

Distr.
GENERAL

WP.9
19 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 (i): Architectures, models and standards

Proposal of a Technological Architecture for a National Statistical and Geographical Information System in Mexico

Supporting Paper

Prepared by Juan Muñoz López, National Institute of Statistics and Geography (INEGI), México

I. Introduction

1. Software architecture provides the blueprints for describing the structure of a system, their components and the relations among them [1]. It's a conceptual artifact that will help to align the efforts of specialists and common users that have different points of view and must cooperate to develop a system. We can view system architecture as a communication instrument used to articulate and explain abstract concepts conforming solutions that can be rationalized and improved by a team that must address different concerns.
2. A well-developed architecture description is a valuable material that is used as a key input in the planning process of an IT system project. Since the beginning software architecture will help to detect inconsistencies and to predict the possibilities to accomplish certain goals or the need to trade some characteristics to get others that are of greater priority.
3. This paper talks about a proposal of an architectural model that the National Institute of Statistics and Geography of Mexico (INEGI) are developing to accomplish its obligations as a public institution. To contextualize this, INEGI is an organization of the public sector in Mexico which have two main responsibilities:
 - (a) Coordinate a National Statistical and Geographical Information System
 - (b) Produce statistic and geographic information that is relevant to the nation
4. Development of a National Information System requires of a solid framework to support the different objectives that it must get. As a coordinator of that System, INEGI must provide a solid plan to develop it, including the environment for its technological support.

5. INEGI is in the way to build the IT platform that will support the processes developed by the institutions of the public sector of Mexico which integrate that National Information System. All those efforts must be lined up to reach the main goals of the system, which are: to produce relevant and official statistical and geographical data, disseminate it and preserve it.

6. The design of the architectural model for this IT platform must consider technical constraints that are inherent to very large systems which must incorporate existent applications in a consistent and coherent way, although they weren't created to cooperate between them.

7. The inclusion of standards in this architecture it's necessary to add characteristics that will be perceived as elements that contribute to determine its capacities and quality. Interoperability, modularity, trustiness, homogeneity, friendliness are examples of this kind of characteristics that designers of the architectural model perceive that are improved when using certain standards, like SDMX, DDI, between others.

8. In this paper some details that picture an overall view of the proposed IT architecture for the support of the system are presented along with some advances in its implementation and an envisioned set of tasks that must be faced in the path to implement a solid technological environment.

II. Proposed IT Architecture

A. Context and Brief description of the objectives

9. The law of the "National System of Statistical and Geographical Information (SNIEG)" [2] describes and regulates INEGI; and is derived from a constitutional amendment that gives autonomy to her as an institution which belongs to the Nation and that doesn't have a direct dependency of any government power.

10. This law defines a National System of Statistical and Geographical Information which name is abbreviated as SNIEG. This system it's conformed by administrative units of institutions from the different levels of government in Mexico (federal, state and municipal) at all its powers (executive, legislative, judicial and autonomous) and is coordinated by INEGI. The purpose of SNIEG can be perceived by its main goals, which are:

- (a) Produce statistical and geographical information
- (b) Disseminate the statistical and geographical information using means to make easier to access it
- (c) Promote knowledge and use of the information
- (d) Preserve the information

11. Information produced by this system is official and must be used by all the levels of the government in Mexico. INEGI has the obligation of coordinate the activities of the system, integrating the information produced by it and improving its operation, taking in account best practices and national and international standards.

12. Under this context, INEGI has envisioned that a large technological system must be developed to support the processes of the SNIEG. Analyzing the main purposes of SNIEG, architects of the IT infrastructure have defined several objectives that will drive its design and development. These goals are:

- (a) Support coordination and collaboration among the members of SNIEG
- (b) Facilitate the exchange and interpretation of information in the national and international context
- (c) Help in the production of statistical and geographical information
- (d) Provide means to make easier the localization, integration, access and interpretation of the statistical and geographical information
- (e) Establish a strong scheme to preserve the relevant statistical and geographical information

13. In a real world, a big system like this one cannot be built from the scratch. It has to be built interconnecting existent systems (most of them cannot be modified) with new modules that must be developed giving coherence and additional functionalities to the whole IT environment.

14. Some examples of the systems that have been developed by INEGI and must be incorporated to this IT platform are:

- (a) The statistical Data Warehouse of INEGI, which is the database that integrates information from all statistical projects
- (b) The Bank of Economic Information, a system that contains more than 160,000 economic indicators
- (c) The Digital Map of Mexico, a geographical information system that can incorporate geographical and statistical layers from different institutions
- (d) The Web Mapping Service, which provides with a service that delivers customized maps from Mexico
- (e) The National Registry of Geographic Information
- (f) Other systems and information repositories, like the National Statistical Registry, the National Directory of Economic Units, the Simulator of Watersheds, etc.

