Eurostat: Standards and open source software for data interoperability

Eurostat collects and publishes huge amounts of data each year, and exchanges many datasets with other large organisations. This exchange was constantly suffering from a lack of interoperability, as data needed to be converted from one organisation's convention into another, a process which consumes both time and money. Different organisations were also using very different tools to work with the data, which caused further problems. In late 2001, Eurostat got together with a number of EU committees to discuss the need for greater interoperability within the European public sector. In 2005, IDABC agreed to fund the SDMX Open DATA Interchange (SODI) project. Thanks to previous cooperations between Eurostat and other international institutions, the SDMX standard quickly found a large group of sponsors, all of which hoped to benefit from the greater interoperability afforded by using a single standard, and the tools built on it. These tools were developed by Eurostat and other sponsoring institutions, and many of them were published under the EUPL license. The SDMX Converter is an example of the successful development and publication of a tool that is essential for working with the SDMX standard.

Quick facts	
Name	X-DIS: SDMX Open Data Interchange (SODI)
Sector	eGovernment
Start date	2005
End date	2009
Objectives	Improve quality, timeliness and interoperability of statistical data
Target group	Mainly public sector
Scope	International
Budget	¬ 3,750,000 over four years
Funding	EU (IDABC)
Achievements	Successful introduction of a standard, development and publication of related software

Table of Contents

Introduction

In 2001, Eurostat's section for Advanced Technologies for the European Statistical System in Luxembourg sat down with various committees composed of the European Member States to brainstorm about the possibilities of developing a European knowledge management system. Although this focus changed quickly, one core issue remained: Europe's need for more interoperability in the information sector. About four years later, these brainstorming sessions were followed by an intensive planning phase which eventually led to the X-DIS program (XML for DATA Interoperability in Statistics). The focus here was no longer on knowledge management, but rather on interoperability of statistics within Europe, and on the usability of these information beyond the organisation which initially gathered it. This was to be achieved by introducing the a standard for all statistical data and related applications, which came to be called SDMX (Statistical Data and Metadata Exchange). In the framework of X-DIS program, the SDMX Open DATA Interchange (SODI) project became an important pillar. Bengt-Åke Lindblad, who is the project officer for SODI explains: SODI is a program to implement data sharing in the European Statistical System using the SDMX standard .

Leonhard Maqua, the program manager responsible for X-DIS further explains: The impulse [for SDMX] came together with the Member States. We have several working groups with the Member States, the most important being the Statistics Telematic Networks and EDI (STNE) workgroup . These meetings proved to be an important framework for the planning of the X-DIS programme and all its sub-projects. Many tools and standards had to be developed, most of which are published as open source software under the EUPL license. From the start, the open source route has been an important strategy to ensure the project's sustainability in the future. Today SDMX is an acknowledged standard by the International Standards Organization (ISO), and many large organization and institutions sponsor the initiative.

Organisation and political background

Eurostat is the statistical institute of the European Commission, producing and illustrating data of a wide range of indicators, from economic ones to social, for example. Within the framework of the IDABC program (Interoperable Delivery of European eGovernment Services to the

Administrations, Businesses and Citizens), Eurostat is in charge of several programs that aim at fostering European eGovernment services. One such program is the X-DIS project, a   project of common interest [1].

Besides Eurostat, the SDMX standard is sponsored by the Bank of International Settlements (BIS), the European Central Bank (ECB), the International Monetary Fund (IMF), the Organization for Economic Co-operation and Development (OECD), the Statistics Division of the United Nations, the World Bank. Maqua explains that the contact between the various institutions has already existed for quite some time, as they had been exchanging data already in previous projects. It therefore only took a a few meeting at the director level to agree on a common initiative for SDMX. The institutions all benefit from the standard, and work together on increasing its use, as well as on improving the applications built on it. The costs for tools developed as part of the SODI project are shared with all institutions, which keeps the financial burden for each organisation at a minimum. The data collected by these institutions can be used by all of them, which facilitates the data collection and its use greatly.

