Distr.
GENERAL

CES/AC.71/2004/17
10 May 2004

ENGLISH ONLY

UNITED NATIONS STATISTICAL COMMISSION and ECONOMIC COMMISSION FOR EUROPE (ECE) CONFERENCE OF EUROPEAN STATISTICIANS

EUROPEAN COMMISSION STATISTICAL OFFICE OF THE EUROPEAN COMMUNITIES (EUROSTAT)

ORGANISATION FOR ECONOMIC COOPERATION AND DEVELOPMENT (OECD) STATISTICS DIRECTORATE

Joint ECE/Eurostat/OECD Meeting on the Management of Statistical Information Systems (MSIS)
(Geneva, 17-19 May 2004)

Topic (iii): Open source and software consortia in statistics

OPEN SOURCE FOR STATISTICS – A STAIRWAY TO HEAVEN?

Invited Paper

Submitted by Statistics Norway[1]

1. In general, there is an increased interest in FLOSS (Free/Libre and Open Source Software). The statistical sector makes no exception in this, but the question is if we are able to draw upon the potential benefits from using FLOSS only as individuals or as a sector. At least we, as for the public sector as a whole, should lean to case studies and analyses done to demonstrate the risk and benefits of open software. Some technical and organisational matters will be discussed in the following, and hopefully add valuable aspects to an open source strategy of any target organisation. A concrete collaborative example formed by a number of statistical institutes, the SOS (Statistics Open Standards) Group will be described.

I. DISCUSSION OF OPEN SOURCE APPLICATIONS

Driving forces

2. In any large IT environment, commercial software has coexisted with open source software for years. The public sector is no exception. Heavy investments in e-government in the 90's made governments to look at open source for the perceived savings of these investments.

3. The belief of quick and inexpensive implementation usually is the main driving force of open source applications. Some see it as a shortcut to their technological independence and as a basis for their future IT capabilities. Further, commercial vendors supports open source software as a business strategy providing for add-on business, as consulting services, and add-on products, which are tactically charged.

[1] Prepared by by Rune Gløersen (rune.gloersen@ssb.no) and Hanne Mette Janson

4. A recent survey (2003) on behalf of the European Commission in European public administrations, found that 63 percent of the respondents said they already use open source in their software architecture. Web and file server features seem to be the most convenient areas of use.

Free/libre/open source software (FLOSS)

5. These are programmes whose licenses give users the freedom to run the program for any purpose, to study and modify the program, and to redistribute copies of either the original or modified program (without having to pay royalties to previous developers). The open source trademark is issued by the Open Source Initiative (OSI), when requirements are compliant. The terms Free Software and Open software are usually used interchangeably.

6. The FLOSS "products" are typically carefully driven and protected by dedicated communities, which seriously want to provide for serious product development.

Pooling software vs. open source software

7. There is no doubt that successful open source projects, as the Linux and Perl revolution, provide for a "share and reuse" best practice. When public-sector agencies stick together in large-scaled collaboration projects running by the same practice, optimised investments may be derived. Such software pooling is valuable and cost-effective provided that the IT management realises that flexibility, concerned with travelling, meetings and "give and take" mentality is provided for.

Interoperability

8. The main reason for retaining the dissemination of floss is the ability to process data produced by proprietary software (word processor components as tables, graphs) which normally are not based on open standards. This self-enforcing process evolves because the proprietary software vendor enforces a kind of de facto standard, despite of being closed.

9. On the other hand, the successful dissemination of Linux enforces the main vendors to provide for this platform, simply to be in business. For example; SAP, IFS, Intentia, and Oracle are all well off on their portfolio adjustments to Linux. Although, 5.000 applications are interoperable with Linux at the moment, most organisations need Unix/Microsoft servers running in parallel, to provide for proprietary software not interoperable with Linux.

Quality and flexibility

10. The open source development traditionally has been mainly conducted by public R&D and university funding. The combination of public and stable funding, highly skilled technical personnel, and unique software requirements of scientific research projects, used to assure for the quality momentum. A recent survey (http://www.infonomics.nl/FLOSS/report/) shows that 65% of FLOSS developers are employed, thus trusts and consortiums with a clear understanding of the FLOSS development model, seem to overtake this leadership. The combination of an emerging new business model and companies with a clear understanding of market needs provides a driving force for high quality software and open standards. Sun, IBM and Novell are heavy driving forces both in the development of core Open Source projects like the Linux kernel and desktop environments like GNOME and KDE.

