Distr.
GENERAL

WP. 30
12 May 2011

ENGLISH ONLY

UNITED NATIONS ECONOMIC COMMISSION FOR EUROPE (UNECE)
CONFERENCE OF EUROPEAN STATISTICIANS

EUROPEAN COMMISSION
STATISTICAL OFFICE OF THE EUROPEAN UNION (EUROSTAT)

ORGANISATION FOR ECONOMIC COOPERATION AND DEVELOPMENT (OECD)
STATISTICS DIRECTORATE

Meeting on the Management of Statistical Information Systems (MSIS 2011)
(Luxembourg, 23-25 May 2011)

Topic (iv): International cooperation/collaboration

Cooperation Models for Software Development

Invited Paper

Prepared by Valentin Todorov, UNIDO

I. INTRODUCTION

1. Collaboration and data exchange between organisations (international and/or national) inspires and at the same time benefits from collaboration in software development and software sharing between these organizations. Using common software solutions results not only in savings in time and money but can enhance mutually the institutional knowledge and promote and enable the implementation of statistical standards. The advances in using statistical standards like SDMX will in turn facilitate the data exchange mechanisms and thus improve the data quality on both sides.

2. Software sharing and collaboration in software development among statistical organizations is nothing new, especially in today's global economy. Different practices and models involving different statistical organizations already exist, as seen in many papers presented at previous MSIS meetings. PC-Axis with Statistics Sweden as the vendor and major stockholder, OECD.Stat statistical data warehouse developed by OECD, as well as several Open Source initiatives are just a few examples. These examples show that in spite of all benefits of the software sharing there are also important issues and socio-technical barriers that affect successful collaboration in this inter-cultural environment.

3. The purpose of this paper is to review and evaluate the currently known models of software sharing in the domain of official statistics and to draft common guidelines and best practices. The focus is on several main issues and opportunities, like licensing, position in the statistical business process model, ownership and governance, sustainability, methods for distributed software development and technical communication advances.

II. STATISTICAL COLLABORATION MODELS

A. The Software Inventory

4. A new Software Inventory platform was launched by the *Sharing Advisory Board (SAB)* with the purpose to collect information about existing shared products, products under development or even products that are undergoing planning. The list was compiled in cooperation with the ESSnet project on a Common Reference Architecture (CORA). The inventory already contains information and contact details for over 50 statistical software applications available for sharing and can be visited at: http://www1.unece.org/stat/platform/display/msis/Software+Inventory.

5. There are several more products for which information was available in the *Sharing Advisory Board* but which are not yet in the inventory. Most of these – *eDAMIS*, *SAM & SEV*, *SDMX Reference Architecture* - are developed by EUROSTAT and are provided for use to its member states free of charge. Since these shared products present one specific model of software sharing we will consider them also in this study. Another product which is not yet in the inventory but is successfully shared as OSS is the developed by ECB *FLEX-CB*. The interesting in this product is that it is one of the most widely used and shared "true" OSS products in the area of official statistics.

6. For each product some basic information is recorded together with more technical specifications like development status, code availability, charges, multilingual support, programming language used, operating system and third party software used. Some of these data are straightforward (the code can be either available or not and charges may apply or not), but others could be quite subjective. For example the development status is given as either *in development* or *in production/stable* but it is not clear at all where the border between these lies. While some of the products are already well established, used in many statistical offices and have a wide user community (*Banff, Blaise, PC-Axis*), others are on the way to this (*OECD.Stat, FLEX-CB*) and thirds, even if used in production in the statistical office that developed them have still to be enhanced to become mature for international collaborative use (*SDMX ISTAT Framework, SAM/SEV*).

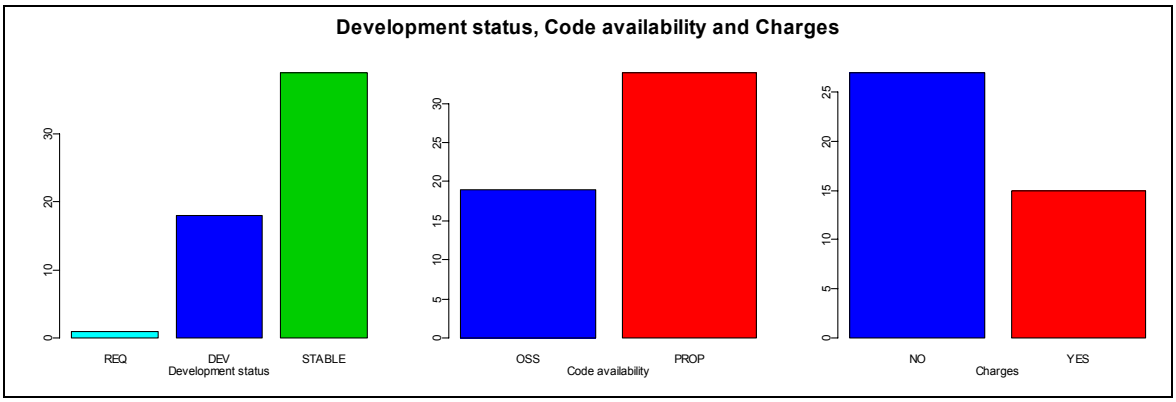

Figure 1: Development stats, Code availability and Charges of the offered software products

7. In Figure 1 the development status, the code availability and the charges applied for the different products offered for sharing are presented. Products for which the required information was not supplied or is not clearly interpretable are not counted. We see that almost half of the products are announced as *in production* (only one is in the *requirements/specification* phase). Also, about half of the products are provided as open source and for most than the half of the products no charges apply.

