

9th National Convention on Statistics (NCS)
EDSA Shangri-La Hotel
October 4-5, 2004

**Web-Enabled Databases as Tools
for Disseminating Statistics**

by
Candido J. Astrologo, Jr.

For additional information, please contact:

Author's name:	Candido J. Astrologo, Jr.
Designation:	Head
Agency:	National Statistical Coordination Board
Address:	Midland Buendia Bldg., 403 Sen. Gil Puyat Avenue, Makati City
Telefax:	(632) 890-8456
E-mail:	cj.astrologo@nscb.gov.ph

Web-Enabled Databases as Tools for Disseminating Statistics¹

by
Candido J. Astrologo, Jr.²

ABSTRACT

In today's world, the Internet has become an indispensable tool for research work. From aardvark to zygote, virtually any topic can be found in the Internet. But how about statistics? How about data on the world's population or the inflation rate in Tuvalu? Although statistical data on social, environment, and economic sectors are readily available on the Web, these are oftentimes presented on websites which are no more than showrooms of statistical information. Users are therefore limited to viewing statistical data not of their own choosing in terms of contents, layout and format.

The changing landscape of I.T. and the convergence of web and database servers present an opportunity for website developers to explore the possibility of web-enabling their databases. Further, the UNSC/ECE Framework on Best Practices in Designing Websites for Dissemination of Statistics calls for a "flexible access to database" to be provided by a statistical office on its website. From the standpoint of a statistical organization, the use of web-enabled databases provides an interactive way of disseminating the statistics they produce where users can view or download only the data they need, whenever they want, and in whatever form they like.

Despite the relatively high cost of overheads in developing and maintaining web-enabled databases, the overriding need of such technology is imperative of statistical offices. On the other hand, a simple yet effective interactive database that can be accessed on the website of a statistical office is essential for them to be responsive to the ever-growing information needs of their clients.

This paper describes how the National Statistical Coordination Board (NSCB) developed and maintains the various web-enabled databases that are accessible from its website. More importantly, this paper presents development strategies using simple and common technology which can be replicated by other statistical offices.

Keywords: Internet; World Wide Web; databases; statistics, statistical dissemination

I. INTRODUCTION

The birth of the Internet in 1991 paved the way for fueling out vast amount of information that can be accessed anywhere, anytime. The Internet provided huge opportunities to organizations to develop innovative and effective dissemination strategies while at the same time cope with the mounting challenges of new technology.

¹ Presented during the 9th National Convention on Statistics at the EDSA Shang-rila Hotel, Mandaluyong City, Philippines, 4-5 October, 2004

² Statistical Coordination Officer VI of the National Statistical Coordination Board. The views expressed in this paper are those of the author and do not necessarily reflect the views of the organizations he represents.

Government offices took advantage of the Internet through the World Wide Web as a public service tool in sharing and disseminating information about their organization; delivering the latest news and developments relating to their functions; marketing, promoting and selling their products and services; and releasing statistics they produce.

Meanwhile, government statistical offices (GSOs) has recognized the Internet as an effective channel for disseminating statistics noting the far-reaching benefits the Internet provides in terms of usability, accessibility and cost efficiency. Some of the advantages that the Internet offers over traditional channels of information dissemination are as follows: [1]

- a wide variety of information can be published and released that can be accessed from one location regardless of time and distance
- users can instantaneously access information as updated
- greater quantity of information can be published than would be feasible in paper
- greater quantity of information can be published in context through hyperlinks to related resources

When the time Internet usage in the Philippines became widespread in 1996, the government and private sectors started to develop their websites with the primary purpose of self-promotion. As the technology on web page development was at its infancy, majority of the websites were made out of an organization's corporate brochure.

Today, GSOs are more concerned in managing their websites than five years ago when most of them were still contemplating of developing a website or not. But the challenges on the GSO do not stop with the launching of its website. In fact, once the GSO has made its web presence felt, more challenges will be encountered, particularly in designing its website to effectively deliver its statistical products and services to its target users.

In July 1997, the National Statistical Coordination Board (NSCB) launched its website using a free hosting account with a local ISP³. The web pages were created in pure HTML4 and initially contained information on NSCB's functions and organizational structure; products and services; press releases on statistical news and developments; and statistical data presented in tabular form. During that time, the NSCB was one of the few government offices and the only GSO which was able to establish a web presence.