Illustration 1: Organizational context of SODI

The SODI project receives funding as part of the X-DIS program, which in turn is financed by the IDABC. The project is set to end in late 2009, and the Eurostat team aims to conclude all the contractual work by then or shortly after. In order to ensure sustainability of the project and the software it has developed, the standard and related applications have been published under the open source license EUPL. The Food and Agriculture Organization of the United Nations (FAO) is just one example of a non-sponsoring institution making use of the SDMX standard and the related open source tools. In the future, besides the already large user groups at Eurostat and the other sponsoring institutions, Eurostat is hoping to further extend the use of the standard to the EU Member State's institutions, and other departments within the EU. In addition, users from the private sector and other institutions will also ensure the standard's use in the future. Already at this point, there are many users that employ the SDMX standard on a daily basis,

---

1   A project of common interest under the IDABC programme is a project in one of a number of policy areas listed in Decision 2004/387/EC of the European Parliament and the Council. One of those policy areas is interinstitutional information exchange.

and further develop the related tools to their needs.

Budget and Funding

Before receiving funding from IDABC, the project had to be carefully planned, which took about 3 years from late 2001 to 2005. IDABC saw the potential of the SDMX standard and agreed to fund the project from 2005 until 2009. In total, the funding for X-DIS is set at ¬ 3,750,000 over those four years. As Maqua explains, the budget is used for all aspects of the program, including the development of applications and the support to the Member States. As an example, he mentions the SDMX Converter, which has a budget of probably ¬ 100.000 to 200.000 . It is however difficult to give clear estimations, as there are not dedicated budgets for individual aspects of the project

The program is very beneficial for Eurostat and the other sponsoring institutions. As the SDMX standard is applicable throughout all institutions, it offers the possibility to improve the interoperability across the institutions, while at the same time speeding up internal processes. It is for this reason that each of them is contributing in some way to the enhancement of the SDMX standard and the related applications. The division of tasks and budgets is decided at high-level meetings of the various institutions, and attempts to give a fair representation of who benefits the most from each development. Moreover, the developments are often tested together, in order to ensure the best possible outcome.

Technical issues

For the X-DIS program, the implementation and the development of the SDMX standard are the central points. The main objectives of the project are:

- to improve quality and timeliness of statistics
- to decrease the reporting burden of enterprises and statistical authorities in Member States
- to improve accessibility of statistics for business users and citizens.

The main technical backbone of SDMX is XML, which stands for Extensive Markup Language. It is a framework in which markup languages are developed, which allows for the for visualization of data in text form. With regard to statistical data, SDMX makes it possible to illustrate statistical data with tables and graphs, which can be updated easily and quickly. Through the use of the standardized SDMX format, the different institutions can make use of data more efficiently, and share their datasets much more easily, as they do not have to convert data into their own format a process which is usually time-consuming and error-prone. Especially in an international context, this substantially facilitates the work, as data has to be collected just once, and can then be shared and used by all. The standard's certification by the International Standards Organisation as ISO

17369 means that the standard meets official criteria and will be applicable in the future as well.

There are seven key actions within the X-DIS project:

- Implement SDMX (Statistical Data and Metadata eXchange, see www.sdmx.org) standards and develop appropriate tools

- Develop an SDMX Open Data Interchange (SODI, interoperability of statistical web sites for economic indicators)

- Set up sectoral networks for exchange of statistical information (initial focus on XBRL - eXtensible Business Reporting Language, which supports information modelling and the expression of semantic meaning commonly required in business reporting)

- Develop and implement advanced visualisation techniques for statistical data in XML format

- Make Eurostat s web site interoperable

- Develop suitable tools to access large statistical XML datasets

- Create a repository of open source software for statistical purposes, using the IDABC OSS repository to collect all tools developed in X-DIS and grant sustainability and re-usage of results