11. In mainstream public-sector deployments, the most important reason for the adoption of open-software seem to relates to quality issues (stability, maturity and compliance with standards) and, to a lesser extent, flexibility (modularity, integration, customisation).

Security

12. As far as the security and reliability level concerning Floss software vs proprietary software is discussed, a slight paradox may run into mind. One should believe that the licensed products took care of backdoor traps, as well as conventional bugs. This may actually never be perceived, simply because the code is hidden, and therefore security inspection is impossible. In this case the users are left on their own to handle loss of data, wrongly executed commands, or defacements. Due to the fact that Floss software is open, code inspection and encountering of security gaps may provide for higher reliability and security. This awareness is about to fortify in military software applications.

Economics

13. The quest of cost savings, due to free up-front license fees and the availability of community based support; tend to be the major force for open source applications. This effect seem to be enforced in organisations where labour and support cost are low, and thus license fees of commercial products will turn to be a more significant cost factor in the budgets. Thus, there is no surprise that (European) public sector organisations have above-average use and planned use rates, compared to other for-profit sectors.

14. An indirect benefit of the consideration of open source in public-sector procurement is that it can be used to boost a user's negotiation position. This is particularly important in product or geographic markets dominated by a single vendor. Enterprises can deploy open source software on a small scale within their organisations to break a monoculture, reduce their vendor dependency, and thereby negotiate better terms for commercial software deployments. Although, this approach likely will lead to increased costs due to management of a more heterogeneous environment. Such costs therefore need to be carefully balanced against potential savings. On the other hand, proprietary software is addicted to “vendor follower” costs. Such may be mandatory upgrades, regardless if there is any internal interest of doing so.

Organisational skills

IT management skills

15. In lack of IT-strategy guidelines, it's likely to believe that the developers introduce different open source products tactically in a case-by-case manner to solve a specific requirement. From there, these products are supported on an "ad hoc" basis. No surprise this will happen, when IT management does not address the open source discussion formally. Thus, minimum management skills of the subject should be required, especially;

- *The IT lock-in situation*; concerning vendor dependencies and the cost and skills required for frequent, and often unintentional upgrades when dedicated to proprietary software

- *The balanced cost aspect*; considering that free software not solely implies cost savings when running a heterogeneous IT-portfolio.

- *The security and reliability requirements*; awareness of the different security levels required in various applications of the organisation. Hereby military and health-care systems may be typically prioritised areas for Floss applications, due to the security inspection and threat encountering possibilities

Developer skills

16. The proportion of developers connected to Floss communities differs considerably within the European countries, bringing France, Germany and UK on top. At the bottom line, this aspect may be the core success criteria related to Floss dissemination, believing in the bottom-up strategy for healthy and effective usage of

open software. Related to this, one should also interpret innovative capabilities and best practice acquirements. Indeed, it is likely to believe that the success of open source projects very much depends on utilizing best practice techniques; related to version management, testing, security and reliability assessment, etc.

17. It may also be the case, that project managers should be encouraged to possess the same open source awareness and strategies as the IT management.

Governmental initiatives

18. There may be various views on to what extent the governmental bodies should guide and control Floss applications in the public sector. France and Germany, for example, have established dedicated coordination and advisory organs in order to assist the different departments in their open source implementation process. According to the FLOSS Survey of Developers (2002), France and Germany also are the European countries with the highest developers' activity into Floss projects.

19. There may be several ways for the Government policies, either on local, national or regional level to encourage standardisation and innovation processes. One can foresee that the development cycle may be driven from the demand side, letting a large firm seed a standard by sponsoring or releasing core code. When a large buyer asks and pays for certain features that may otherwise be commercially unavailable, an existing developer base may be commercialised. It seems to be the fact that open source systems that have developed into standards, tend to be developed by small groups. Having a large base of open source development therefore helps a region's companies involve themselves early in the de facto standardisation process.

20. In broad outline, the European countries, regarding both policy and implementation seem to be quite heterogeneous. Single governmental institutions decide on themselves whether, and to what extent they use open source software. Mainly, the same intentions exist: interoperability, security and cost reduction. As a consequence, the reflections regarding Floss are to be made by the different Floss users on their own. The Statistical sector as a whole is not in the position to take specific stock on the use of Floss. The reason for this is mainly the lack of a common business model widely used and implemented. Any advantages must be considered upon the specific economic, technological, knowledge based and architectural conditions of the respective organisations.