8. It is interesting to compare the code availability and the charges to the development status and this comparison is presented in the mosaic plots in Figure 2. The source code is given and no charges are applied to much higher extent for the products which are in development than for those that are stable, in production. Does this show a new trend of openness and collaboration, for the newly developed software or is it so that as soon as the new products become stable the policy will change?

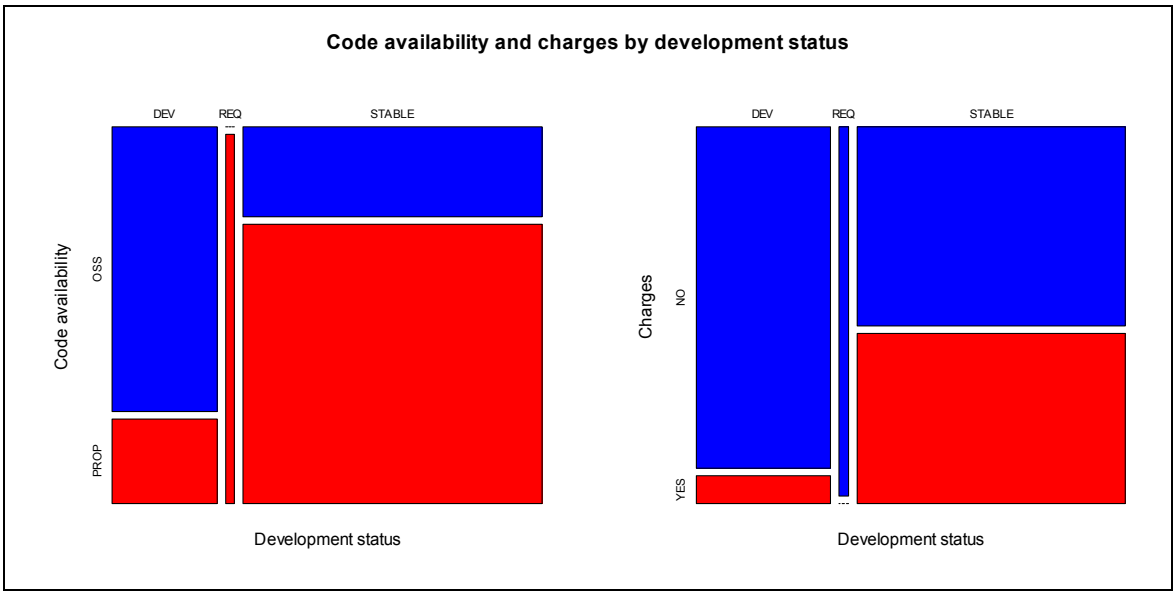

Figure 2: Cross-tabulation of the Code availability and the Charges of the offered products vs. their development status

9. When talking about sharing and collaboration it is also important to consider the programming environment used. The distribution of the products according programming language or environment is shown in the leftmost panel of Figure 3. There are three main groups of products – developed in SAS, in JAVA or in anything else (VB, C, C#, .Net). There are only three products written in R and only one of them states to have a stable version in production. (Of course there are many more R packages at CRAN – see C for details).

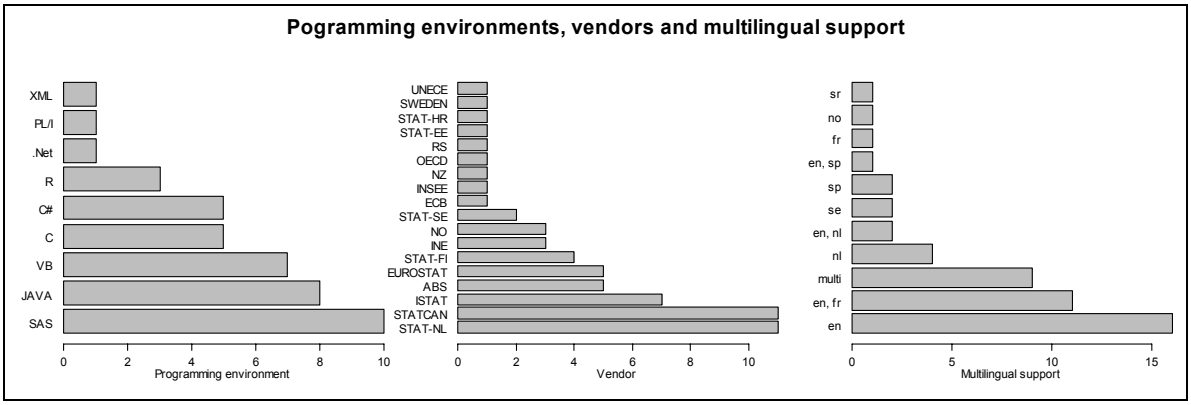

Figure 3: Programming environment, and multilingual support of the offered products