15. In the same context, designers of the architecture must think in a project that will take a long time to be developed, even when some of their components will be integrated and in a short time will be replaced, yet before having finished the whole system Many reasons will raise the need to modify or replace the software components of the system, like: adaptation to new needs, change of procedures and methodologies, evolution of technologies, ageing and obsolescence, etc.

16. Members of SNIEG are institutions with non-equally developed IT capacities. Some of them can be at the top in the application of technologies, but others may be exactly the opposite. The design of the system must consider different levels of service requiring different amounts of resources.

17. Those concerns are key restrictions that must be addressed by a modular and low coupled system with a consistent architectural model, and are part of a bigger set of non-functional requirements that also includes other issues like: stability, performance, security, interoperability, etc.

18. The premises that have been described above; are the main drivers for the design of an IT architecture to conform an IT environment for the support of the activities that must be developed by the members of SNIEG to accomplish the obligations that have been mandated by the law.

B. Description of the Architecture

19. According to the standard IEEE1471-2000 [3], a well-documented architectural description must contain enough information to interpret in a correct way all the aspects of the system and including information for the management of the document itself. This is a good practice that only few institutions follow to document its systems.

20. In the core of an architectural description, different views organized by different perspectives called points of view contain the blueprints that focus the interest of the project participants. Some views are used to integrate a global overview of the system or to show a conceptual description of the system; others are used to describe aspects that are of the interest of software developers, aspects about the information structures and its distribution can be defined in other group of views; and aspects related to the environment in which the system will be deployed are included in a different point of view [4].

21. This paper only contains a brief conceptual description of the system architecture of the IT support environment for the SNIEG, but it's necessary to remark that this is according only to the scope of this document and that always is recommended to develop a well-documented architectural description of the systems.

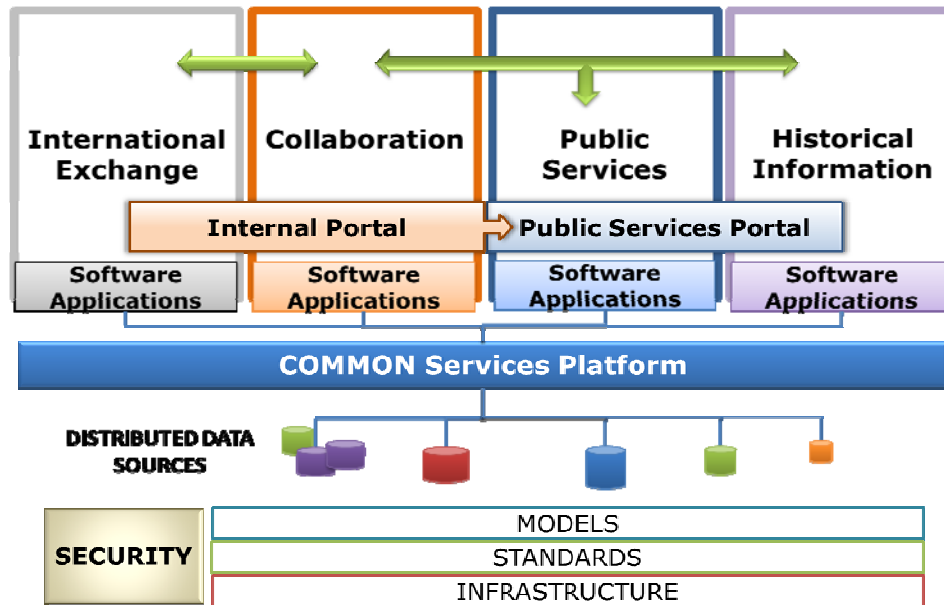


Figure 1; Proposed Conceptual System Architecture of the IT Support Environment for the SNIEG

22. From a very abstract perspective, this conceptual architectural model describes a federated network settled on a common support structure integrated by three stacked layers denominated: infrastructure, standards and models. This support structure also has a fourth layer which deals with security, but this is transversal and interacts with the other three.

23. Infrastructure is the bottom layer, describes the basic elements that must be common to all members in order to make possible to connect their systems or to provide information services. Some of those elements when are mentioned could sound too obvious, but for its relevance it's necessary to be sure that they aren't omitted, per example: any institution that wishes to provide information services in the system must have a permanent and stable connection to Internet with enough bandwidth to support user demands.

24. Standards, is the second layer and it defines the standards that must be adopted by the members in order to harmonize the IT environment. As an example: one of the standards that have been adopted in the system for the exchange of information among the members is SDMX.