II. THE SOS GROUP

21. Statistical institutes have a long and fruitful history on exchange of best practices concerning organisational and professional issues. Nevertheless, there are few examples of binding collaborative initiatives that have been proven fruitful over time. Both were NSIs have been forced to collaborate, or where people or professions within NSIs have entered into cooperation on a voluntary basis, these initiatives have had a tendency to vanish into thin air. In most cases people bring with them best practices as ideas for implementation in their home office. The ability to implement and cost of implementation will vary dramatically from NSI to NSI, because of the difficulties to apply to internal architecture, technology and knowledge.

22. However, there is a need for the management level of the NSIs to follow up initiatives made from discussions in different sectors and professions within the institutes. The SOS group was established from this need, as described in the following.

The birth

23. The idea of the SOS Group was from the beginning nothing else but an initiative to cooperate and specifically to share best practice examples between the Nordic NSIs within the field of IT. Over the past years, IT experts had been meeting each other in different groups, making there own decisions based on their interpretation of both NSI strategies and common technological trends. There was a clear lack of

communication on the IT management level, in order to catch up with internal opinions without to much filtering of information. In 1998 there was a first meeting in what we then called the Nordic IT Group (Nord-IT). Typical fields of common interest were the Swedish/Danish solution for Statistical databanks, metadata strategies, GIS and especially the common obligations on international data reporting.

24. In June 1999, Nord-IT had a joint meeting with Mr. Wouter Keller from CBS in the Netherlands. Focus was set on organising a joint venture on software development and software sharing, perhaps closer to a principle of software consortia than according to open source principles. The name was changed to Statistics Open Source Group, and CBS Netherlands and the statistical office of Switzerland joined the group.

25. It was decided to start the SOS co-operation stepwise. The purpose was to increase the re-use of statistical software components for the total statistics production model. There was a hope that available software packages or solutions should dissolve into components that could be (re)used together.

26. The method was to take stock of the potentially exportable components and to define interfaces. Other principles were: Steal as much as possible, get quick results (!)

27. But we also have to mention the main restriction, which is quite eccentric in terms of open source discussions; the co-operation was delimited to development for the PC/LAN platform, based on the Microsoft software family. However, not all NSIs were completely on a Microsoft platform, in Statistics Norway for example, the statistical production is mainly carried out on a Unix platform, but we joined the group anyway, leaving any platform problems to be solved in practise. In the initial phase we concentrated on metadata and output (dissemination).

28. It was agreed that each component should be the sole responsibility of one NSI or partner: One partner is committed. The basic idea was that this country developed what was important for its own use, what it would have developed anyway. The components should normally be free of charge for the partners. We were not planning for components developed in consortia.

29. he group were organised by key persons in the respective NSIs, and with no secretariat or other commonly funded personnel. Furthermore, there was agreed upon two levels, one *technical level* where experts should meet and report to the *high level meeting* (Steering Committee).

30. In practice, the cooperation concentrated on the following:

- A common project for the future developments of the Swedish/Danish Statistical Databank solution. Statistics Norway entered into the cooperation. The database model is commonly agreed between the partners. A number of interfaces exist as commonly used components. This covers especially a number of PC-Axis components that are reused in different user interfaces, while web end-user interfaces are locally developed and managed.
- Further development of the Neuchâtel model for statistical classifications.
- Investigation in the Dutch Cristal model to look at potential implementation within the different NSIs.

31. The meetings of the Steering Committee have taken place approximately twice a year.

The Puberty

32. The SOS Group has proven successful for the partners in some areas, while the outcome has been less than expected in others. A key element already mentioned while establishing the partnership was to define standard interfaces, or to take stock on such already defined. This is a key to success in any field of sharing knowledge or technology. The Gesmes/TS message for reporting aggregated data is one example of such interface, where the SOS Group have joined forces in order to develop solutions for common use.

33. If we take a look at the fields of cooperation mentioned above, the success of the cooperation of the Statistical Databank is due to the fact that we stick to a common database model as a core, and that we are able to share components based upon standard interfaces to this model, i.e. for reporting data to international organisations. Nevertheless, the cooperation is only between three of the partners, while the others have their own solutions.

