

Distr.
GENERAL

CES/AC.71/2005/28
21 March 2005

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)**

**ORGANIZATION FOR ECONOMIC
COOPERATION AND DEVELOPMENT (OECD)
STATISTICS DIRECTORATE**

Joint ECE/Eurostat/OECD Meeting on the Management of Statistical Information Systems (MSIS)
(Bratislava, Slovakia, 18-20 April 2005)

Topic (iii): XML and web services

**AN APPLICATION OF SDMX-ML FOR THE JOINT DISSEMINATION OF NATIONAL
CONTRIBUTIONS TO EURO AREA AGGREGATES ON THE WEB SITES OF THE EUROPEAN
CENTRAL BANK AND NATIONAL CENTRAL BANKS**

Supporting Paper

Submitted by the European Central Bank, Germany¹

Keywords

Publication of statistics, time series, Internet, XML, XSL transformation, SDMX

Summary

We present a framework for the distributed publication of statistics. The European Central Bank provides time series on its web site in SDMX-ML syntax. The central banks in the European System of Central Banks then integrate customised tables populated with these data into their web sites.

I. INTRODUCTION

1. Since the European Central Bank (ECB) has first started to publish euro area aggregate statistics in late 1998², several groups of users have voiced their demand for an easier access to the national contributions behind these aggregates. Among those users have been journalists, researchers, market participants, and also some committees of the European System of Central Banks (ESCB), which comprises the ECB and the national central banks (NCBs) of all European Union member states.

2. Up to now, national data have been published only by the responsible national institutions. This situation puts a considerable burden on the consumer of statistics who tries to compare data across countries. In fact, locating and accessing the relevant tables turns out to be quite difficult due to the differences in

¹ Prepared by B. Bodenstorfer and G. Salou (bernhard.bodenstorfer@ecb.int; gerard.salou@ecb.int).

² P. Bull: "The Development of Statistics for Economic and Monetary Union", ECB publication, 2004, http://www.ecb.int/pub/redirect/pub_5874_en.html.

classifications, details, metadata, languages, presentation, download options, and data formats applied on the various web sites.

3. In response to this situation, the Statistics Committee of the ESCB has mandated a newly created task force to improving the accessibility of euro area statistics in late 2003. Since then, the task force has formulated a solution and evaluated it in a pilot project. After approval by the Governing Council of the ECB on 3 March 2005, the technical implementation of the proposal is under way.

4. The purpose of this paper is to present the techniques behind the adopted solution. The web sites of the central banks in the ESCB will include tables showing euro area statistics next to the relevant national contributions. While the numeric content of these tables will be synchronised, the presentation can significantly differ between the central banks. It is up to each of them to decide on a layout which optimally integrates the tables into its web site.

5. The remainder of this report is organized as follows: Section II introduces the technical standards used and sketches the basic idea of the proposed technical framework. It also highlights important functional features. Section III reports on the issues dealt with in the pilot project and hints on ways to respond to them. Section IV concludes this paper and briefly draws attention to some topics outside the technical scope.

II. TECHNICAL PRINCIPLES

A. Standards Used

6. The proposed framework is based on XML and web technologies. In particular, it relies on the standards adopted by the SDMX³ initiative on 30 September 2004: The SDMX-Markup Language (SDMX-ML) is the XML standard to represent multi-dimensional statistical data, in particular, arrays of time series and associated metadata. Among the SDMX-ML data formats, the so-called compact data format has been chosen in order to obtain maximal performance.

7. The second important standard applied is the extensible stylesheet language (XSL) defined by the W3C⁴. XSL can express transformations from XML to other formats; in particular, to different types of XML and to HTML.

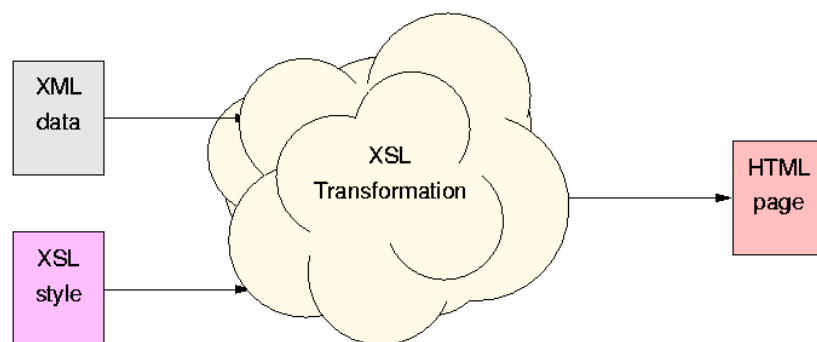


Figure 1: XSL Transformation of XML into HTML.

