Distr.
GENERAL

Working Paper
22 April 2013

ENGLISH ONLY

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

EUROPEAN COMMISSION
STATISTICAL OFFICE OF THE
EUROPEAN UNION (EUROSTAT)

ORGANISATION FOR ECONOMIC COOPERATION
AND DEVELOPMENT (OECD)
STATISTICS DIRECTORATE

UNITED NATIONS
ECONOMIC AND SOCIAL COMMISSION
FOR ASIA AND THE PACIFIC (ESCAP)

Meeting on the Management of Statistical Information Systems (MSIS 2013)
(Paris, France, and Bangkok, Thailand, 23-25 April 2013)

Topic (i): Architecture

Envisioned Technological Architecture of INEGI
Towards an Enterprise Architecture

Prepared by Juan Muñoz López, INEGI, México

I. Introduction

1. At INEGI, as in other institutions, we have been developing IT solutions following different approaches, from the development of complete systems tailored for specific projects (stove pipe systems), through combination of commercial and in-house solutions, to the use of multiple common systems which can be reused across different projects to give support to specific stages of statistical projects. We think that next step in the evolution of these solutions is to establish a common platform for sharing technological services for all statistical and geographical projects that we are running.

2. This tactic will facilitate the development of integrated solutions that will let us to develop systems with an expanded scope and value. This common platform is an enabler to solve requirements that use to be very difficult to solve, like geo-referencing, integrating and comparing information coming from different projects. Another advantage comes when looking to implement new technologies and trends such as cloud computing, taking advantage of big data, supporting Open Goverment OpenData initiatives, going to social networks, implementing solutions under mobile technologies, etc.

3. In this paper we describe an envisioned path to develop a Technological Architecture under the wide vision that is provided by an enterprise architecture approach. Section II contains a brief description of the context of INEGI, necessary to interpret the point of view from which the proposal has been made. Section III provides a set of needs, goals and aims that the project of designing a technological architecture is pursuing, the fourth section contains the logic of the proposal that has been conceived at this moment, and in the last section some conclusions are presented.

II. Business Case

4. Next paragraphs provide a succinct description of some aspects of the context from which the technological architecture for INEGI is envisioned.

B. Business Context

5. The National Institute of Statistics and Geography of Mexico (INEGI) is an Autonomous Federal Government Body of the Mexican State that is in charge of two main functions:

(a) Coordination of a National Statistical and Geographical Information System
(b) Production of statistical and geographical information for official use

6. The System (named as SNIEG) coordinated by INEGI is integrated by offices which produce statistical or geographical information and belong to different government levels. The System must produce and disseminate statistics about the fields of socio-demography, economy, Government and Justice and environment; as well as geographic information.

7. As an information producer, INEGI provides part of the information that the SNIEG must deliver from the mentioned fields, and also is in charge of the National Accounts System and the Price Indexes. For gathering that information INEGI make use of census, surveys, administrative registries; topological, geographical and geodesic studies, satellite photography, etc.

8. To accomplish its goals INEGI has seven specialized business units. Statistics are produced in five of them, accordingly with different fields (socio-demography, economy, government and justice, environment and special or integrated indicators). Geographical information is produced by the unit which produces statistics from the environment. National accounts and price indexes are generated by the unit which produces economical information. Nationwide operations are carried by 10 regional offices and 34 state offices geographically dispersed in all the Mexican territory.

C. Information

9. INEGI's information is contained in different structures, attending different purposes. Those structures goes from raw data files to multidimensional databases.

10. Raw data (encrypted or unencrypted) files are generated or gathered at the first steps of the information production process. Those files are integrated to different databases for each project. Additional administrative information, metadata and other supporting files are also produced as sub-products of the process.

11. After data is processed, new databases are created. Those databases are inputs for several statistical and geographical products. These databases are standardized and integrated to the main databases of the Institution which support the information systems of the INEGI, like the statistical data warehouse, the bank of economic information and the geographical database. New relational, multidimensional and hierarchical databases with data and metadata are created as a result of this process.

12. INEGI's legal framework says that a historical record of the information must be kept, so some of this information is archived.