25. The layer of models is somewhat more abstract, it describes models, styles and patterns that are followed by the architecture of the whole system or parts of it, and they serve to define its functionalities and the characteristics of its services. By example, a big part of the infrastructure is based on a Service Oriented Architecture model.

26. Security layer defines schemas, rules and practices that all the members of the System must adopt to ensure a solid, stable and trusty technological environment. For each layer, a set of rules must be declared; as an example, one of those rules could be that all the software applications integrated into the network must have a common authentication system.

27. In a logical way, the network can be divided into four modules, each one oriented to support specific needs of SNIEG.
28. *Collaboration Module*: Is the core module of the network, it addresses the coordination and collaboration among all the members of the System. It is an extranet with restricted access and it is planned to incorporate tools to support task like:
- (a) Planning and project management
 - (b) Meeting scheduling and following of agreements
 - (c) Integration and exchange of information
 - (d) Development of common projects to produce statistical and geographical information
 - (e) Collaborative work to develop technical instruments
29. *Public Services Module*: It is oriented to disseminate statistical and geographical information to society, providing different kinds of tools to support a diversity of audiences and needs. It's envisioned to provide access to information in a multimodal way, not only using the traditional website interfaces accessed by users using computers, but providing other means of access to mobile devices like smart phones or tablets; or systems directly connected by web services providing or consuming electronic data and metadata flows.
30. *International Exchange*: Its objective is to bring support for the accomplishment of international agreements related to the exchange of significant statistical and geographical information from Mexican dependencies. It's a mechanism that will help to improve coordination in this topic among national institutions and with international organizations helping to provide a better picture of the reality of the country.
31. *Historical Information*: This module must preserve statistical and geographical information over time. It is a tool to ensure that new generations will have information that they will be able to use to analyze the country's development in different epochs. By its nature, it should aim to keep the information current and accessible, even facing permanent technological changes.
32. The external view of all the modules is composed by different applications that act as an interface for the users to deliver them the services of the system. The Internet pages of the network function as containers for applications that help focus the satisfaction of needs of different types of users, but because of the architecture they all have a common source of information.
33. Control and location of each one the components of the system is an essential concern that must be addressed to maintain it operable and manageable. Implementation of the mechanism of registries provides a practical solution to this issue. The registry service of the system provides the functionalities to maintain a directory similar to those implemented by operating systems, but with some particularities.
34. The Directory is envisioned as a mean to locate, characterize and manage different kinds of resources like: institutions, people, services, software applications, data repositories, digital documents, policies and information. This service will be the core of the IT supporting infrastructure and must be the result of the conjunction of specialized modules.
35. Some examples of rules that will be applied to the directory are: only registered resources are recognized as components of the IT environment that supports the SNIEG; all software components of SNIEG must attend to the authorizations and restrictions registered on it for each resource; all records in the directory that point to inactive resources are subject to a deputation process.

C. Role of Standards in the Architecture

36. As it can be seen, the architecture of the system for the support of SNIEG describes an IT environment that works as an integration framework of different information sources. Statistical and geographical information are the blood of the system and must reach each of its components in a fluid way.

37. Unfortunately for the purposes of the system, information has been produced, processed and stored by institutions in many different ways and using diverse formats. The systems of those institutions have been developed to solve just their own needs and they have not been designed to be part of an environment in which they must interoperate. Also, in a lot of cases systems in a single institution cannot communicate directly between them.

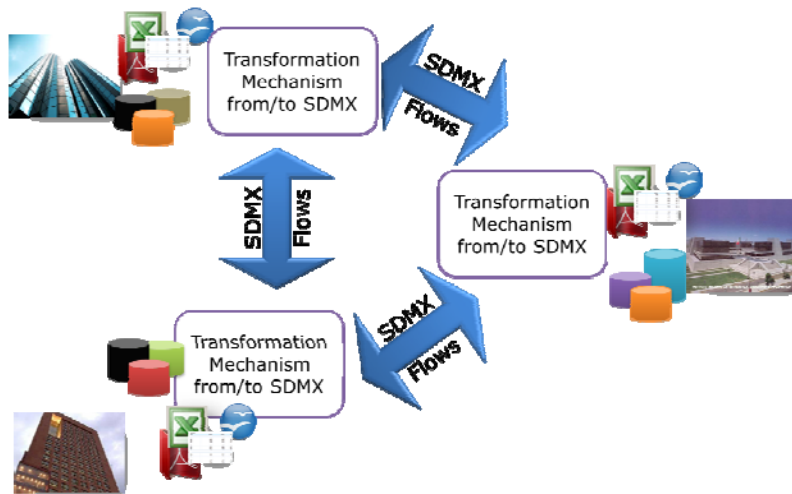


Figure 2; Example of the application of SDMX to exchange information among institutions

38. Standards like SDMX can be seen as a powerful instrument to enable interoperability among systems. Statistical information that is in different formats can be translated to the common language of SDMX and be exchanged with other systems. The transformation of information can be made by a mechanism that will be external to existent systems, so they don't need to be modified.