B. Types of Collaboration Projects

10. If we consider the information about licensing, code availability and charges applied as well as the development history of the products which are offered for sharing we can identify roughly three distinct models:
    a) Software developed by one organization (either for own needs or for sharing) and released either as Open Source Software or as Free Software (without access to the source code);
    b) Software developed by one organization and offered for collaborative sharing for a fee;
    c) "True" Open Source Software.

11. Examples for the first case are most of the tools developed by Eurostat. *eDAMIS* is Eurostat's system for data exchange with countries and all its components are developed by Eurostat. Part of them (mainly *eDAMIS Web Application EWA*) are to be installed in Member States. The

development is driven by user needs. A more complex case is the product *SDMX Reference Architecture*, whose main objective is to provide a specification of a generalized architecture to be used partially or as whole by member states interesting in starting SDMX projects and willing to participate in Eurostat's *pull* data transmission (an example could be the Census Hub).

12. The second group comprises the most prominent shared products like *PC-Axis, Banff, Blaise* and recently *OECD.Stat*. These packages have different development history, licensing models, user groups, etc. and here we will consider examples.

13. ***PC-Axis*** is a suite of software designed for disseminating and visualizing statistical data (GSBPM 7.x). The core components (*PC-Axis* and *PX-Web*) were developed by Statistics Sweden in late 1980's and some statistical organizations have developed and contributed additional components (for example, PX-Edit - Finland, PX-Map - Norway, PAX-Admin - UNECE). There are currently around 40 members of the PC-Axis Reference Group, mostly national and international statistical organizations. Organizations using PC-Axis components for web dissemination are required to pay a *license fee* which is calculated relative to the GDP and population and varies from country to country. The money is used to help fund PC-Axis development and support activities in Statistics Sweden. Additional components are available at no cost from the developers, and can be downloaded from http://www.pc-axis.scb.se/. All PC-Axis license holders are entitled to attend the annual meetings of the PC-Axis Reference Group. This group reviews new developments, and determines priorities for development work over the coming years. The use of PC-Axis tools is growing rapidly. This has led Statistics Sweden to start considering how to spread the burden and responsibility for developing and maintaining the core software. A task-force of Nordic countries is considering this issue, and will shortly make recommendations for new governance arrangements.

14. ***OECD.Stat*** is relatively new software which was initially developed for OECD's own needs but has since been adopted by other organizations on the basis of bilateral agreement. Examples of sharing are the National Statistical Institutes of Australia, New Zealand and Italy, and the International Monetary Fund (see [16] for further details). The software is designed for disseminating and visualizing statistical data (GSBPM 7.x). Interested organizations enter into a Memorandum of Understanding (MOU) with mutual benefits for both OECD and the collaborating partner. Two types exist: one with no end date, and one with a fixed period. Contributions to the community are made by way of financial, secondments and joint developments. A collaboration community is currently being developed with the objective to build a community spirit of sharing and working together for a common goal. Most statistical organizations have very similar needs and although OECD does not promote *OECD.Stat* as the only option, it does facilitate the use of common standards (e.g. SDMX) and tools. An annual workshop takes place each year to bring the user community together and discuss developments, share ideas and knowledge. Partners are able to help shape the future roadmap by providing input both on developments and community activities. The solution is constantly being reviewed and maintained to ensure it remains relevant to technology and meets user needs. OECD provides initial training and support, usually remotely, but on site visits can be arranged. Collaborating partners are also given access to an online community portal where all members can respond to support requests. OECD provides a dedicated support person to manage the collaboration partners' requests. System and user documentation is also made available through the portal.

15. ***Generalized Systems Suite (StatCan):*** Another example from this group is the suite of generalized statistical systems of StatCan which the agency has maintained for many years and now is reviewing these products. The new suite will comprise products with names prefixed by *G-*, i.e. *G-SAM* for *Sampling*. Details about the suite and the plans of its revamping can be found in [12]. StatCan is inviting interested agencies to participate in the development of the requirement of the new system. The licensing of the suite and the conditions for sharing are bound by the fact that by law StatCan must license the use of any software product developed by the Government of Canada. Currently StatCan is reviewing its licensing practices to see if co-development and open source models are feasible and to modernize the approach used to determine the fees applied to our supported products. The products of the suite are fully supported and are offered for a fixed fee (if they are to be used for production purposes) which will be used to cover the costs associated with managing the sharing of the products.

