for everything else:http://oss-watch.ac.uk/

DOI: *10.6084/m9.figshare.1434044*

Open Source Software is Free...

www.software.ac.uk

Photo “free beer tap” by jakob fenger (CC-BY)

Photo “Speech”by Quinn Dombrowski (CC-BY)

DOI: *10.6084/m9.figshare.1434044*

Four essential freedoms

www.software.ac.uk

- Freedom for anyone to run the program as they wish, for any purpose ***(non-discriminatory)***
- Freedom to study how the program works, and change it so it does your computing as you wish ***(source code available, modifications allowed)***
- Freedom to redistribute copies so you can help others ***(redistribution allowed)***
- Freedom to distribute copies of your modified versions to others to give them a chance to benefit from your changes ***(derivatives allowed)***

DOI: *10.6084/m9.figshare.1434044*

Open source, open science

- Non-discriminatory
  - Research should not be restricted or siloed
- Access to source code
  - Research should be transparent, robust, and accessible
- Redistribution of software
  - Providing access to the widest possible community
- Removing barriers to reuse
  - Research should encourage building on the work of others, and giving them credit

DOI: *10.6084/m9.figshare.1434044*

Types of open source license

www.software.ac.uk

- Permissive ("Use with attribution")
  - Simple (e.g. Modified "3-clause" BSD, MIT)
  - Grant Patent Rights (e.g. Apache, Eclipse)
- Copyleft ("Share modifications under same license")
  - Strongly copyleft (e.g. GPL)
  - Weakly copyleft (e.g. LGPL, Mozilla, EUPL)
- All OS licenses allow private *and* commercial use; modification; distribution; limit liability; retain copyright
- More information
  - http://choosealicense.com/licenses/
  - https://tldrlegal.com/

DOI: *10.6084/m9.figshare.1434044*

Other common licenses

- Closed (“Proprietary”)
- Restricted (“Academic” / “Non-commercial”)
- Public domain / CC0
- Informal license
- No license
- Also
  - Larger works comprising code with different licenses (“license compatibility”)
  - Software made available under more than one license (“dual licensing”)

DOI: *10.6084/m9.figshare.1434044*

Licenses vs Governance

www.software.ac.uk

- Open source software is more than the license
  - Some open licensed software is not produced openly
  - Some closed licensed software benefits from open source project principles and processes
  - Most open source contributors are paid to do so
- Best practice identified from OSS projects useful for governance and project/product management of all types of software
  - http://producingoss.com/

DOI: *10.6084/m9.figshare.1434044*

Pros and Cons of Open Source

www.software.ac.uk

- Pros
  - Easier to evaluate
  - No lock-in
  - Easier to attract contributors and staff
  - Frictionless code sharing
  - Free advertising
  - Able to take with you
  - Generally better modularisation
  - Reduced duplication
  - Can work together on common platforms
- Cons
  - Harder to corner cut when shipping to deadline
  - Harder to gain income from selling the software
    - But most sell services
  - Cannot restrict to non-commercial use only
    - But what is non-commercial use?
  - Once licensed, cannot revoke license
    - Though can relicense
  - Can increase competition

DOI: *10.6084/m9.figshare.1434044*

Building out new functionality

www.software.ac.uk

- Dropbox moved to Go from Python to benefit from better concurrency support and execution speed
- Go did not have the depth of libraries that Dropbox needed to build larger systems
- Dropbox team started to build its own libraries
  - connection management and memcache client
- Open-sourced work to kickstart community and help create better production systems
- https://blogs.dropbox.com/tech/2014/07/open-sourcing-our-go-libraries/

DOI: *10.6084/m9.figshare.1434044*

Building community

- Spark started as academic project at Berkeley in 2009
  - Code released under BSD in 2010
- Started Apache Incubator process in June 2013
  - Gained contributions from 120 developers in 25 organisations, including Intel, Yahoo!, Cloudera, Alibaba
  - Relicensed under Apache 2.0 license
- Became full Apache project in February 2014
  - Most active Apache project in 2014
  - Source code on GitHub
  - Used globally by Alibaba, Amazon, IBM
- Company formed to commercialise in Sep 2013
  - Based around hosting and consulting

DOI: *10.6084/m9.figshare.1434044*

Creating widely used platforms

www.software.ac.uk

- Git originally written and released by Linus Torvalds under GPL license
  - Git trademarks enforced for consistency
  - Easy for third parties to create services on top
- GitHub contribute to Git and open source some of their own tools
  - Business model exploits paid hosting / support for enterprise installations
    - http://tom.preston-werner.com/2011/11/22/open-source-everything.html
    - http://todogroup.org/blog/why-we-run-an-open-source-program-github/
- Gitlab go further and release their commercial EE edition under MIT license, but code must be requested