34. On the other hand, the cooperation on the use of Cristal vanished. The main reason for this is perhaps that Cristal represents a core business model, where you must change a number of existing data- and production models in order to take stock on the potential advantages. Any solutions where you need to redesign your business model is not implemented nor accepted over night. One other reason was perhaps the binding to the Microsoft platform.

35. The Neuchâtel standard was originally limited to classifications, but has been extended to cover other areas like variable definitions. However, the cooperating members are limited, and such standards are competing with a number of others, like ISO 11179, DDI, Dublin Core etc. None of them really represents a break-through regarding usage. The Neuchãtel standard is implemented in a common Classification database developed by Statistics Norway and Statistics Denmark.

36. Another area that turned out to be quite successful, even though outside the original scope of the group, was a common investigation of the SuperStar software from SpaceTime Resaerch Ltd. In this case the Group (all except one member) acted as one body during evaluation of the software, and finally during negotiations with the vendor.

37. The common usage of software, combined with the common model for statistical output databases and to some extent common use of metadata standards have shown to be very useful for the Group. Concrete interfaces have been developed, and there is increasing number of examples where one country develops modules that easily can be reused by other partners.

Becoming adult

38. During the first years, the SOS Group changed its focus from components to interfaces. This has been reflected in the name, which currently is The Statistics Open Standards Group. This reflects a more or less adopted experience that the key to cooperation is on the development and usage of common interfaces. All concrete examples of cooperation have been on developments either of specific interfaces, or software modules that can be used by others because they are developed in accordance with such interfaces.

39. The SOS Group has established a technical group on XML. For the moment, we expect a lot from sharing expertise and examples of implementations between the NSIs. Especially interesting is the examples of XML based standard interfaces that support the production cycle of statistics, and the use of some common available XML standards, like XBRL and XML4DR within the different NSIs.

40. Another area of cooperation in the near future will be connected to the growing demand for web services. A web service however, is a standard interface to a non standard world. Each implementation will be specific to the underlying source of information. This means that cooperation could be limited to sharing knowledge and best practise. However, if more of the statistical production is based upon elements of a core business model, NSIs could to a larger extent take stock on web services as completely reusable modules.

41. Moreover, the main reason that the SOS Group still exists, is because there is an understanding between the partners that successful projects will come and go. The long-term perspective of such collaboration, where you will be in position to react correctly and in due time on each others ideas, projects and products, is probably regarded as most valuable. However, such cluster of interested parties should be kept small and targeted, also enabled to act as necessary meeting places between the respective NSIs, and the

number of formal Working Groups, Committees and other bodies where NSIs meet according to 3. parties agendas.

42. Be aware that any conclusions and reflections on the experiences from 5 years of work in The SOS Group are the opinions of the authors only.

III. SOME CONCLUSIONS

43. From the general discussion, a set of strategy aspects should be addressed concerning Floss, including:

- License conditions of software (extent of source code access, reproduction, interoperability, amongst others);
- Matters of exceptional usage, when license conditions will not be fulfilled;
- Preferred application areas within the organisation's infrastructure (i.e. security and reliability applications, internet applications);
- Organisational skills requirements;
- Adaptation to governmental overall strategy, if any.

44. The SOS Group is an example of a joint interest group among certain NSIs, where the principle of software sharing and exchange of best practise is the core business. The specific NSIs have joined forces out of their own interest, because they have identified a number of areas of similarity, which is regarded as a basis for share and reuse of software and knowledge in practise. Once assembled, the group widely draws upon the opinions, strengths and knowledge of each other.

45. The SOS Group represents an initiative according to a prerequisite for an open source share and reuse best practice as it was mentioned in the general discussion:

There is no doubt that successful open source projects provide for a "share and reuse" best practice. When public-sector agencies stick together in collaboration projects running by the same practice, optimised investments may be derived. Such software pooling is valuable and cost-effective provided that the management realises that flexibility, concerned with travelling, meetings and "give and take" mentality is provided for.

46. However, the statistical sector is still heterogeneous, and whether open source is a stairway to heaven or not, is a question to be answered by each NSI, or clusters of NSIs running under similar conditions, rather than for the statistical sector as a whole.

- - - - -