16. ***FLEX-CB**:* This is the most typical representative of the third group, "true" OSS. The purpose of this project is to create useful visualizations of statistical data from institutions that employ SDMX. The development started by reusing the ADOBE FLEX source code already developed by the ECB. One of the key goals of the project is to collaboratively make improvements to the visualizations that may include improved interoperability with different data sets (real time, calculators output etc.) and improved information support for an expanded user base. A more general purpose is to advance the creation, evolution, promotion, and support of this source code while cultivating an ecosystem of complementary products, capabilities, and services. The main and most impressive show case of this product is the inflation dashbord at the ECB web site. The deliverable is a library and API, based on Adobe FLEX, for visualization of statistical data in SDMX format (phase 7. Disseminate of GSBPM). The product is released as open source under the New BSD License and the source code is available at Google-code: http://code.google.com/p/flex-cb/. The following link titled "Where Is It Used" provides more details and links to the available implementations of the library.
17.
18. ***R packages at CRAN***: to the third type of projects belong also many R packages which are available at the CRAN repository (see C below and [11]). Including them in the present study is out of the scope of this paper but it worths the effort finding the links to other software for official statistics in the Software inventory.

C. The CRAN Task View for Official Statistics

19. A number of packages developed in the R statistical environment [6] for tasks relevant for the activities performed in the National and International Statistical Offices exist and are available at the R repository CRAN, but it is not easy to cope with the fast growing number of packages (more then 1500 currently). In an attempt to structure these wealth of statistical functions the so called CRAN Task Views were created – see Zeileis (2005). This mechanism allows creating collections of packages which belong to the same area so that an interested user can easily search for a given functionality. The Task Views are supplied with relevant descriptions of the packages which are further grouped in functional areas. It is possible to install or update all packages from a given task view or simply to browse and read the descriptions in search of the desired methodology.
20. Recently a Task View dedicated to Official Statistics was added which covers many of the areas of interest. This CRAN task view contains a list of packages that include methods typically used in official statistics and survey methodology. Many packages provide functionality for more than one of the topics listed. Therefore this list cannot be considered a strict categorization and some packages can be listed more than once. Actually all topics that are of interest in official statistics are listed there, even if no related R packages are available so far. The following list gives an idea of the topics covered. Currently more than 40 packages are listed there:
    - Complex Survey Design: General Comments
    - Complex Survey Design: Details
    - Complex Survey Design: Point and Variance Estimation
    - Complex Survey Design: Calibration
    - Editing and Visual Inspection of Micro data
    - Imputation
    - Statistical Disclosure Control
    - Seasonal Adjustment
    - Statistical Record Matching
    - Indices and Indicators
    - Additional Packages and Functionalities
21. Some of the most well established (in the respective area) and widely used packages are: *Amelia, impute, mice, RecordLinkage, robCompositions, rrcovNA* (see [11]), *sampfling, sampling, sdcMicro, survey (core), VIM, x12*. Other task views like TimeSeries, Econometrics and SocialSciences could be also of interest.

D. The Role of GSBPM

22. In 2009 the joint UNECE/Eurostat/OECD group on Statistical Metadata (METIS) adopted and published a Generic Statistical Business Process Model (GSBPM) as a tool for describing and benchmarking statistical production processes. GSBPM is establishing itself as a reliable tool for blueprinting the structure of the processes in the national and international statistical organizations and more and more organizations are adopting it for modeling their organization and production process (see Work Session on Statistical Metadata, Geneva, 10-12 March 2010). Thus GSBPM is rapidly becoming a *defacto* global standard. The main purpose of the developing of the GSBPM was that it should provide a basis for statistical organizations to agree on standard terminology to aid their discussions on developing statistical metadata systems and processes, however it is apparent that the model can be very useful in other areas like (for many more applications see [5]):

- To harmonize the statistical computing architectures since the GSBPM identifies the key components of the statistical business process, promotes standard terminology and standard ways of working across statistical business processes. The potential of the GSBPM as a model not only for the statistical production process but also for statistical computing architectures was evaluated further in the European Union *ESSNet* project on a Common Reference Architecture (see Vaccari, 2009 [1]).
- To facilitate the sharing of statistical software. By fostering the standard terminology and defining the components of statistical processes GSBPM encourages the sharing of software tools also between different statistical organizations that apply the model. It therefore provides an input to the *Sharing Advisory Board*.
- To provide a basis for explaining the use of SDMX in a statistical organization.
- To provide a framework for process quality assessment and improvement. The GSBPM provides a mechanism to standardize processes as much as possible in order to apply a benchmarking approach to process quality assessment.

