briefing paper

digital preservation europe

Open Source in Digital Preservation

Open source denotes the principles of promoting open access to a good's production or design process and the product itself. It is mostly used in the context of computer software, meaning that the knowledge assembled in software programs and operating systems is available. The Mozilla foundation's web applications like Firefox and Thunderbird are very prominent examples of open source software. Open source is often mentioned in the digital preservation context for open standards play an important role here. File format specifications and document formats can be also open source, and related to open standards. Together they satisfy quite a number of preservation requirements but for a number of reasons they cannot be proclaimed as a one-fits-all solution for digital preservation.

Free as in 'Free Speech'

Free software grants everyone the right to use, copy, and distribute software with or without modifications, either free of charge or for a certain fee. Not all Open Source Software is free of charge and not all software that is free of charge is Open Source. There are many potential ways to make a program free. Although the term 'free software' should refer to a number of freedoms, proprietary software vendors tend to use this term to refer to price. According to the definition of free software by the Free Software Foundation (FSF), a software product that is offered free of charge is not free software as it doesn't allow the free use, redistribution or modification of the software.

Open Source Software Licensing Models

It is often assumed, wrongly, that Open Source equates to free of charge. However the availability of source code can be subject to a wide range of licensing models. The term itself only means that the source code will be made available, making it easier for other developers to implement interfaces to or extensions of the software.

Due to the Berne Convention software is automatically copyright protected unless it is released to the Public Domain. Therefore the easiest way to declare a program to be free software is to release it to the Public Domain. Under the Public Domain all copyrights are abandoned and gives individuals the right to convert the software program into proprietary software which they could subsequently be marketed for profit.

The two licensing models, shareware and freeware, should not be mistaken for open source. A product that is declared shareware allows for redistribution but not for any modifications. After a certain period of time the user of the software is required to pay a licence fee which very often is claimed by dramatically limiting the functionality of the product. Freeware products are made available freely in binary form but modifications are prohibited by withholding the source code.

Creators of commercial or proprietary software often use copyright to restrict the users' freedom. Therefore the term 'copyleft' was invented to assure that no-one can add any additional restrictions once software has been modified or redistributed. This means that all freedoms associated with free software are guaranteed to everyone who owns a copy. Whichever model you chose, all open source software licences share one important right, the right to access and modify the source code. Differences occur in relation to rights available after modification of the software of linking it to products under licence.

GNU Public License (GPL)

One of the most well-known licensing models for open source software is the GNU General Public Licence (GPL). The GPL grants the right of redistribution, be it commercially or free of charge, as long as the new software is licensed by the GPL as well. Source code under the GPL is allowed to be modified but the derived work must again be published under the GPL. Because of this very strong copyleft the GPL is often considered too restrictive. A company that uses software or libraries published under the GPL also has to publish their software, i.e. the software it created with the help of the GPL'ed source code.

GNU Lesser General Public License (LGPL)

Due to its 'viral' effect the GPL makes the use of free software in commercial software quite unattractive and thus constrains the applicability of open source products. In contrast to the GPL, companies using source code that is published under the

LGPL in their commercial product don't have the obligation to publish the whole product under the LGPL again. Thus the LGPL is especially suitable for software libraries such as the GNU C libraries.

Berkeley Software Distribution (BSD)

A software product under the BSD licence allows commercial use as well as being included into other commercial products. The derived work of a BSD product does not have to be free itself. It can be redistributed under a commercial software licence.

Mozilla Public License (MPL)

The source code of the well-known internet browser Mozilla is released under the MPL. It grants the right to incorporate an MPL protected product into another software product which then can apply any licence. Modifications however have to be licensed under the MPL again.

Open Source and Digital Preservation

Open source is not necessarily confined to software. Open standards, for example, can also be regarded as open source, in the sense that they are freely available and open to the public. Assets conforming to open standards are more qualified for being preserved over a long period of time inasmuch as they give access to the file format, making it easier to develop a tool which migrates this format should it become obsolete.

In addition to this, many of the file format specifications like the OpenOffice.org spreadsheet and document formats are themselves open source. However, proprietary solutions can also provide satisfying results, having the advantages of, continuing and guaranteed customer support. Further, industry and user acceptance for standards and tools is often more important than the actual openness itself. The best standard does not necessarily have to be the most feasible when considering support, availability, and acceptance by industry partners.

Open standards for files to be preserved, and also the implementation of the preservation software and its parts under an open source licence brings advantages. Other institutions can use components developed using open software and adapt them to their needs. Furthermore, especially with respect to trust, open source software is much easier to evaluate then proprietary software.

Conclusions

Open source should not be mistaken for the solution to all problems. Like commercial software, open source applications differ in quality of implementations. Unlike commercial software, however, open source applications come without warranties. Therefore, it is essential to gather all requirements when considering the use of open source software.

Due to source code availability the main advantage of open source software lies in the possibility of customising and tailoring it to an institution's needs. Also shortcomings in the software can be corrected independently by either taking corrective actions oneself or by mandating some other company. With open source software one does not have to wait for some update of the vendor which might never come.

Further information and resources

[1] Berne Convention.
The Berne Convention for the Protection of Literary and Artistic Works. This convention gives an internationally agreed definition on copyright. Unless the author explicitly declares differently, everything written down automatically falls under copyright law. This protection also applies to computer programs in whatever programming language they are expressed.
http://www.wipo.int/treaties/en/ip/berne/trtdocs_wo001.html

[2] Open Source The Unauthorised White Papers, Rosenberg, 2000.
This book addresses various aspects of open source software. Besides a clear definition of what commercial and non-commercial software is and all the licensing concerns it also deals with business issues such as reliability and performance. Furthermore it also analyses the costs of an Open Source solution.

[3] Free Software Foundation.
The Free Software Foundation was established in 1985. It maintains the Free Software Definition, promotes the usage of free software and promotes the rights to use, study, copy, modify and redistribute computer programs.
http://www.fsf.org

[4] The four kinds of freedom.
It precisely defines and explains the four kinds of freedom to which free software refers: the freedom to run, study and adapt, redistribute copies and improve the program. A piece of software is only free if these rights are granted to the users of it. Among other things this means that the users neither have to ask nor to pay for the permission to do these things.
http://www.gnu.org/philosophy/free-sw.html

[5] Open Source Initiative.
The Open Source Initiative was founded in 1998 and is devoted to promote open-source software and dispel misunderstandings. It gives a precise definition on what Open Source means.
http://opensource.org

Robert Neumayer, Vienna University of Technology
neumayer@ifs.tuwien.ac.at