Open Source communities for long-term maintenance of digital assets: what is offered for ODF & OOXML?

Jonas Gamalielsson and Björn Lundell

University of Skövde, Skövde, Sweden,
{jonas.gamalielsson, bjorn.lundell}@his.se

1   Introduction

Many organisations have requirements for long-term maintenance of their software systems and digital assets. In this paper, we explore sustainability of Open Source Software (OSS) licensed office applications (hereafter referred to as OSSOA) implementing the ISO-standardised XML-based document formats ODF (ISO/IEC 26300:2006) and OfficeOpen XML (ISO/IEC 29500:2008). We draw from prior research conducted in the Swedish public sector context where different applications for the two formats are used (Lundell 2011, Lundell & Gamalielsson 2011).

In a number of countries there are governmental policies mandating use of ODF in the public sector (Lundell 2011). Previous research in the Swedish public sector found that "there is little or no evidence of consideration given to document formats when municipalities procure software" (Lundell 2011). The same study found significant confusion concerning the concept of standard and differences between file formats and software applications.

Implementations of file formats that are distributed under an Open Source license clearly contribute to the desired economic effect of standards (FLOSSPOLS 2005), in that such implementations stimulate competition in the marketplace and minimise the risk for different types of lock-in effects (Lundell 2011). For example, organisations have had concerns for avoiding vendor lock-in for decades (Guijarro 2007).

Different Open Source licenses have different effects on longevity of tools, and licence selection is a critical issue for companies and communities (Engelfriet 2010). Previous results on Open Source implementations of document formats show that community commitment and choice of licenses may significantly affect long-term maintenance of digital artefacts (Lundell & Gamalielsson 2011). Further, company commitment and choice of software licenses affect longevity of tool support for different file formats (Lundell et al. 2011).

Before an organisation adopts an Open Source project it is important to evaluate its communities in order to make sure that it is healthy and that the project is likely to be sustainable and maintained for a long time (van der Linden et al. 2009). From this, our goal is to investigate the availability of effective and sustainable OSSOAs for creation and editing of documents.

2   Research approach

Firstly, we identify effective and sustainable OSSOAs implementing the two XML-based formats ODF and Office Open XML (OOXML). Hence, we undertook: a

systematic investigation of existing OSSOAs through a literature analysis, interviews, search in OSS forges, and an analysis of information on existing OSS available through framework agreements provided by the Legal, Financial and Administrative Services Agency (*swe.* Kammarkollegiet). Our analysis identified several OSSOAs for creation and editing of documents in the ODF format, whereas we only identified one OSSOA that provides support for creation and editing of documents in the OOXML format (docx4all). For the OOXML format, our selection of docx4all for investigation was informed by a previous analysis covering OSSOAs that provide support for creation and editing of documents (Garshol, 2010). We note that our study covers all tools using ".docx" as its internal format (although the ECMA, instead of the ISO version). It was decided to explore the LGPL-licensed LibreOffice (LO) project as a representative of an effective OSSOA implementing the ODF format. Our review revealed that both LO and OpenOffice.org (OO) are in professional use with the ODF format in the Swedish public sector, whereas there is no OSSOA in professional use for the OOXML format. For the study we selected LO as an analysis would encompass an investigation of consequences of a fork. Further, amongst organisations already using OO we observed an interest in exploring LO as a potential alternative. We establish the sustainability for the selected Open Source projects through an assessment of contributions to the Software Configuration Management system[1] (SCM) over time.

Secondly, for each document format for which there exists an Open Source implementation with a sustainable community, our goal was to extend the analysis with a deeper characterisation of the Open Source community. For the ODF format, we analysed the community of the LO project in order to reveal its potential for long-term maintenance. This project constitutes a fork from the OO project. Our investigation addresses both the base project (OO) until the fork and the branch project (LO) after the fork. Specifically, we investigated the LO developer community over time with the view to obtain insights concerning long-term sustainability by means of a quantitative analysis. In so doing, we extend a previous analysis of the OO developer community (Meeks 2008) by elaborating on the effects of the fork.

To investigate the sustainability of Open Source communities, we adopt the approach from an earlier study (Gamalielsson et al. 2011) in order to analyse the contributions in terms of committed SCM artefacts of the Open Source projects over time. The data for the LO project was collected from the LO website (www.libreoffice.org/get-involved/developers, accessed 8 Aug. 2011), and for docx4all the data was collected from the docx4all website (dev.plutext.org/trac/docx4all, accessed 8 Aug. 2011).