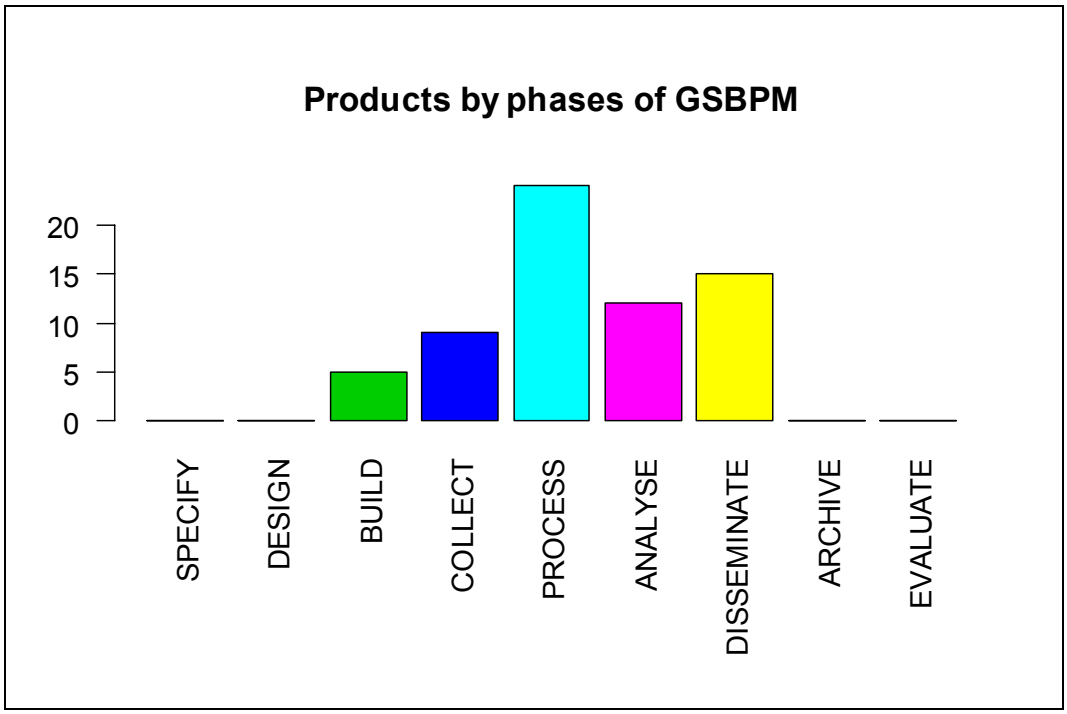

Figure 4: Diagram of the distribution of the products by the phases of the GSBPM

23. The distribution of the products according to their role in GSBPM is shown in Figure 4. No products are offered covering the phases *Specify, Design, Archive and Evaluate* and the majority of the products lie in the phases *Process, Analyse* and *Disseminate*.

E. Licensing Issues

24. For successful sharing of software between the different national and international statistical organizations it is necessary to define the commercial and legal foundations for the exchange of software. Since these offices are not primarily intended to do marketing for their software packages – although some of them do so – it is obvious that the Open Source Software (OSS) approach is a viable alternative for such exchange activities. However the use of OSS could be limited by different factors like the particular legislation of the country (see paragraph 15 and [12] for an example of the case of Canada), the Total Cost of Ownership (TCO), some other legal questions of "compatibility" between OSS rules and the rules of the organizations, etc.

25. One of the work packages of the ESSNET CORA project [13] studies the possible licensing models of sharing applications among the European statistical organizations, explains and discusses the current models of software distributions (open source models, joint development consortia, ownership limited to executable, free executables etc.). In detail are described and explained the available licensing models - the GNU type licenses (http://www.gnu.org/licenses/gpl.html) and the Apache/BSD type licenses (http://www.apache.org/licenses/) as well as the European Union Public License (EUPL) (http://ec.europa.eu/idabc/en/document/7330.html). The document considers also the questions about further development and support including training for OSS software and describes different ways to use and improve OSS like bartering, co-development, freeware and proposes modifications to the traditional OSS approach to make it suitable the statistical organizations.

F. Multi-lingual Support

26. When software is developed in the national context for internal use in an organization, usually due to the budget, resources and time constraints, multi-lingual support is either not a requirement or its priority is very low. Therefore the multi-lingual support rarely becomes a key part of the software architecture from the beginning. The lack of multi-lingual support is usually a major barrier to sharing (statistical) software. Exceptions are countries with official bilingualism like Canada where all software systems used by public servants or by citizens must be offered in both official languages (English and French).

27. The experience of *StatCan* with some considerations that development teams should take into account when designing and implementing a bilingual application was presented in a recent article by Karen Doherty [12]. Three scenarios are considered – from the simplest when all controls are displayed in the both languages, through a more complicated when the user once (by starting or installing the application) selects the language to the most complicated when the user can switch between languages at any time without loosing the context. It is clear that the second and third scenarios are not limited to the bilinguality but can be easily extended to multiple languages. It is stressed that the multi-lingual issue must be approached early in the design of the system and become a key part of the architecture thus avoiding major refactoring later. The preferred way of implementing multi-lingual support is to apply frameworks and relevant tools. The experience of StatCan is reflected in the right panel of Figure 3 which shows the multilingual support of the products from the software inventory. There are 11 bilingual (Englis/French) products offered by StatCan. There are also several other bilingual products (English/Dutch, English/Spanish) as well as about ten products which are declared *multi-lingual* – not a wonder that *Blaise*, *OECD.Stat* and *PC-Axis* are among them.