8. In this project, XSL is used to transform SDMX-ML data into HTML. XSL transformation facilities are nowadays readily integrated in the programming languages that are familiar for web development, for example in PHP, JSP, and ASP. This makes it easy to merge dynamic content created from XML into web pages.

³ Statistical Data and Metadata Exchange: <http://www.sdmx.org>.

⁴ World Wide Web Consortium: <http://www.w3.org>.

B. Principles of the framework

9. Out of its statistical data base, the ECB produces a set of data files in SDMX-ML compact data format. The files contain time series for euro area aggregates and national contributions to those (ESCB data) together with relevant metadata. Then the ECB makes the files centrally available on its web site.

10. Each central bank defines an XSL stylesheet containing parameters for the presentation of the ESCB data on its web site. It then creates dynamic web pages by means of widespread technology to show the data using the stylesheet.

11. When activated, the code of the web page retrieves the necessary XML files from the ECB, performs the XSL transformation based on the stylesheet, and presents the resulting content on the generated page. Thus, each NCB can apply its own look-and-feel and presentation criteria to show the ESCB data. Like all NCBs, the ECB uses XSL, too, to integrate ESCB data into its web site (not shown in Figure 2).

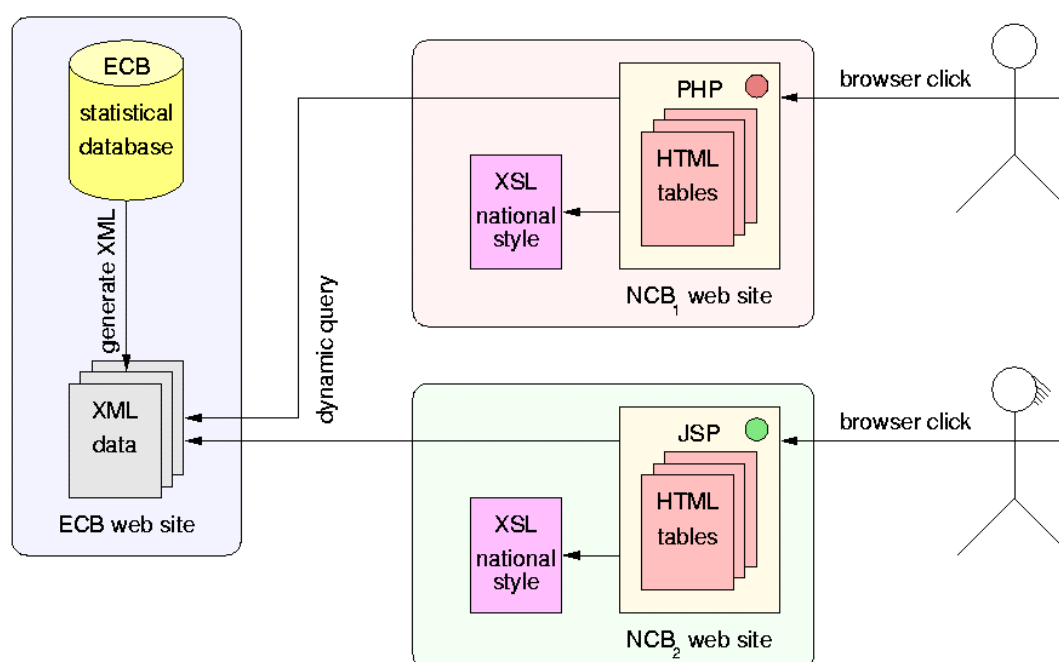


Figure 2: Upon request, two NCBs present the same data in an individually chosen style.

C. Functional features

12. ESCB statistics are shown on the web site of each central bank, so users do not have to exit the web site to access them.

13. ESCB statistics appear as a co-ordinated network. Inconsistencies in data will not occur because there is only one data source. Still, the original ownership of all data is not undermined because the appropriate information is packed into the SDMX-ML data file.

14. ESCB statistics are integrated into each web site with the appropriate look-and-feel. Each central bank can abide by the style of its web site. For example, the XSL stylesheets can present the tables in the relevant national languages.

15. The use of widespread XML technology is flexible and extremely cost-efficient. Of course, NCBs may still choose to link to the web pages of the ECB for ESCB statistics.

16. Relying on SDMX-ML leverages the statistical data model commonly understood within the ESCB, which is based on GESMES/TS⁵ key families⁶.

