

Criteria for the non invasive transition to OpenOffice

Bruno Rossi
Free University of Bozen
Bozen, Italy
Bruno.Rossi@unibz.it

Alberto Sillitti
Free University of Bozen
Bozen, Italy
Alberto.Sillitti@unibz.it

Marco Scotto
DIST – University of Genova
Genova, Italy
scotto@dist.unige.it

Giancarlo Succi
Free University of Bozen
Bozen, Italy
Giancarlo.Succi@dist.unige.it

Abstract – Open Source Software (OSS) is receiving an increasing attention as a possible alternative to proprietary solutions. There are supporters of both the alternatives that stress advantages and disadvantages, but what is missing is an empirical view of a transition with the aid of case studies and controlled experiments. The aim of the paper is to report the results of an empirical investigation in the field of office automation in the Public Administration (PA). The available OSS in the field is introduced in the existing environment while preserving the proprietary solution. The analysis is supported by both qualitative and quantitative data. The effects on productivity and on users' attitude towards OSS and the emerging criteria for a possible transition are exposed.

I. INTRODUCTION

Some people suggest that the usage of OSS in PAs guarantees higher levels of data security and privacy and ensures a major openness of the Public Administration that does not constrain citizens to buy any kind of license to read or edit the public available files. On the other side, the introduction of a new software to handle the documents may affect negatively the productivity. The reason is that the software may not be fitted to the requirements, requiring too much time for personnel training or it may be not easy to integrate in the existing system environment.

An experiment in a Public Administration (PA) has been deployed to understand deeply the effects of this introduction and develop a possible plan of action of a possible large-scale diffusion. The experiment has been deployed inside the COSPA project (Consortium for the Open Source in the Public Administration), a consortium that aims at analysing the effects of the introduction of Open Data Standards (ODS) and Open Source (OS) software for personal productivity and document management in European PAs [1].

The office automation field was taken in consideration for several reasons: the importance of this kind of software in the activities of a PA and the presence of two strong rivals in the field, the proprietary solution represented by Microsoft Office¹ and the OS solution offering similar functionalities, OpenOffice.org². Furthermore, there is an

¹In the following we will call the package "Microsoft Office" or simply, "Office".

²In the following pages the application will be called with the name that is usually used: "OpenOffice" omitting the prefix ".org".

already available study of the different levels of usability offered by the two solutions performed by the Berkeley University [2].

II. METHODOLOGY

The Goal Question Metrics (GQM) model was employed in every phase of the project, from the overall design to the creation of the questionnaires [3].

The experimental design followed three different paths:

a) *Classical experimentation*. On one side there has been the direct experimentation on 22 users of the PA under study, defined following the guidelines of the classical experimentation with strict numerical analysis [4]. It has been based on an experimental pre-test-post-test control group design of the type represented in Fig.1:

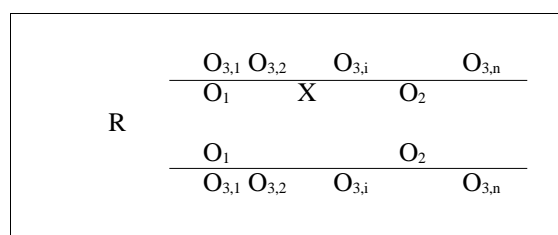


Fig. 1. Experimental design

This model is usually defined as "multiple time series design with randomisation" [5].

At the beginning a sample of 22 people participating to the experiment has been identified on the basis of the representativity inside the Public Administration and on the effective availability of the directors of divisions and offices. The sample has then been divided in an "almost" random way in two groups, the only constraint being the fact to have in the same group people from the same office: one group of 11 people experimented the introduction of OpenOffice (our treatment X), while the other group was used as a control group.

The possible presence of artifacts, should not have influenced significantly the results. A questionnaire has been submitted to both groups before (O_1) and after (O_2) the introduction of OpenOffice to evaluate the effects of

the experimentation on the attitude towards OSS. The activities of both groups have been constantly monitored by an automatic system for data collection that permitted the collection of a series of objective process data (the series of observations $O_{3,i}$).

b) *Quasi-experimental design*. On the other side more qualitative evaluations on the usage of OpenOffice and on Open Source have been made following the guidelines of the so-called “quasi experimental” methodologies [5]. The quasi-experimental design is used to evaluate the implications of a treatment (in our case the introduction of OpenOffice) without the typical requirements of a classical experimentation, like the randomisation and the presence of a control group. Clearly the results obtained are less significant than the one of a real experimentation, less generalisable and more sensible to exogenous effects. However, they can still give an indication on the effects of the experiment, in particular if the sample is not of negligible size, as in our case.

c) *One-shot evaluation*. The evaluation is based on a punctual evaluation of a situation without the presence of a valid sample. This evaluation is used when the presence of a sample of relevant dimension would be too onerous to be used in a single experimentation. The resulting data is in this way very “noisy” and subject to change in future replications of the experiment.