28. The *Sharing Advisory Board* is preparing Principles and Guidelines for Developing Multi-lingual Statistical Software which will be presented at the MSIS'2011 meeting and will be published at the wiki.

G. Support, Documentation and Training

29. An important component of the total cost of a product or system including all direct and indirect costs is the user support, the available documentation and training activities which could reduce the steepness of the learning curve. This is one issue on which the open source software products are

mostly criticized. Although there are mailing lists which could provide help within minutes like the R-Help list of the most prominent statistical open source product, the organizations prefer dedicated support by the software vendor. The availability of *training materials* and *training activities* is also important for the quick introduction of a new system. If the software product is not free of charge, the availability of evaluation version will help the potential user in deciding about the applicability of the software for the particular organization needs.

30. If we consider our list of products offered for sharing at the Software inventory we see that for less than the half of the products support and training are available. Of course those that offer support and training are in first line the prominent *Blaise*, *Banff*, *PC-Axis*, *OECD.Stat* and *eDamis*. Support and training are offered for most of the products comprising the *Generalized Systems Suite* of StatCan. *FLEX-CB* acts in the same way as most of the open source projects - the source is available for anyone to use, modify and contribute and a help forum provides for quick support granted by the few involved developers. However this cannot compare with open source projects with very large user base. Its documentation also follows the trend of the OSS movement – by being quite sparse.

31. All of the products state that documentation is available and in most of the cases links to the documents are provided but we cannot evaluate this criterion without looking at the particular content.

III. STATISTICAL COLLABORATION GROUPS

32. The software sharing and collaborative software development is closely related to the existing active international groups in the fields of official statistics. Recently an inventory giving an overview of the different international groups whose work is related to the enterprise architecture of statistical organisations was published at the MSIS Wiki page. The groups have been asked to complete a template outlining their activities and outputs which information is also available at this page. These groups and their relationships are shown in Figure 5.

33. Several of them are especially interesting for the topic considered in the present study because their main objective is to foster collaboration approach in striving for improvement in Statistical Information Management and to facilitate better cooperation in the field of development of software and uptake of IT related standards within the respective statistical production systems. These are the *User groups* of different statistical software products, the *Statistics Open Standards (SOS) Group* of Nordic countries and the informal CSTAT group *Statistical Network* including ABS Australia, StatCan Canada, SCB Sweden, SSB Norway, ONS UK and SNZ New Zealand.

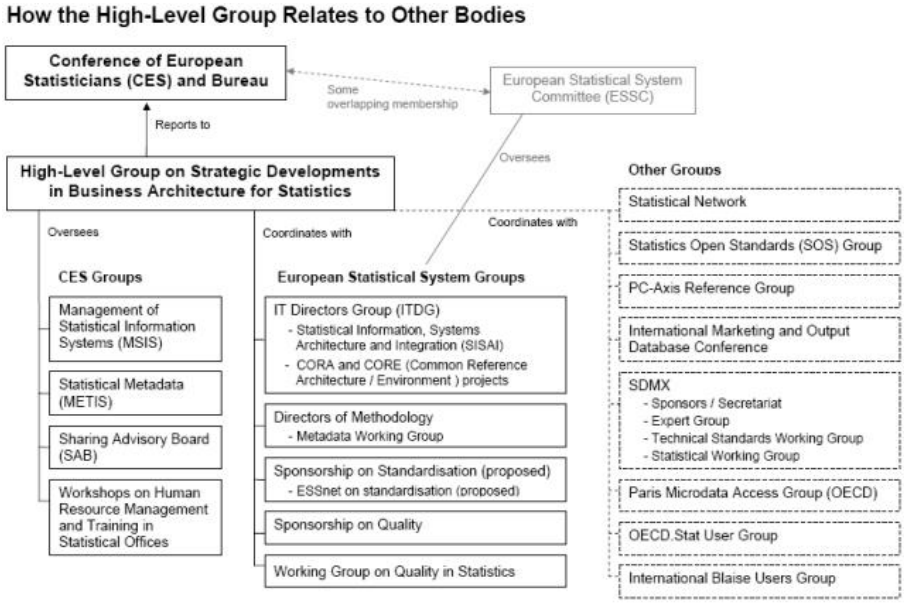

Figure 5: Relations between International Statistical Collaboration Groups

a. Statistical Network

34. The *Statistical Network* was formed under the leadership at the *Informal CSTAT Workgroup on Stronger Collaboration on Statistical Information Management Systems* in Paris last June and its main objective could be formulated as: "Working together with pace and passion to better meet our societies' information needs while driving down costs" or "Harmonising statistical methods, systems and capabilities across statistical agencies". The participating countries, Australia, New Zealand, Sweden, Norway, UK and Canada, agreed to share knowledge on the following five opportunities - see Value Creation Group (2010):

a) Innovation in Dissemination (New Zealand),
b) Confidentiality and Disclosure Control (Sweden),
c) Common Metadata/Information Management Framework (OCMIMF) (Australia),
d) Editing (Norway) and
e) Web Data Collection (Canada).