D. Systems

13. As each project defines its own strategy to produce information, it becomes necessary to create specific systems to support them. Since some years ago, the Central IT office of INEGI has established a strategy to develop flexible software systems that can be reused in several projects. Under that strategy

similar activities from different projects use the same components and software tools. Questionnaires design, collection of data in field or by Internet, management of field work, codifying and validation, are examples of systems which are used to support information production on different projects.

14. Although those systems cover a good part of the process, they are unconnected, so the areas that use them need to use commercial software or to develop small apps and tools to translate information from one process to the other.

15. Systems for information dissemination are built under the same strategy, one system is used to show information from different projects. Dissemination systems are oriented to satisfy needs of specific groups of users or to deliver information which shares some particularities.

16. Management of all resources from INEGI is made by means of one integrated system developed in-house, which acts as an ERP. This system covers all the administrative process.

E. Technology

17. The computational environment of INEGI supports three platforms: Solaris UNIX, Red Hat Linux and Windows. Most of the servers running those OS are virtualized, just few of them are physical and are used to bring certain general services that are mainly constrained by security issues. Servers that support the main systems are centralized and physically contained in three datacentres. General IT services, like e-mail, are controlled from the main datacentre but in some cases they by servers located at the 10 regional offices.

18. Oracle and PostgreSQL are used to support geographical information; Oracle and MS SQL supports almost the rest of database needs of INEGI, just few specific projects under restricted conditions use other databases like Sybase.

19. The environment provides support for desktop or web systems developed under .NET or Java technologies. Other kind of technologies are discouraged. Emerging mobile platforms like iOS and Android are in an adoption stage.

20. An integrated communication network connects all the offices of INEGI and provides voice, video and data and also access to Internet. The resources of the network are controlled under an active directory divided into three domains accordingly to three geographical regions. Active directory is connected to the management system so the movements made by the Human Resources areas are applied as they happen.

III. New Perspectives and the Need of a New Technological Architecture

21. INEGI's technological environment has been changing and evolving since a long time ago, but new business conditions and new technologies make us review the existent strategies and develop new ones to keep the competitiveness of the Institute facing the future.

22. From our perspective, the architectural design of the new IT environment must provide efficient IT solutions for the support of at least the following business requirements:

(a) Ease the coordination of the national statistical and geographical information network.

(b) Improve efficiency of the internal processes, with emphasis on those directly related to the production and dissemination of statistical and geographical information. Efficiency must be evaluated in terms of reductions in time and resources employed to produce valuable outputs for the Mexican society.

(c) Deliver information to users in a timelier manner and in ways that will facilitate their localization, visualization, use and interpretation.

(d) Support international information comparability, application of domestic and international standards and implementation of best practices.

(e) Flexibility and responsiveness to information requirements from diverse users.

(f) More integration of statistical and geographical information. That integration must go beyond geo-referencing statistical information, to provide new analytical tools to help analysis and decision taking.

(g) Provide society with communication and participation channels to increase its participation in the production of relevant information.

(h) Offer statistical and geographical information services in ways that will facilitate their use by the society, supporting open data principles.

(i) Build and organization that will be in a continuous cycle of evolution and skill growing, making use of previous experiences and incorporating new knowledge to enrich the institution.

(j) As in other Statistical Offices, use of big data to complement statistics is a hot topic, experts from various backgrounds coincide that this will be an important issue in the production of information.

(k) Preserve the relevant statistical and geographical information to provide an historical record that in the future will help society to interpret and to analyse transcendental decisions made in the past (our present time).

23. In addition to that business requirements, we foresee a set of technical requirements to be satisfied, and several relevant technologies that must be supported and used in order to design an architecture that will enable the Institute to develop new high value added services to offer to his users, like:

(a) Support for cloud computing to create flexible, scalable and dependable IT services with low costs is a promise of this technology, but some concerns about security and information owning still being an important issue to consider, in addition existent commercial implementations have some technological limitations when dealing with specialized systems dealing with large databases.