3 Results

The developer activity in the selected OSSOAs is presented in Figure 1, which shows the number of commits for each month over the four year time period from August 2007 to July 2011 for LO (blue trace) and docx4all (red trace). Our SCM analysis of the LO project includes the development in OO before the fork in September 2010. We note that activity in the LO project varies, with a distinct peak in connection with the OO 2.4 release in March 2008 (59804 commits in April, which is not visible in the diagram for scaling reasons). Since October 2008 (with the release of OO 3.0) there

---

[1] Examples include Git, SVN and CVS.

have been 2851 commits each month on average. We also note that the activity is much lower in docx4all (red trace in Figure 1), and that there has been very limited SCM activity in the project since September 2009. Further, there have been no contributions to the SCM since September 2010 (when only two commits were provided) except for a single commit in April 2011. The maximum number of commits in docx4all during one month is 195 (during May 2008). Since the start of the project (October 2007) there have been 29 commits each month on average. We acknowledge that docx4all is a much smaller project than LO, and therefore it is not unexpected that docx4all has had fewer commits. Totally, there have been 664 committers (unique IDs in the SCM log) in the LO project over the studied four year period, whereas only two committers have contributed to docx4all since the creation of the project (where one of these committers has contributed 79% of all commits).
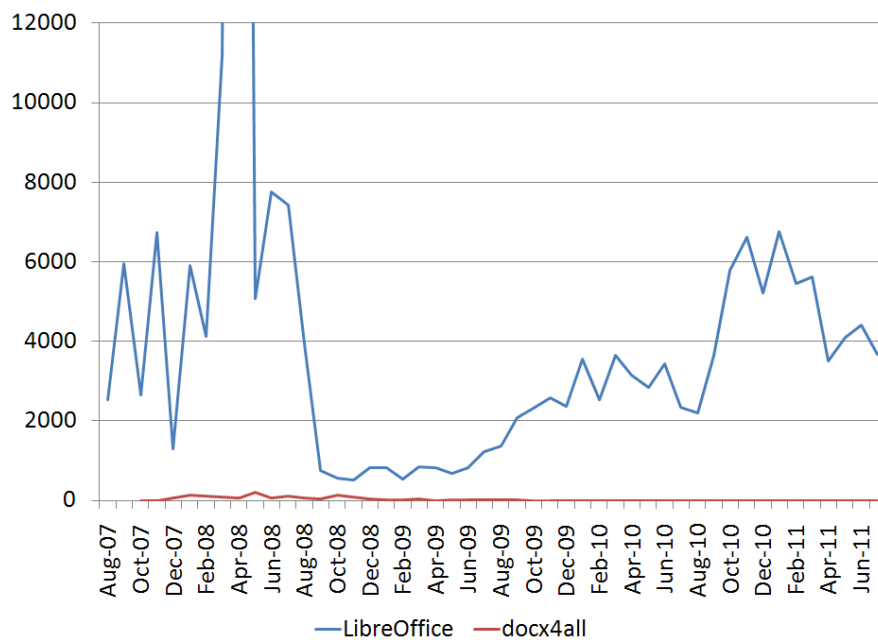

Fig 1. Number of commits over time for LO (blue) and docx4all (red)