35. The fourth opportunity – the industrialization of editing – has already started to take shape. The partners shared their vision and a business case will be presented as part of agenda topic (vi) *International collaboration*, of the 2011 UNECE data editing work session. These projects have been selected to test the Statistical Network collaboration approach and are being conducted with the following principles:

- Maintaining commitment to the scientific principles underlying the statistical business,
- Using existing models and standards and evolving these where necessary,
- Early sharing of relevant plans,
- Building projects with sharing and re-use across the whole community,
- Share designs,
- Start small, learning, sharing and delivering frequently,
- Exploration of the strengths of all disciplines within statistical industry.

36. Collaboration will be in practical small steps to industrialize methods and processes to quickly and effectively benefit all participating NSIs. The objectives of the selected opportunities (projects) are to conduct a small exploration at the operational level to trial this approach. It is important to note that a number of the Statistical Network NSIs will contribute to each of the elected projects, although there is no requirement that every NSI in the Statistical Network needs to contribute actively to every initiative. Each of the initiatives is being led and co-ordinated by a different NSI. Whilst early collaboration has commenced by participating NSIs on these initiatives, it is recognized that these collaborations do not exist in isolation and NSIs have other collaborative work / partnerships within the standards communities for example. The Statistical Network will complement other collaborative activities rather than duplicate them. It is anticipated participation in collaborations associated with the informal CSTAT workgroup will not only among other benefits help progress the Statistical Network, but, each NSIs goals and aspects of its own capability.

b. Statistics Open Standards (SOS) group

37. Statistics Open Standards (SOS) Group is a consortium of National Statistical Institutes who have agreed that they wish to and are able to contribute to common development of their statistical production environment. It is a prerequisite for this model, that the members constitute a homogeneous group in relation to use of IT, that their underlying production models are similar, and that they share a set of common visions. The members are: Denmark, Finland, Iceland, Netherlands, Norway and Sweden. The Terms of Reference for the SOS Group will be revised annually if needed, and the consortium will continue its work until decided otherwise by the Chief Statisticians. Areas of common interest and cooperation are *dissemination databases, metadata, data collection, architecture and tools, standards.* Although the member institutes serve under quite equal conditions and produce similar statistics, harmonized according to EU regulations, the production processes are implemented differently and the IT architectures are in no way harmonized which makes the implementation of software components for a specific purpose is rather cumbersome and expensive. For more details see [2].

c. User Groups

38. There are several other international statistical groups organized around a particular software product:
    - PC-Axis Reference Group
    - OECD.Stat User Group
    - International Blaise Users Group

39. The key objectives of these groups are of course promoting the implementation and use of the software family in national statistical offices and other organizations and to serve as forums for discussion and exchange of ideas and experiences.

IV. SOCIO-TECHNICAL ISSUES OF COLLABORATIVE SOFTWARE DEVELOPMENT

40. The advantages of the collaboration in software development (the so called *global software development*, see [7]) are recognized by many organizations including the National and International statistical offices. However there are issues, social as well as technical, that prevent organizations from successful collaboration in the anticipated inter-cultural environment.

41. There are four main practices which usually are associated with collaboration (see [7]):
    1) Identify common goals, objectives and rewards
    2) Collaboratively establish and maintain the product ownership boundaries among the interfering teams/organizations
    3) Collaboratively establish and maintain interfaces and processes among the interfering teams/organizations for exchange of inputs, outputs and work products
    4) Collaboratively develop, communicate and distribute among the interfering teams/organizations the commitment lists and work plans.

42. Each of these practices requires effective communication among the teams in order to establish a common understanding of the project and the relevant issues. At the same time each of these practices depends on mutual trust, mutual agreement and consensus rather than authority. Many barriers to these practices have been identified, the main of which are: geographic distance, temporal distance (locations in different time zones), language and culture differences (including corporate culture), infrastructure and product architecture.

43. In the current era of globalization, solutions to these barriers are emerging, addressing the language and cultural differences, social communications, communication infrastructure and distributed development processes. In this context *virtual teams* and *global teaming model* were introduced, extensively studied and successfully applied in the practical software development (see [7, 8, 9 and 10]. Virtual teams have the same goals and objectives as traditional teams and in the same way they interact through interdependent tasks. The essential difference is that virtual teams operate across geographic, temporal, and organizational boundaries and the environment could be multicultural and multilingual [14]. The infrastructure is essential for communication between virtual team members which is normally electronic and often asynchronous. The Global Team Model was particularly developed for managing virtual teams and is a model that represents the key practices involved in software development. Organizations should consider this model when operating in a geographically distributed environment. The GTM is a descriptive process model, so to implement the practices would require organizations to tailor them to their own specific needs.