III. PILOT PROJECT

17. A pilot project was conducted to validate and elaborate the proposed framework. Several NCBs participated together with the ECB. The results were generally positive. Nevertheless, some issues were identified for further attention:

A. Performance

18. The proposed solution involves a query to the ECB web site. This implies network overhead. Additionally, the XSL transformation uses processing resources and takes its time. Performance appeared quite acceptable for XML files containing up to 50 series, which usually is enough for a table, but it gradually deteriorates for larger volumes. Of course, available XSL software has been markedly improving and hardware will also progress. Still, file size and performance must be further monitored during the implementation phase.

19. Caching the HTML pages between page hits can provide a remedy: The first time, a page is requested, it is computed from XML. The result is stored for later requests. These then do not have to run the XSL transformation until the XML data changes. To detect such changes, a special request could be sent to the ECB's web server, namely an HTTP⁷ standard "HEAD" request. It avoids the retrieval of the whole data file unless the file has changed.

20. Network traffic for downloading the XML file could in future be reduced using a compressing HTTP content encoding such as gzip. However, since file sizes are moderate so far, this was not tested.

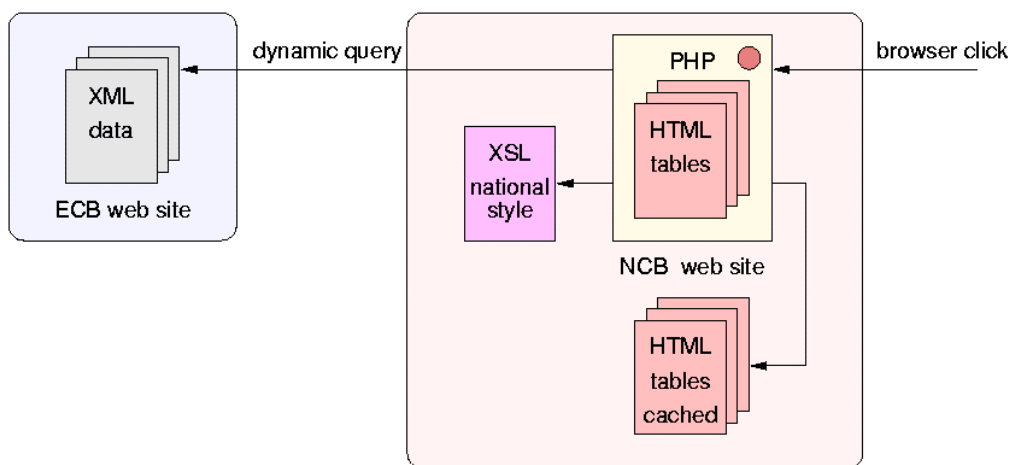


Figure 3: Caching HTML pages increases performance and availability.

B. Security

21. Since the XML file from the ECB enters the NCB via the Internet, it is important that the transfer uses a secure protocol. The data is anyway intended for the public, so encryption is not required. But authentication

⁵ The SDMX standards are backward compatible with GESMES/TS, the generic statistical message for time series. GESMES/TS, formerly called GESMES/CB, is based on the UN/EDIFACT standard GESMES, and is used throughout the ESCB for the exchange of statistical data and metadata.

⁶ Key families assign the dimensions and attributes to identify and qualify the data in a multi-dimensional array. Statisticians can define key families tailored according to the requirements of their subject and draw on them to set up the intended dataflows within the ESCB at minimal cost.

⁷ IETF: "Hypertext Transfer Protocol -- HTTP/1.1", <http://www.ietf.org/rfc/rfc2616.txt>.

is essential in order to make sure that the XML data is really coming from the ECB. Therefore, the standard HTTPS⁸ protocol was tested during the pilot. This protocol is frequently applied in online banking portals, for instance. The technical costs incurred for HTTPS are low, since it is a standard protocol. There is some cost, though, in terms of organizational overhead because HTTPS requires a certain discipline for key management.

22. In very practical terms, NCBs may need to customise their firewall configuration so that their web servers can retrieve the ESCB data from the ECB.

23. As another issue, contemporary XSL software generates noticeable server load and thus raises the proneness of the web sites to denial-of-service attacks: Too frequent request of the dynamic pages may eventually degrade the server's overall responsiveness. Caching HTML tables may be the best way to minimise this risk.

C. Availability

24. Since the ESCB statistics pages of the NCBs rely on the online retrieval of data from the ECB, problems there or in the network could affect the availability of the tables on the web sites of the NCBs. While a service level agreement could help to deal with the risk that the ECB fails to provide the data, the risk of interconnection problems is harder to defeat. Again, caching HTML tables promises a solution. If access to the XML files fails, the cached HTML pages could be exhibited anyway.

