

Open Source Software Migration in Integrated Information Systems in Public Sector

Bruno Rossi, Barbara Russo, and Giancarlo Succi
Free University of Bolzano-Bozen, Faculty of Computer Science,
Domenikanerplatz 3, 39100 Bolzano-Bozen, Italy
{bruno.rossi,barbara.russo,giancarlo.succi}@unibz.it
WWW home page: <http://www.case.unibz.it>

Abstract. This paper reports about a study on the introduction of Open Source Software (OSS) in a Public Administration located in Europe. The Public Administration examined has introduced OSS as a means to save on the license costs and to have a larger space for customisation purposes. The adoption of new software may have an impact on the employees' productivity that need to be addressed. In this article, we compare the usage of OpenOffice.org and Microsoft Office. Data about the usual office activities performed by the users participating to the experimentation have been collected by means of an automated non-invasive data collection tool. The result of this study reports a similar usage pattern of both suites in terms of workload, but a different approach in using functionalities provided by each software. A further analysis on the life cycles of documents elaborated with the office suites seems to validate the similarities among the software solutions examined.

1 Introduction

The introduction of Open Source Software (OSS) in substitution or in parallel with Closed Source Software (CSS) is an argument that acquired recently great relevance. The proposed savings in terms of license costs and the broader opportunities for software customisation are arguments that interest particularly private and public companies. There are many studies and market researches on the convenience of the migration that privilege one of the two solutions, depending mainly of the factor of cost considered [1]. A complete migration is not an easy step, especially in working environments where the interdependencies and the vertical integration is a key issue

Please use the following format when citing this chapter:

Rossi, B., Russo, B., Succi, G., 2006, in International Federation for Information Processing, Volume 205, Research and Practical Issues of Enterprise Information Systems, eds. Tjoa, A.M., Xu, L., Chaudhry, S., (Boston:Springer), pp.683-689.

[2]. Supporters of OSS also stress the importance to avoid the realisation of phenomena of lock-ins, situations in which a company commits itself to a single supplier or single data format [2, 3]. There are cases of successful migrations, like the city of Calgary in Canada [4] or the region of Extremadura in Spain [5]. But there are also cases of unsuccessful deployment, like the city of Nürnberg [6] or delays and over expenditures like the case of the city of Munich [7], both cases in Germany. The case study discussed in this paper concerns a European Public Administration. For two months a successful migration to OSS on the desktop side has been monitored. The analysis reported has focused on the software for office automation. The contribution to the field of this study is an evaluation of the similarities and differences in usage patterns of OpenOffice.org and Microsoft Office. This can shed some light on the effect of a transition on the routine of the office work in a PA.

2 The Study

2.1 Study description

The study has been based on the data collected from a Public Administration in Europe during a migration to the OpenOffice.org suite. This office automation suite offers similar functionalities as the ones offered by Microsoft Office. It is composed by a word-processor, spreadsheet software, presentation software and a drawing tool. The Public Administration (PA) examined has been adopting OSS for some time; the analysis we report refers to a situation, where the proprietary and open solution coexisted in the working environment. To monitor the time spent on different solutions, data have been collected with the aid of the PROM software [8]. With a non-invasive impact, the software gives the opportunity to register for every document the time spent, the name of the document and the functions used. This last feature is at the moment still limited, but can give useful insights of the different patterns of usage between the two solutions. The monitoring of users as we report in this paper has been performed during 2 months with both solutions installed in parallel. 100 users have been involved in the experimentation.

2.2 Dataset distribution

To have an idea of the evolving situation during the first two months of the experimentation, we can see in Figure 1 the comparison between Microsoft Office and OpenOffice.org usage. The figure refers to the average number of documents worked by all users on a specific day. As these numbers seem to report the daily averages are very similar ($\sigma_{OOo}=1,6$; $avg_{OOo}=4,42$; $min_{OOo}=1$, $max_{OOo}=6,82$; $\sigma_{MSO}=1,6$; $avg_{MSO}=4,2$ $min_{MSO}=1,5$, $max_{MSO}=8,5$).

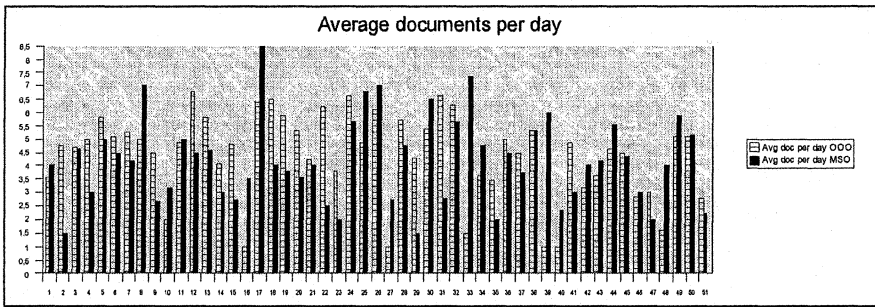


Fig. 1. Average number of OpenOffice.org documents per day (outlined) and MS Office documents per day (in black).