Due to the absence of a sustainable developer community for docx4all, we decided to only consider LO in our further analysis. The proportion of commits for the 10 most active affiliations over time in the LO project is shown in Figure 2 (like in Figure 1, the peak in April 2008 is not visible for scaling reasons). It can be observed that "openoffice" is dominating until August 2010, and that other affiliations break the dominance from September 2010 (the month of the fork) and onwards. It is also noted that "sun" is most active in the period from October 2009 to July 2010, and that "oracle" is most active from August 2010 to March 2011. Further observations are that "novell" and "suse" have been active for the entire four year period with an increased activity after the fork, and that "redhat" has become the major contributor ever since the fork.
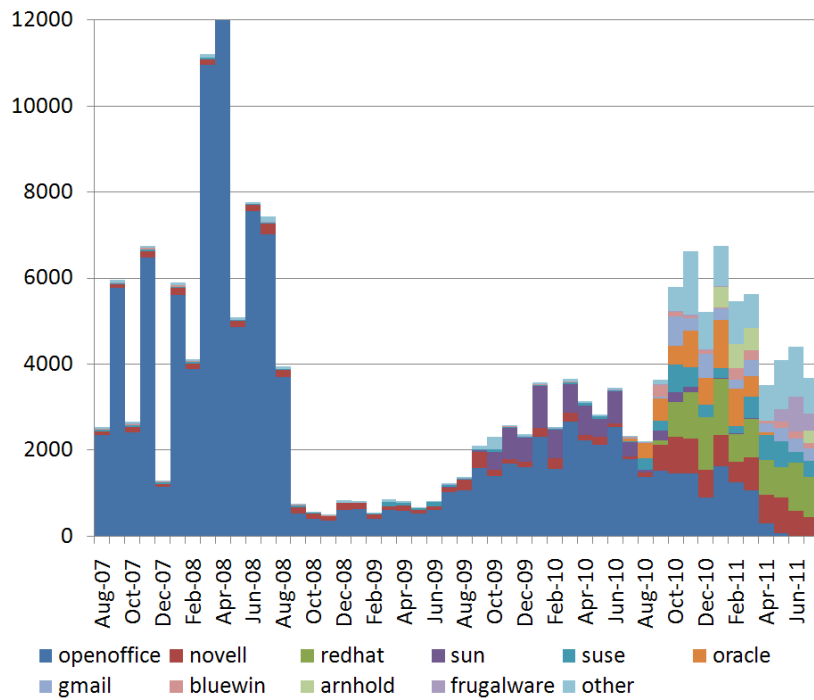
Fig 2. LO: Proportion of commits per affiliation over time

Figure 3 illustrates the total affiliation commit influence for LO 10 months before and after the fork on 28 September 2010, and further emphasizes the shift from "openoffice" domination to a more diversified developer community after the fork.
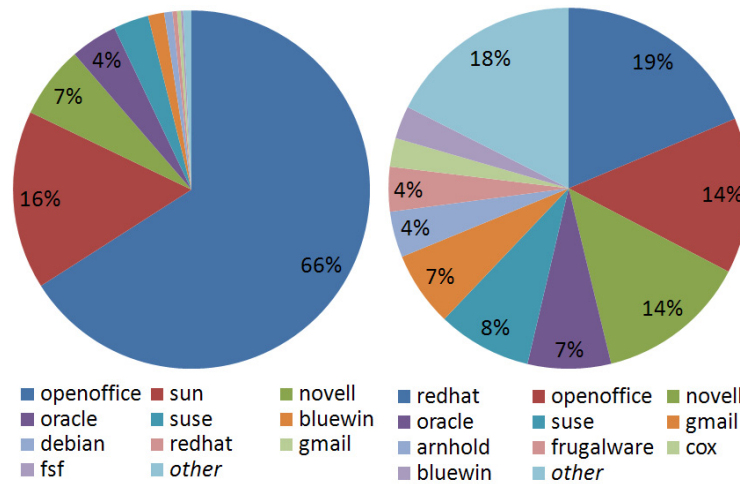

Fig 3. LO: total affiliation commit influence (left pie: during 10 months before the fork, right pie: during 10 months after the fork)

4   Conclusion and discussion

We find that there is a sustainable and effective OSSOA for the ODF document format but not for OOXML. Further, despite the relatively short time window after the fork, our analysis indicates that the LO project has an active and diversified developer community. This is in contrast with worrying observations made three years earlier for the OO project in which Meeks (2008) claims that the OO project "is a profoundly sick project, and worse one that doesn't appear to be improving with age." However, we acknowledge the inherent complexity in quantitatively analysing the OO and LO projects for a number of reasons, such as governance models and differences in practices regarding use of affiliation during project interactions.

Long term maintenance of digital assets and supporting OSSOAs is a significant issue for both private and public sectors. It is closely related to longevity of file formats. In a number of usage scenarios it is common that organisations need to preserve their systems and associated digital assets for more than 30 years (Lundell et al. 2011, Lundell 2011). It is important to recognise that any action in the private and public sectors must be based on a long-term vision and the recognition that no provider will remain on the market for the life-span during which digital assets must be maintained. For achieving a long-term sustainable ICT ecosystem it is important to recognize that for each file format used in the public sector there must exist a supporting Open Source implementation. Further, before a file format is used and referred in a public sector procurement, we argue that there should exist a sustainable implementation that is licensed under a "share-alike" or "keep-open" Open Source license. Such a requirement would be one effective strategy for minimising the risk for vendor lock-in and lack of interoperability in the public sector. It is an open question to what extent the Document Foundation may establish long-term sustainability for LO with its use of a "keep-open" license.