D. No dynamic web pages

25. Some external web sites do not foresee dynamic web pages. The concerned central banks may have to rely on static HTML pages that are updated in a timely manner. Generation of such pages could be triggered by regularly polling the XML files. As above, the HTTP "HEAD" request can minimise network traffic for the iterative queries.

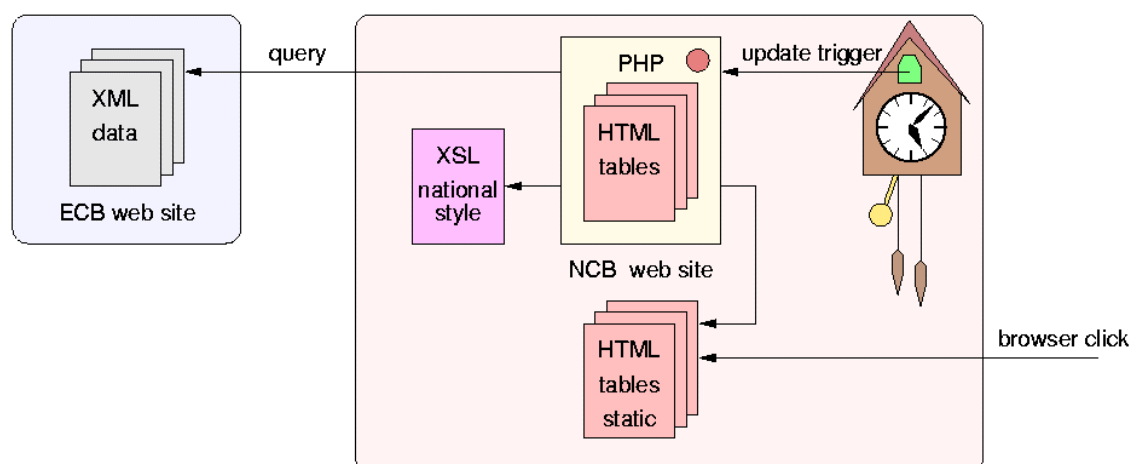


Figure 4: Regular polling and calculation of static HTML pages.

26. Instead of polling each XML file on the web site of the ECB, a single additional file could be provided there to contain a list of recently updated XML files. Then only this single file would have to be periodically checked.

27. To avoid polling completely, the ECB could itself send to the NCBs a notification upon a change in the XML files. Also the strict adherence to a release calendar would eliminate the need for polling. However, all

⁸ HTTP via a transport layer security protocol. See IETF "The TLS Protocol", <http://www.ietf.org/rfc/rfc2246.txt>. Also compare IETF "The Secure HyperText Transfer Protocol", <http://www.ietf.org/rfc/rfc2660.txt>.

these workarounds would considerably raise the overall system complexity and thus the risk to occasionally lose synchronisation of the table content with the XML file.

E. Natural language translation

28. Most metadata in the ECB's statistical database comes in the form of codes. National XSL stylesheets can rely on the relevant code lists and confine themselves to translating the codes. However, some free-text metadata is available only in English language. XSL can still contain translation tables, but the maintenance of those would require a continuous effort.

F. Adjustments of metadata

29. A limited set of metadata related to publication and confidentiality need to be revised before the data can be put on the web site in the form of SDMX-ML. And some metadata are not at all available in the ECB statistical data bank, for example, the national release calendars and contact addresses to which consumers of statistics should refer concerning the data under the responsibility of the NCBs.

IV. CONCLUSION

A. Summary

30. This paper has described the framework for the joint publication of national contributions to euro area aggregates from the technical point of view. It has used this example to highlight the opportunities offered by the SDMX standards in a realistic application.

B. Outside the scope

31. On the other hand, issues which are only marginally or not at all related to the technology have been largely left out. As examples should be mentioned the difficulties to choose the right scope of data and the need to harmonize the concepts behind them before numbers from different countries can be published next to each other in a meaningful manner. Also new questions may arise as a result of cross-comparisons by consumers of statistics. Some coordination may be required to respond to difficult questions that do not clearly fall in the domain of a single central bank. And at a later stage, coordinated pages should be established to answer frequently asked questions.

32. The paper has left very abstract the criteria for a suitable presentation of ESCB statistics: The XSL stylesheet should just do it. To improve the accessibility of the national contributions, data must be shown clear and simple. It has been suggested that a minimum set of metadata should be identified for this purpose, as appropriate metadata would reduce the need for user support.

C. Acknowledgements

33. The authors would like to thank their colleagues at the NCBs who participated in the pilot and the contributors to the Apache Jakarta⁹ project who provide open source XSL transformation software.

⁹ The Apache Jakarta Project: <http://jakarta.apache.org>.