44. The software development infrastructure and the relevant tools are essential to collaboration among the team members, enabling the facilitation, automation, and control of the entire development process. Especially in global software engineering when distance aggravates coordination and control problems through its negative effects on communication the adequate software tools and technologies will be of crucial importance. A useful overview of such tools can be found in [15].

V. SUMMARY

45. The presented study provides analysis of collaboration activities for sharing of software among national and international statistical organizations through a review and evaluation of the currently known use cases. The focus is on several main issues and opportunities, like licensing, position in the statistical business process model, ownership and governance, sustainability, standards utilization, methods for distributed software development and technical communication advances. From the product perspective three main approaches were identified: a) software developed by one organization (either for own needs or for sharing) and released as Open Source or Free Software; b) software developed by one organization and offered for collaborative sharing for a fee and c) "true" open source software which was developed as OSS and follows the OSS models.

46. The collaboration adds new dimensions of complexity in the technical, administrative and legal aspects and imposes more strict requirements on the analysis, specification and design of statistical software. However, based on the considered experiences of sharing statistical software between organizations we believe that it will substantially contribute to the harmonization of the software architectures, to the adoption of international standards like SDMX and will be beneficial for reducing the costs for developing and maintaining of statistical software, for user support and capacity building.

47. The study is considered a work in progress which will be further extended to cover the experience of emerging forms of collaboration between national and international statistical organization as well as the open source communities.

VI. ACKNOWLEDGEMENTS

48. We would like to thank to all people that contributed to this study, especially the members of the *Sharing Advisory Board (SAB)* for provided information, comments and suggestions.

VII. REFERENCES

[1] Vaccari, Carlo (2009) COmmon Reference Architecture (CORA) ESSnet: history and next activities, paper presented at the Meeting on the Management of Statistical Information Systems (MSIS 2009), Oslo, Norway, 18-20 May 2009, URL: http://www.unece.org/stats/documents/2009.05.msis.htm

[2] Rune Gløersen (2008) The SOS Group, paper presented at the Meeting on the Management of Statistical Information Systems (MSIS 2008), Luxembourg, 7-9 April, 2008, URL: http://www.unece.org/stats/documents/2008.04.msis.htm

[3] Value Creation Group (2010). Informal CSTAT Workgroup on Stronger Collaboration on Statistical Information Management Systems – Draft Output Report. International Conference, Paris, June 2010.

[4] Achim Zeileis (2005). CRAN task views. ***R*** *News*, 5(1):39-40.

[5] United Nations Economic Commission for Europe (2009). Generic Statistical Business Process Model, Version 4.0, URL: www.unece.org/stats/gsbpm.

[6] R Development Core Team (2009). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL: http: //www.R-project.org/.

[7] J. Noll, S. Beecham and I. Richardson (2010). Global software development and collaboration: barriers and solutions, *ACM Inroads* **1** 66-78.

[8] S. Theußl and A. Zeileis (2009). Collaborative Software Development Using R-Forge, *The R Journal* **1** (1) 9-14.

[9] S. Beecham, J. Noll, I. Richardson, N. Ali, (2010). Crafting a Global Teaming Model for Architectural Knowledge Management, In: *5th IEEE International Conference on Global Software Engineering (ICGSE) 2010*, 55-63, URL: http://ulir.ul.ie/handle/10344/705.

[10] V. Todorov (2009). Virtual Teams: Practical Guide to Wikis and other Collaboration Tools, UNIDO Working Paper, URL: http://www.unido.org/fileadmin/user_media/Publications/RSF_DPR/WP102009_Ebook.pdf

[11] V. Todorov, M. Templ and P. Flizmoser (2011) Software for multivariate outlier detection in survey data, paper presented at the Work Session on Statistical Data Editing (Ljubljana, Slovenia, 9-11 May 2011), URL: http://www.unece.org/stats/documents/2011.05.sde.htm
[12] K. Doherty, Building bilingual applications at StatCan (2011). Software Sharing Newsletter, 4, URL: http://www1.unece.org/stat/platform/display/msis/SAB+Newsletter.
[13] EUROSTAT (2010). CORA: Models for licensing, support and legal properties, Work package 4.1 URL: http://cora.forge.osor.eu/WP4__Design_of_the_Organizational_Architecture.htm.
[14] J. Lipnack and J. Stamps (1997), *Virtual Teams: Reaching Across Space, Time, and Organizations With Technology*, John Wiley & Sons.
[15] F. Lanubile, C. Ebert, R Prikladnicki and A. Vizcaino (2010), Collaboration Tools for Global Software Engineering, *IEEE Software* **27** (2), 52-55.
[16] T. Feltcher (2010) Sharing Statistical Software – an update on the OECD experience, paper presented at the Meeting on the Management of Statistical Information Systems (MSIS 2010), Daejeon, Korea, 26-29 April 2010, URL: http://www.unece.org/stats/documents/2010.04.msis.htm.