All the three designs have been applied, however mostly of the analysis of this paper comes from the classical experimentation design.

Two types of questionnaires have been submitted to users. The first identical before and after the experimentation, to understand the attitude towards Open Source and the effects of the experimentation on such attitude; the second has been submitted only at the end, where all the final results of the project have been collected and more information for the replication of the experiment has been determined.

As for the tools used, two different non-invasive systems for data collection have been employed:

- FLEA (FiLe Extension Analyzer) and DepA (Dependency Analyser) to analyse the different typologies of files present on the target system, eventual macros and the reciprocal dependencies among different applications installed;
- PROM (PRO Metrics)[6], to monitor the usage of OpenOffice and Microsoft Office.

III. THE EXPERIMENT

The experimental protocol of the introduction of OpenOffice used in this study was defined as a predefined set of steps:

1. A seminar to give a motivation on the reasons of the experimentation.
2. Installation of OpenOffice and translation of the most used documents in the OpenOffice format.
3. Training.
4. Usage help with support in case of need through a forum and an available hot-line.

5. Periodic verification meetings with the OpenOffice users to identify possible problems.
6. Automatic start of OpenOffice with the files carrying a Microsoft Office extension.

Regarding the protocol, the selection of the experimental groups has been done in a way to enable that the participants were, when possible, in some way in relation one with each other, physically near and if possible coming from the same organisational units, to take advantage of possible network externalities that arise in terms of document exchange, reciprocal help, creation of informal centres of competence and so on [7].

The experiment lasted for 32 weeks, the first 10 only Microsoft Office was monitored and the system dependencies were collected. OpenOffice was introduced and monitored until the end of the project, with week 23 delimiting the introduction of the OpenOffice automatic start with Microsoft Office files.

IV. DATA ANALYSIS

In this section we analyse three aspects derived from the data collected: the productivity of the new solution, the users' acceptance of OSS and the system interoperability issues.

A. Interoperability issues

The study of interoperability constraints is an important step before a migration. In our case, we performed a dynamic and a static analysis of the system dependencies.

The static analysis was performed at the beginning of the experimentation through surveys submitted to IT personnel, in particular constraints with Oracle and SAP software were thought to be relevant by the interviewed personnel.

The dynamic analysis applied with the aid of automatic tools, reported no serious dependencies in the systems involved in the experimentation. In fact the dependencies reported by the static analysis were found, but not so relevant in numbers.

These results cannot be generalised to the whole PA under exam and can be also be seen as a sort of limitation to the experiment, as in more turbulent environments, the results of the whole experiment may be different.

B. Productivity evaluation

The productivity has been evaluated using the data coming from PROM.

In Fig. 2, on the X-axis there is the week of the experimentation, OpenOffice has been inserted during week 10 and the automatic association has been activated during week 23. In blue the percentage of opened files using OpenOffice is represented and in orange the percentage of average time devoted to OpenOffice among users that effectively used OpenOffice.

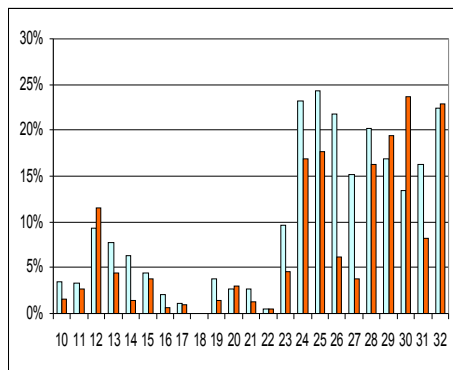


Fig.2. Increase in OpenOffice usage

In Fig.3 the effective number of users adopting OpenOffice is increasing. At the end of the experiment, every user was employing OpenOffice together with Office. Also in this case, the effective increase of usage happened after the introduction of the automatic start in week 23.

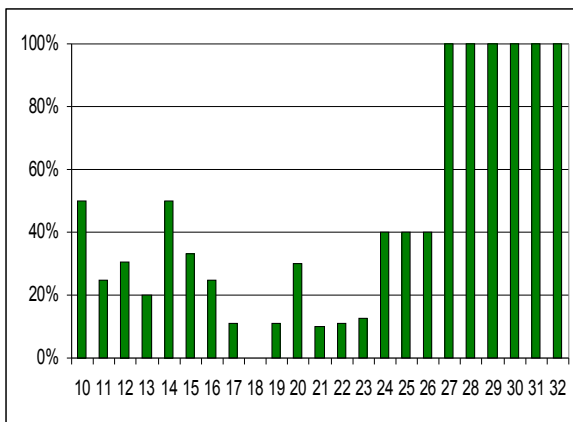


Fig. 3. Increase in the effective number of OpenOffice users.