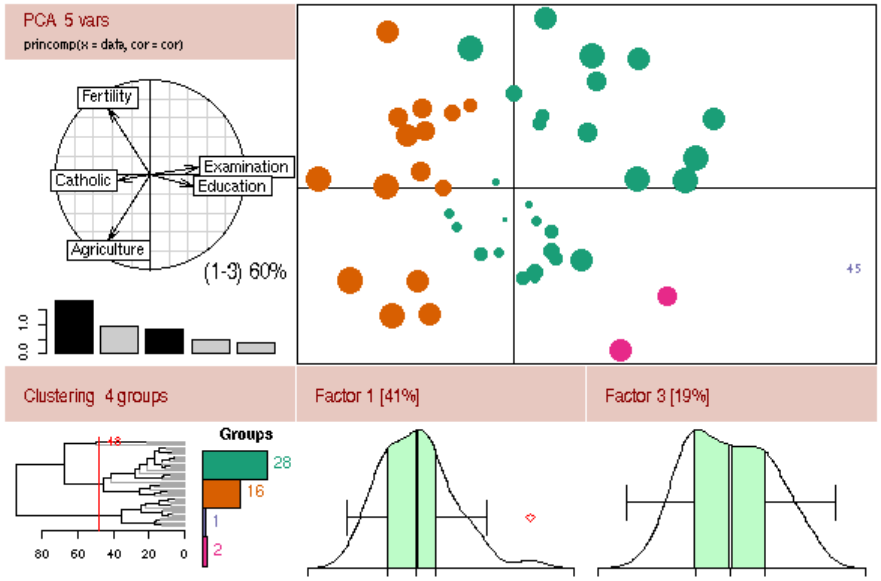
DOI: *10.6084/m9.figshare.1434044*

Underpinning research

www.software.ac.uk

- Basics: Website, mailing list, code repository, issue resolution
- Remove barriers to participation, increase efficiency

- 1993: First public release; 2 devs
- 1995: Code open sourced; 3 devs
- 1996: r-testers list set up
- 1997: lists split: r-announce, r-help, r-devel; public CVS; 11 devs
- 2000: CRAN split and mirror
- 2001: BioConductor
- 2003: Namespaces
- 2005: I8n, L8n
- 2007: R-Forge
- Today: BioConductor (33 core devs), R-Forge (532 projects, 1562 devs), CRAN (1400+ packages)



http://cran.r-project.org/doc/html/interface98-paper/paper_2.html

DOI: *10.6084/m9.figshare.1434044*

Case studies in academia

www.software.ac.uk

- OSS Watch have produced case studies of several open source software projects in academia
  - Koha: a case study in project ownership
  - Wookie: a case study in sustainability
  - ATutor LMS: a case study
  - TexGen: a case study
  - MediaWiki: a case study in sustainability
  - WebPA: the road to sustainability
  - Apache Cocoon: a case study in sustainability
  - Moodle: a case study in sustainability
  - Exim: a case study in sustainability
- http://oss-watch.ac.uk/resources/casestudies

DOI: *10.6084/m9.figshare.1434044*

www.software.ac.uk

“[T]here's been a stunning and irreversible trend in enterprise infrastructure. If you're operating a data center, you're almost certainly using an open source operating system, database, middleware and other plumbing. No dominant platform-level software infrastructure has emerged in the last ten years in closed-source, proprietary form.”

- Mike Olsen, Cloudera founder

GPL vs Apache

- Software freedom vs developer freedom
- Copyleft licenses have increased license management costs
- Industry is generally more supportive of permissive licenses
- Open source software no longer written as an alternative to closed, but as the platform to drive service provision
- Software no longer the competitive differentiator, but the ability to operate it at scale
- http://timreview.ca/article/650

DOI: *10.6084/m9.figshare.1434044*

Summary

- Open source now de facto for infrastructure software
- Open source encourages
  - Exploitation
  - Reproducibility and Robustness
  - Reuse
- Open source helps support open science
- *"Publicly funded research […] is a public good, produced in the public interest, which should be made openly available"* – RCUK
  - http://www.software.ac.uk/resources/guides/epsrc-research-data-policy-and-software

DOI: *10.6084/m9.figshare.1434044*