In acknowledging that our research has not addressed the extent to which the specification of each ISO standard is actually implemented in an OSSOA, we note that for the OOXML standard it has been claimed that "it is unclear whether anyone is able to implement the ISO standard" (documentfoundation.org 2011). Further, at time of writing, we note that there is uncertainty concerning when support for this standard will be provided by an office application[2] under any software license.

For further work, we plan to broaden our investigation of long-term sustainable communities to include both branches of the OO project, and also extend our analysis of sustainability and commitment of individuals to the projects over time. In addition, we plan to undertake a qualitative analysis of the projects, including investigation of: external events, choice of license, foundations, governance and work practices. A related issue for further investigation concerns the extent to which it is possible to successfully migrate documents between implementations of the two document formats. It is not uncommon that office applications provide warnings when a user attempts to save a document in a non-native file format. For example, in addition to its native ODF support, the most recent release (version 3.4.2) of the office application LibreOffice also provides options for saving a document in several non-native file formats. When attempting to save a document using the "Office Open XML Text" option in the office application LibreOffice, the application issues the warning: "This document may contain formatting or content that cannot be saved in Office Open XML Text file format. Do you want to save the document in this format anyway? Use the

---

[2] In fact, it has been estimated that full support for creation and editing of documents in OOXML will be provided in a proprietary office application 'no later than Office "15."' (Mahugh 2010).

latest ODF file format and be sure all formatting and content is saved correctly."
(documentfoundation.org 2011). Such warnings may cause users to vary and the extent
to which successful migration between formats actually is (and can be) provided
during long life-cycles beyond any single office application needs further exploration.

Acknowledgement

References

documentfoundation.org (2011). LibreOffice OOXML, The Document Foundation Wiki,
     http://wiki.documentfoundation.org/LibreOffice_OOXML, accessed 15 September 2011.
Engelfriet, A. (2010). Choosing an Open Source License, IEEE Software, 27(1): 48-49
FLOSSPOLS (2005). An Economic Basis for Open Standards, Deliverable D4, December 12,
     2005, http://flosspols.org.
Gamalielsson, J., Lundell, B. and Mattsson, A. (2011). Open Source Software for Model
     Driven Development: A Case Study, In Hissam, S. (Eds.) Open Source Systems: Grounding
     Research, IFIP Advances in Information and Communication Technology, Vol. 365, ISBN:
     978-3-642-24417-9, Springer, Boston, pp. 348-367.
Guijarro, L. (2007). Interoperability frameworks and enterprise architectures in e-government
     initiatives in Europe and the United States. Government Information Quarterly, 24 (1), 89-
     101.
van der Linden, F., Lundell, B. and Marttiin, P. (2009). Commodification of Industrial
     Software: A Case for Open Source. IEEE Software, 26 (4): 77-83.
Lundell, B. (2011). e-Governance in public sector ICT-procurement: what is shaping practice in
     Sweden?, European Journal of ePractice, 12(6), http://www.epractice.eu/en/document/
     5290101
Lundell, B. & Gamalielsson, J. (2011). Towards a Sustainable Swedish e-Government Practice:
     Observations from unlocking digital assets. In Proceedings of the IFIP e-government
     conference 2011(EGOV 2011), Delft, The Netherlands, 28 August - 2 September 2011.
Lundell, B., Gamalielsson, J. & Mattsson, A. (2011). Exploring Tool Support for Long-term
     Maintenance of Digital Assets: a Case Study, In Fomin, V. and Jakobs, K. (Eds.)
     Proceedings: 16th EURAS Annual Standardization Conference, European Academy of
     Standardisation, The EURAS Board, pp. 207-217.
Mahugh,   D.   (2010).   Office's   Support   for   ISO/IEC   29500   Strict,
     http://blogs.msdn.com/b/dmahugh/archive/2010/04/06/office-s-support-for-iso-iec-29500-
     strict.aspx, accessed 15 September 2011.
Meeks, M. (2008). Measuring the true success of OpenOffice.org, http://people.gnome.org/
     ~michael/blog/ooo-commit-stats-2008.html, accessed 15 September 2011.