(b) New mobile technologies are changing the way we do the processes. They not only are an important dissemination platform to deliver information and services, but they constitute a new way to develop other tasks like collecting information substituting paper questionnaires or creating ways to recollect data on real time, and enabling new concepts like “crowd-data”. The way we deliver statistical and geographical information is expanding, we have gone from paper to computers, and now to devices of daily use like the tablets, the smartphones and in coming days to other internet connected devices that we will find in our homes.

(c) Desktop virtualization provides a way to deliver customized computational environments to users with independence of a particular physical device. This technology make easier to go from one place to other, and in combination with other technologies enables access to powerful systems on devices with limited resources.

(d) Probably augmented reality will be a way to deliver valuable information focused in satisfy needs of the user in a specific time and under certain context. Maturation of the devices used

to support this technology in the coming years will enhance its practicality and will let them be more commonly used.

(e) We have found that social networks are becoming a fast, effective and low cost way to promote our services, and to go to the people and obtain feedback, but they are evolving and specializing. A next step that doesn't look too far, is their integration to productive processes, in an evolution path in which they will become a coordination tool that will expand the capacities of the technologies that we have today to do that job.

(f) Of course, the need of reusing, sharing and evolving software systems, will be a must; stretched budgets and need to complement specialized skills to build new robust solutions will drive specialist from IT staffs from different organizations to work under common standards and share software and knowledge to develop common agreed solutions that will be customizable to provide effective support to similar process in different contexts.

24. Rational and efficient use of technology will need of the guidance of a planning instrument that will serve as a focal point to align efforts and resources, this is the aim of the technological architecture that now a day is under design and refinement at INEGI.

25. We think that this proposal is a logical evolution step of the way we have been following in the maturation of the use of IT technologies, but at this point it provides elements to consolidate a common way to keep evolving in a fast pace environment. Systems that are currently working like those described in Section II let us to perceive the architecture implementation as a viable project. Other projects like the incorporation of the SDMX reference infrastructure from Eurostat to our processes, the implementation of systems integrating common tools and reusable blocks let us perceive some of the benefits that we can get.

IV. A Technological Architecture to be Integrated in an Enterprise Architecture

26. Accordingly with TOGAF, an Enterprise Architecture is an instrument that "structures and gives context to all enterprise activities delivering concrete business outcomes primarily but not exclusively in the IT domain".

27. Under TOGAF framework there are four architecture domains:

(a) The Business Architecture defines the business strategy, governance, organization, and key business processes.

(b) The Data Architecture describes the structure of an organization's logical and physical data assets and data management resources.

(c) The Application Architecture provides a blueprint for the individual application systems to be deployed, their interactions, and their relationships to the core business processes of the organization.

(d) The Technology Architecture describes the logical software and hardware capabilities that are required to support the deployment of business, data, and application services. This includes IT infrastructure, middleware, networks, communications, processing, standards, etc.

28. As it was formerly described, INEGI's organization corresponds to its main business lines, and on each one, a whole information production process is sustained. In the statistical side, processes to produce information for different domains which share several commonalities appears more than once. A straight

approach to modify organizational structures and consolidate activities would be a way to maximize the use of resources in the whole institute. Common activities would be developed in just one area avoiding duplication and those that are too specialized would be developed by different departments delimited in their functions by very well defined borderlines. New and existent processes would be defined as work flows and coordinated by management groups responsible of specific projects.

29. Although the optimization of the business processes following the method described in the last paragraph could bring a lot of benefits for the Institute, we recognize that it would be a project that could take a long period of time to be done as it depends of a lot of variables that obviously are not controlled by the IT department. Technology cannot wait that long to provide effective solutions.

30. Use of GSBPM to characterize statistical processes can provide us a key tool to solve the problem. Besides is desirable or not to change the organizational structure, we can classify tasks made by each department as activities from the second level of the mentioned model. In a very simple IT logic, those activities that are similar, can be supported by similar tools which just need to be customized to the specific needs defined by each project. Also, software systems can be characterized in this way transforming them in reusable building blocks that can be connected to conform working flows which can be automated.