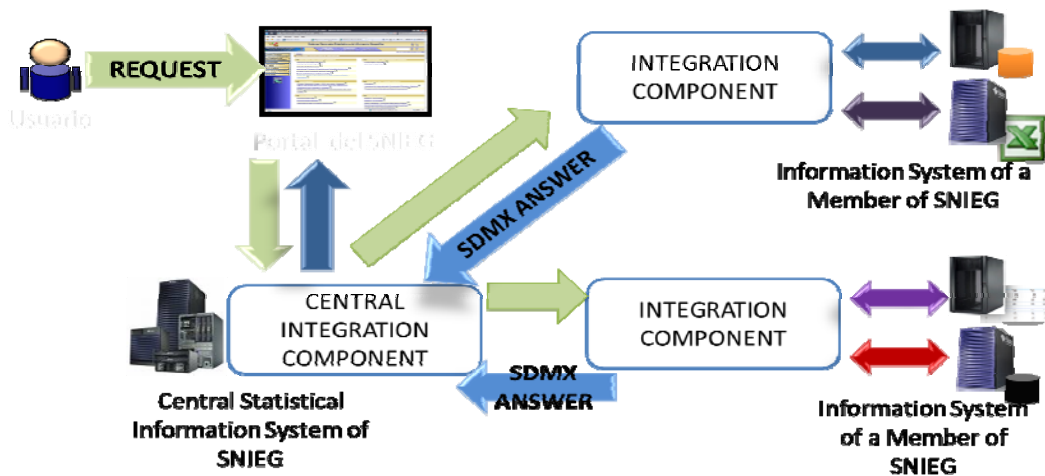


Figure 3: Scenery describing the mechanism to integrate distributed data to answer requests of the user in real time

39. This is a powerful mechanism can be applied in several ways, like: to collect information from administrative registries, to integrate distributed information in real time to answer queries that implies data produced by different institutions, to provide information to general domain statistical applications, etc.

40. Reuse of software is another benefit that comes with the application of standards, like this one. Existent applications which are working in an institution to solve specific problems can be used in other institution to help in similar problems with little effort. When interfaces between applications are based in standards modularity is enhanced and they can be replaced when needed with minimal or no impact to other components.

41. Going more in depth, use of other standards helps to harmonize production of information that is in charge of each member of SNIEG and make the outputs trustier to users. As an example, DDI has been adopted by INEGI and it's being used to document the stages of statistical projects, so users can review how information was produced and determine in which grade they want to trust in their results.

42. A technical norm issued by INEGI describes activities that must be considered to develop a statistical survey project in the context of the SNIEG, so it's possible for users to know in advance what they can expect, and developers can determine which tools can be shared among projects and how they must be connected to support working flows.

III. Work Ahead

43. The proposal of architectural model that has been presented in this document shows details of a large system in an abstract level. Although some other works to provide details about interactions between the components of the IT platform have been done, there are a lot of aspects that must be reviewed and concretized.

44. Integration of existent systems, like the ones that were mentioned in this paper, is another hard task which needs to be planned to develop specific strategies. Areas that are in charge of some of those systems has started the efforts to communicate them using strategies established by the architecture, per example, incorporating SDMX and web services facilities to exchange information or integrating metadata using DDI.

45. But most efforts to construct the IT platform to support SNIEG that have been already started are considered internal to INEGI, so it's necessary to establish working groups with other Mexican institutions to start the coordination of the development of this federated network that will support with IT a system to produce, disseminate and preserve statistical and geographical information of special interest for the society.

IV. References

- [1] Bass L., Clements, P. and Kazman R., Software Architecture in Practice. Second Edition, Addison Wesley, U.S.A., 2003
- [2] Congreso General de los Estados Unidos Mexicanos, Ley del Sistema Nacional de Información Estadística y Geográfica, INEGI, México, 2008
- [3] IEEE, "IEEE 1471-2000: IEEE Recommended Practice for Architectural Description for Software-Intensive Systems", Institute of Electrical and Electronics Engineers , ANSI ,USA, May-2000, pp 1-23
- [4] Muñoz J., Architectural Design Methodology For Reference Models That Integrate Legacy Systems, UAA, México, 2008