In Figure 2 the total daily number of documents per solution is reported. This is the global sum of all the documents handled daily by all users participating to the experimentation.

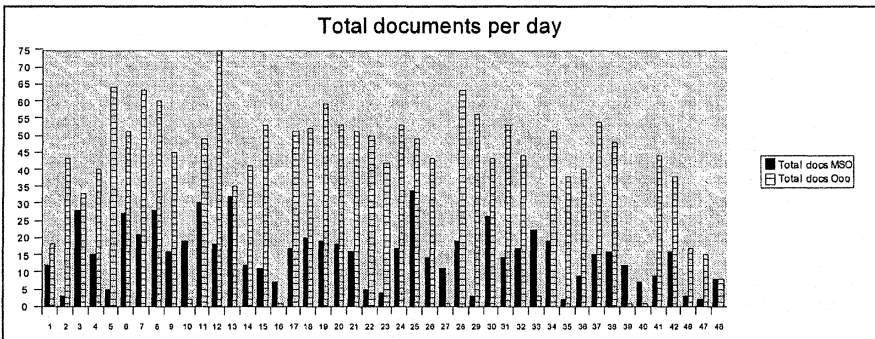


Fig. 2. Total OpenOffice.org documents per day (outlined) and MS Office documents per day (in black).

From this table can be derived that more users are in fact adopting the open solution. At the end of the period considered, the documents handled with OpenOffice.org have been 4.032 against 1.206 opened with Microsoft Office. This to justify that the migration examined is already in a mature state, in which the new technology introduced is taking over the old one.

The purpose of these figures is to give an idea of the existing situation in the PA that is the focus of our study.

2.3 Functions used

In this section we report the functionalities used on the different suites, divided per software. At the moment, the version of the software used for the data collection does not allow a more fine-grained analysis. In Table 1 the total numbers of functions adopted during the study. These were selected being the most representative one.

Table 1. Total functions used according to application type

	Microsoft Office	OpenOffice.org
Open	145	1038
Save As	205	1321
Print	170	1109
Spelling	178	578
Insert table	2	43
Find and replace	7	39

In Table 2, the same functions are reported, this time normalised with the number of documents handled per solution.

Table 2. Total functions weighted per document handled

	Microsoft Office	OpenOffice.org
Open	0.15	0.24
Save As	0.21	0.31
Print	0.18	0.26
Spelling	0.18	0.13
Insert table	0.00	0.01
Find and replace	0.01	0.01

The impact of activities like inserting tables and finding and replacing words seem very low in both solutions. We could not trace keyboard shortcuts, so the analysis in this sense is limited. In general the usage pattern of OpenOffice.org seems different, with more actions performed, like opening documents, saving and printing. Spelling instead had a higher impact in Microsoft Office than in OpenOffice.org.

2.4 Documents life cycle

To deepen the analysis of the differences between the two solutions, we derived a concept similar to the one of Product Life Cycle (PLC), in our case applied to documents. The concept of Product Life Cycle was first introduced by Theodore Levitt in 1965 [9]. Typical stages in a product life cycle are: Introduction, Growth, Maturity and Decline. There are many different variations of the Product Life Cycle

model that differ mainly in the names used to describe the different stages and the purpose of the underlying analysis. A similar model has been used to study the diffusion of new technology [10]. In Figure 3 a typical product life cycle is depicted, together with the different phases of maturity.

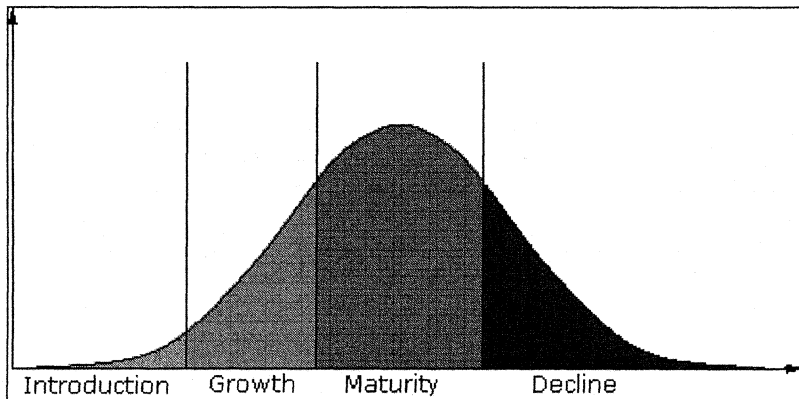


Fig. 3. A typical product life cycle, with four phases of maturity (Introduction, Growth, Maturity, Decline).