31. As described for the organizational areas, each system need to be developed with a very well defined scope, and delimited by standardized interfaces that make them easy to connect. Inputs and outputs of each piece of software need to be defined following a set of rules to make easier the construction of an information flow across a complete process. In this case, each information object needed by the systems could be described with a common language and structure like the one provided by GSIM.

32. If a new institute was born and did not have any system, the required software solutions would be designed following the described principles, then we would have a complete, flexible and homogeneous environment that could be adapted to different contexts and circumstances without too much effort. But reality is different, the Institute has invested a lot of resources on systems although not the "optimal", they provide support to existent processes and it would be very hard to get rid of them before getting the benefits that has been calculated to be obtained during its lifecycle. It’s necessary to give them a place in the new architecture and facilitate their eventual replacement by more optimized parts.

33. As the systems that now a day are running at INEGI were not built to provide support to just one GSBPM activity. We will need to identify their functionalities and then map them to the GSBPM activities that are being covered. We will find that in some cases some needed functionalities of the activity are not covered and it will be necessary to develop pieces of software to complement the system. In other cases, some functionalities will be beyond the scope of the activity, we will need to take the decision on completely covering the other activity that is partially covered by complementing the system or discontinue the use of that functionality replacing it by other piece more suited to the process needs.

34. After we have designed all needed functionality we will design the interfaces that will be used to connect this kind of “super-brick” to the others. Sometimes it will be possible to connect the new interface to the code of the existing system and have the work done, but in most of the cases we will need to rely on the information layer to do this job. Interfaces will need to take information directly from the databases and files that has the system to provide the necessary “pipes” to connect the “information flow”.

35. Separation of interfaces from functionality will bring us an additional benefit, it will be easier to adapt pieces of software to different technologies and devices.

36. A lot of flexibility in the pieces of software to deal with different information structures will be needed, so it will be necessary to keep a separation between software functionality and persistent information structures. The use of connectors to access different databases and files will be necessary to provide that separation and provide a way to replace them in the future.

37. It is necessary to remark the importance of a well-designed information layer. Maybe at the beginning we can support some redundancy and inconsistencies that need to be corrected with software, but at the end we need to have an optimized system of databases and files that will serve as a foundation for all the services provided by the IT environment. The capacity of the system to have a good performance, to be reliable, to satisfy information needs and to be adaptable to new requirements in great (maybe the greatest) part depends of very well designed database models that have been conceived to provide an optimized and integrated support to multiple systems.

38. The technological layer is defined at the end but is not less important, it provides the necessary elements to make a sustainable implementation of all the architecture. A coherent and solid IT platform constituted by a consistent set of interoperable technologies will provide the fundaments to bring all conceptualized systems to reality. But in the logic of layers separation the implementation of the systems, we will obtain a high grade of independence from the solutions provided by a specific vendor. That separation will be feasible by the extended use of technological standards and open software solutions where possible.

39. Computational support, communications and security, are elements that must conform a standardized common platform composed of software and hardware where systems and information will be relying to perform its job.

V. Conclusions

40. The fast evolution of the IT technologies has led us to the point where adaptability and flexibility to adapt new ways of doing things are a constant. Without a clear vision and a planned path our role will become more reactive, and as the complexity and speed of change increases, it will be more difficult to remain competitive. A technological architecture provides us with planning tools to be proactive, and put ourselves in a position where we can make better use of both existing technologies as those that emerging along the way.

41. An Enterprise Architecture would provide a way to improve efficiency in the whole organization, but it is difficult to implement in all its extension, so we need to develop strategies to use the concept in a landed and contextualized environment to get at least some of the benefit that it could provide.

42. The logic of the majority of standards tell us that benefits are greater as long as more individuals make use of them. As long as more institutions follows a similar route to build its architectures it will be more easier to share common solutions, the effort that is being drive by the HLG to set the principles for an standardized architecture will be very important to make feasible a practical collaborative development and sharing of solutions.

43. This paper pictures a first approach that is under refinement and may not be enough detailed to directly be applied to other contexts that are different to the one of INEGI, but it provides some ideas that maybe are useful for other Institutions sharing some commonalities with ours.