Especially the repository named in the last point is important to ensure the use of the SDMX standard and the related tools after the project's end. Achieving this sort of sustainability was a major consideration in the design of the X-DIS project.. Maqua explains that the tools are usable with your own data, or with the data we provide on our website . The tools are usually more interesting for larger bodies than for individuals, although they may also be of use in companies. As an example Maqua refers to the Eurostat SDMX converter, which converts statistical data between different version of SDMX. It is an important tool, which is essential for anyone working with SDMX , he states. This may also explain its popularity in terms of SDMX-related downloads at the OSOR.eu platform and repository. [The converter] has been downloaded from the OSOR platform alone over 300 times within four month , Maqua further adds.

In addition to OSOR, the SDMX standard is also available on the

Leonhard Maqua. Head of Section Standardisation and advanced information technologies for statistics

SEMIC.eu platform (Semantic Interoperability Centre Europe). Here also, a frequently updated release of the standard is available for download. By joining the forum on the website, one can further engage in discussion about the standard and related tools, which proves to be an important input for the project. The presence of the standard on multiple websites is therefore very important for the project, as this not only ensures more users, but a broader group of user experiences which can be shared and discussed.

The SDMX converter serves as an example of the typical way in which a tool is developed in the X-DIS project. For the development of this tool, Eurostat contracted a third party consortium comprised of a Hungarian, Greek, and Luxembourgian development team. Eurostat clearly specified its requirements, and closely monitored the development process, Maqua says. Lindblad adds that   the software is developed with Java technology, it requires the Java Runtime Environment and it is platform independent   .This was very important, seeing that the software had to function on many platforms.

Legal issues

In order to receive the IDABC's approval and ultimately funding for the X-DIS program and the SODI project within this, Eurostat had to bring forward justifications for the need of such a program. Clear explanations of the aims of the project, and the existing shortcomings at the time where therefore necessary to get the IDABC's approval. Eventually however, the program targets were clearly formed, and it was clearly in the IDABC's interest to approve and fund the program.

Publishing applications as open source software does not pose any problems for Eurostat. When Eurostat contracts third parties to develop software, it makes sure to obtain the copyright and related rights in the product. As the funding for the development came from IDABC, the European Commission is ultimately is the owner of the software and can decide what to do with it. Once a tool such as the SDMX converter is fully functioning and ready to use, the Eurostat team may decide to publish it under EUPL license, if it is considered useful for the public and other institutions.

Change management

For the introduction of the SODI project at Eurostat, the team gradually implemented the standard and the related tools and applications, says Lindblad.

The X-DIS team took a step-by-step approach to introducing the use of the SDMX Open Data Interchange (SODI). This allowed Eurostat to adopt its own procedures. Ensuring a smooth

transition was also necessary since Eurostat works closely with the Member States in collecting data, so that a significant number of organisations were affected by the introduction of the new standard and the related tools. Fortunately, this process went smoothly, and the team did not encounter significant problems.

Specialised tools such as the SDMX converter and other software developed by the X-DIS project usually do not require significant changes in the IT infrastructure of the user organisations. This is particularly the case for applications built on platform-independent technologies such as Java.

Effect on government services

The use of the SDMX standard has greatly accelerated the process of publishing statistical data provided by Member States. Where this process was previously hindered by the use of diverging standards, SDMX allows for a much quicker publication, since Eurostat no longer has to invest time in converting the data of the Member States from various formats into its own mode of display.

In addition, the use of a common standard amongst various international institutions is very beneficial. It makes significant savings possible, as it becomes very easy for an organisation using the standard to use data which has already been gathered elsewhere for its own purposes, with minimal effort. The data is collected only once, and then shared between the various organisations, which reduces the amount of funding needed for collecting data. The process of sharing is not only enabled, but also accelerated by the use of SDMX as a common standard.