- The Introduction phase is where the product is introduced on the market; in this phase marketing has the strongest importance than in other phases.
- The Growth phase is where the increase in sales supported by marketing becomes considerable.
- In the Maturity phase the sales are stabilised and the speed of increase of sales is slowly reduced until it begins to become negative.
- In the Decline phase, the product is no more attractive to possible customers that may prefer a more technological concurrent product.

In the Product Life Cycle, the duration of each phase may be different between different products, as the adoption curve may have a different aspect.

We decided to model the DLC as a measure of software usage to further discover existing analogies between CSS and OSS. Following an analogy with PLC, we considered the life of a single document, as composed from different phases: its creation phase that starts with the generation of the underlying file, the growth phase where the document usage increases, the maturity where the usage reaches maximum levels and the decline phase, when the document's usage begins to decrement to the complete halt. To perform this task we analysed all 5.238 documents and divided them in two groups, the ones handled with OpenOffice.org and the ones handled with Microsoft Office. Subsequently, we reported the life of each document into the PLC model framework. The last step was to analyse the distribution of the derived

DLCs. In Table 3 we report the results obtained. The average length of the documents is very similar in a comparison between both solutions.

Table 3. Comparison of DLC of all documents, the scale is in days. Averages marked with a * have been obtained by excluding documents with a life cycle of one day

Application	Avg DLC length	St. Dev.	Avg DLC length*	Max DLC length
Microsoft Office	1.83	3.45	4.64	51
OpenOffice.org	1.78	3.74	4.06	46
All documents	1.82	3.47	4.11	51

As we should expect, the results of the DLC analysis are comparable, the difference between the two software solutions are minimal. As a side note, we discovered that only 25% of all documents have a life cycle greater than one day.

3 Conclusions

While the study is still limited, we are getting a clearer picture on the interactions of users with their desktop software. The study reported refers to a Public Administration where OSS has already been adopted in parallel with the closed solution for some time. In this sense we are in a more mature moment during the technology adoption, not the early phases of a complete migration. The next step would be the complete adoption of OSS for office automation, if the feasibility study performed show favourable. The results of our analysis show that proprietary and open solution can coexist on the working environment on the desktop side. Also the average number of documents per day seems comparable. Focusing on the functions used, some activities seem to be more adopted by users with the open solution than the closed one. During our study, some function we thought at first important were rarely used. A more fine-grained analysis will be possible with more accurate software for data collection, collecting more measures necessary to evaluate fully all the functions used. The analysis and evolution of the documents' life cycle, a concept borrowed from economics, seem also to justify the strong similarities between the two solutions offered.

References

1. The Yankee Group, North American Linux TCO Survey (2005); www.yankeegroup.com
2. C. Shapiro and H. R. Varian, *Information Rules: A Strategic Guide to the Network Economy* (Harvard Business School Press, 1999).
3. S. Liebowitz and S. Margolis, *Winners, losers & Microsoft, Competition and Antitrust in High Technology* (Oakland, CA, The Independent Institute, 1999).
4. Real World Linux 2004, Conference and Expo Reports (May 2004); http://www.realworldlinuxbiz.com/artman/publish/printer_rw104rp.shtml

5. Extremadura Linux Migration Case Study (2005); <http://insight.zdnet.co.uk/software/linuxunix/0,39020472,39197928,00.htm>
6. Stadt Nürnberg, Strategische Ausrichtung im Hinblick auf Systemunabhängigkeit und Open Source Software (2004); <http://online-service.nuernberg.de/eris/agendaItem.do?id=49681>
7. Landeshauptstadt München, Clientstudie der Landeshauptstadt München (2003); http://www.muenchen.de/aktuell/clientstudie_kurz.pdf
8. A. Sillitti, A. Janes, G. Succi, and T. Vernazza, Collecting, Integrating and Analyzing Software Metrics and Personal Software Process Data, Euromicro 2003, Belek-Antalya, Turkey, 1–6 September 2003.
9. T. Levitt, Exploit the Product Life Cycle, *Harvard Business Review* **43**, 81-94 (November-December 1965)
10. M.B. Frank, A New Product Growth Model for Consumer Durables, *Management Science*, **15**, 215-227 (1969)