To further study the loss of time and efficiency caused by the adoption of OpenOffice, we tried to find an answer to the following questions:

a) The usage of OpenOffice caused a reduction in the number of documents used per day? The correlations studied are the number of documents used each day and number of documents opened with OpenOffice. A negative effect on the usage of OpenOffice has to produce a negative impact on the usage of OpenOffice and a significant negative correlation between these two variables, that is the more documents are handled with OpenOffice, the less are globally handled;

b) The usage of OpenOffice caused an increase in the time devoted to each document? The time devoted to the management of the all the documents and to the OpenOffice. A negative effect on the usage of OpenOffice has to create a significant positive correlation, that is it should be evident that the more time is spent with OpenOffice, the more time is spent globally managing

documents, as OpenOffice required more time to accomplish the same tasks.

As for question a), the correlation between the number of documents opened with OpenOffice and the total number of documents has been $-0,08$ with an inexistent significance, the probability of error of the correlation is 66%, therefore it has to be excluded that the usage of OpenOffice has reduced the number of documents handled daily.

Relating question b), the correlation between time devoted to documents handled with OpenOffice and the total time devoted to document handling has been $-0,04$ with a null significance, the probability of error of such correlation is 82%, therefore it has to be excluded that the usage of OpenOffice has increased the global effort to handle documents.

The comparison with the control group confirmed furthermore that the evolution of the usage of documents among the test group and the control group has been consistent, excluding in this way exogenous factors.

C. Change of attitude

By analysing the results of the submitted pre- and post-test questionnaires on the attitude towards OSS, some interesting effects have been detected.

After the experiment, all users have an opinion about OSS, also those that were neutral have split in positive and negative towards OSS. The importance to know the instrument they are using has increased, as the importance given to training.

Half of the users reported OpenOffice as having the same functionalities offered by Office, while the other half considered it with less functionalities. However, when asked about a possible substitution, almost all said that it was suitable with no or very few problems.

V. LIMITATIONS

The data collected represents the results of a single experience and as it is the results cannot be systematically generalised, as the essential comparative aspect is missing.

The selection of the sample followed also some constraints imposed by the PA under exam.

The field of office automation in particular may not be comparable fully to other desktop environments where the open solution is not as strong as OpenOffice. As already reported, the focus of the study has not been on the substitution of the old solution, as the proprietary application has been maintained together with the new one. A scenario of a complete substitution would need a separate experiment.

VI. CONCLUSION

The placement of OpenOffice side by side with Office was evaluated with a gradual and measured approach, that includes techniques to ease the OpenOffice diffusion, like the automatic association with the Office extensions. The good results can bring in our opinion to a positive large-scale installation of OpenOffice in the PA considered.

However, a generalisation to other PAs or to other fields different from the office automation are to be taken with care, keeping in mind the limitations of the experiment. As a final note, the insertion of the experimental data in a TCO (Total Cost of Ownership) framework, like the one that is being developed by the COSPA project, can better help in evaluating the convenience of a full transition.

VII. ACKNOWLEDGMENTS

We acknowledge Dr. Hellmuth Ladurner for his precious help and support. Acknowledgments also to all the users involved, participants to the experiment, technical personnel, supervisors: without their constant effort and their availability this study would not have been possible.

VIII. REFERENCES

- [1] The COSPA Project, *Consortium for the Open Source in the Public Administration*, <http://www.cospa-project.org>
- [2] Everitt K. e Lederer S. (2001) *A Usability Comparison of Sun StarOffice Writer 5.2 vs. Microsoft Word 2000*, <http://www.sims.berkeley.edu/courses/is271/f01/projects/WordStar/>
- [3] Basili V. (1995) “*Applying the Goal/Question/Metric Paradigm*,” in Experience factory, in Software Quality Assurance and Measurement: A Worldwide perspective, ITP An International Thomson Publishing Company, chapter 2, pages21- 44
- [4] Lloyd J. (1999) “*Statistical Analysis of Categorical Data*”, Wiley
- [5] Campbell D.T. e Stanley T.D. (1990) “*Experimental and Quasi-Experimental Design*”, Houghton Mifflin Company
- [6] A. Sillitti, A. Janes, G. Succi, T. Vernazza, “*Collecting, Integrating and Analyzing Software Metrics and Personal Software Process Data*”, EUROMICRO 2003, Belek-Antalya, Turkey, 1 – 6 September 2003.
- [7] Shapiro C. e Varian H.R. (1999) “*Information Rules: A Strategic Guide to the Network Economy*”, Harvard Business School Press