Cooperation with other public bodies

SDMX is an initiative that is sponsored by the Bank of International Settlements (BIS), the European Central Bank (ECB), the International Monetary Fund (IMF), the Organisation for Economic Co-operation and Development (OECD), the United Nations (UN), the World Bank and Eurostat. The cooperation between these organisations originated before the SODI project, as they were had already been exchanging data and expertise. The common initiative came into existence after a series of executive meetings, which were further prepared in technical staff meetings in Paris, Maqua explains. Eurostat is the main beneficiary of the greater interoperability provided by the SDMX initiative, because in contrast to the other institutions we receive data*and* pass it on, which is basically only the case with the ECB although it has a much narrower target area, Maqua explains. Nonetheless, the initiative also brings many benefits to the other institutions as well. As Maqua says, we do something, they do something, and afterwards we can all use it together, because SDMX is a common standard.

In the case of the SDMX converter, Eurostat was in charge of the development of the tool. It contracted a consortium of firms, which developed the converter in close cooperation with the Eurostat team. For testing purposes, the team worked closely together with the Bank of International Settlements (BIS) in order to test the software thoroughly and in areas where Eurostat could not have put the program through its paces by itself. As Maqua explains: Especially in the beginning we didn't have too many data sets to work with, so it came in very helpful to use their input   .

Another example can be seen in the European Central Bank's Statistics Dashboard tool, which visualises statistical data in a way that is very clear and easily understandable. Although this tool was not developed by Eurostat, it   will probably be used intensively by us   , says Maqua. This in turn will help in further improving the software. Cross-institutional developments are thus not considered an obstacle for the initiative, but rather help all participating institutions in some form or another.

Beyond the group of sponsoring organisations, the United Nation's Food and Agriculture Organisation has also started using the standard   something which Maqua says he only learned about when a representative of the body told him so at a a conference.

Evaluation

Achievements / Lessons learned

For the success of the X-DIS program and the SDMX initiative, careful planning was one of the most important ingredients. Maqua highlights that   we invested one and a half work years into the planning phase of the project   , which enabled the project team to give clear predictions of the requirements for such a program. With a certain pride in his voice, Maqua adds: We basically managed to stick completely to the project plan   .

With basically no advertisement and only with the help of the Open Source community and the widespread use across all sponsoring institutions has the SDMX standard achieved relatively large popularity.   Many opportunities even beyond the purely statistical area have been created through the SDMX standard and the the tools that come with it   , says Maqua.

Conclusion

Compared to many other tools in the Open Source ecosystem, the SDMX standard and tools such as the SDMX Converter are backed by a number of very large institutions. Although one may assume

that cooperation on such a high level could be problematic, it appears that the SDMX initiative and the SODI project work very efficiently. Through the SDMX standard, all tools can be employed by all institutions, which saves a lot of financial resources, and further ensures a fairly large user group. This is in turn helps in developing adaptable and well functioning software, as it can be tested in many situations and contexts.

Use of the standard and the related tools is spreading beyond the group of sponsoring organisations. Assuming that this process continues, it will further add to the combined weight of the large institutions already backing the standard and the related tools. This demonstrates the power of free software licenses like the EUPL to enable sharing between a large number of partners without the need for significant contractual overheads.

With regard to the future, the Eurostat team's approach to publish the software under the EUPL license appears very successful. As the software is already used extensively by a large group of users even beyond the sponsoring institutions, early since of success are already visible before the funding of the project runs out.

Links

-[SDMX.org](#)

-[Eurostat SMDX Converter on OSOR.eu](#)

-[SDMX Content-Oriented Guidelines](#)

-[X-DIS - XML for Data Interoperability in Statistics at IDABC](#)

-[SDMX Open Data Interchange (SODI)](#)

-[The SODI Pilot Project](#)

-[SODI Demonstration of the results of the SODI Pilots](#)

-[SEMIC.eu platform](#)

Author: Gregor Bierhals, UNU-MERIT