II. GOOD PRACTICES IN DISSEMINATING STATISTICS

The policy of GSOs in the dissemination of statistics is in accordance with the United Nations Fundamental Principles of Official Statistics. The first of the 10 principles stipulates that "... official statistics that meet the test of practical utility

³ ISP - Internet Service Provider - A business that supplies Internet connectivity services to individuals, businesses, and other organizations.

⁴ HTML - HyperText Markup Language - The markup language used for documents on the World Wide Web. A tag-based notation language used to format documents that can then be interpreted and rendered by an Internet browser..

are to be compiled and made available on an impartial basis by official statistical agencies to honor citizens' entitlement to public information." In effect, the first principle implies that for information to be widely known and available, dissemination activities should provide information in the form required by users. Further, the presentation of statistics should follow sound statistical principles so that they are easy to understand and impartially reported [2].

In relation to the said principle, the NSCB Executive Board issued NSCB Resolution No. 10 in 1997 calling for the adoption and implementation of the Government Statistics Accessibility Program (GSAP) as formulated by the NSCB Technical Staff. The Program primarily aims to improve the accessibility of users to statistical information and services through a series of standards and policies on information dissemination among producers of statistical information [3]. The first of this series is the General Standards on Statistical Information Dissemination (GSSID) which was issued through NSCB Resolution No. 8 in 1999. Among the standards set forth which relate to the dissemination of statistics are as follows [4]:

- provision of all appropriate metadata
- dissemination of all available disaggregations
- making available the use of the most micro level summary statistics
- development and maintenance of a data archiving system
- immediate dissemination of statistical information in any form, upon availability or even before the release of the formal publication
- extensive dissemination of statistics using both print and electronic media.

One appropriate technology that can be utilized to meet the standards is a database. A database system can integrate along with statistical data their corresponding metadata. It can likewise store enormous amount of information even up to the micro level with all available disaggregations. Further, a database can act as backend⁵ in processing queries of users and producing tables for publications. And lastly, a database can serve as an archiving application to store and preserve statistical information and documents for use of current and future generations.

The integration of databases and Internet technologies has spawned a powerful and effective tool in information dissemination - web databases or web-enabled databases⁶. As a dissemination tool for statistical data, web-enabled databases present several advantages over publications and even CD-ROMs, to wit [5]:

- accessible via any browser, regardless of location and time
- provides interactivity and variety of options to users
- produce query results instantaneously
- allow simultaneous access by several users

⁵ In applications, software or a feature of software that provides an interface to another application or tool compared to "backend" which refers to the part of the program that runs on the server

⁶ A data store or information repository that can be accessed via a query language or on application program accessed via other Web applications; especially forms that are developed using standardized HTML tags.

- offers wide access to information in both public (web) and private domain (intranet)
- ensure instant capture of latest data
- give opportunities to users for feedback and reactions

During the UNSC/UNECE⁷ Conference of European Statisticians in 2001, the Conference published the methodological material “Best Practices in Designing Websites for Dissemination of Statistics.” The material listed four major aspects that should be taken into consideration by GSOs in designing and developing their websites, to wit [6]:

- Users – users are the crucial purpose of a website and should be the center of attention; it is important to know who the users are, what they need, how they use the data, what their competence levels are; there is a need for a user’s feedback
- Content – basic and key statistical data; fixed statistical tables; web-readable statistical publications; pre-defined tables, time series data; flexible access to databases; downloadable data/documents in user’s desired format; profile of the GSO including contact persons, products and services; metadata support; links to other statistical and related sources
- Architecture and Functions – fast response time; comprehensible navigation structure; site-wide consistent layout and design; search function; compatibility with most browsers and screen resolutions
- Development and Maintenance Issues – competence of developers; support of top management; editorial board; application of new technology

While the NSCB has followed most of the standards and recommendations set forth in the aforementioned document, the limited manpower and financial resources place a constraint in adopting the rest of the guidelines. On one hand, continuous efforts are being done by the NSCB not only to conform to statistical web publishing standards and comply with the guidelines but more importantly in addressing the online statistical data needs of its stakeholders.

III. THE NSCB WEBSITE

A. Short History

Developed in 1997 to establish an identity for the NSCB on the Internet, the NSCB website has become one of the most important services offered by the NSCB to its stakeholders. The website started with postings on basic information about the NSCB: its mandate, functions and organizational structure; articles on statistical news and developments; products and services; and a few statistical tables, all in not more than 50 static web pages.

During that time, the NSCB was one of the few government offices and the only statistical office which had a website. Although the website was developed

⁷ United Nations Statistical Commission/United Nations Economic Commission for Europe

in-house and the pages were purely made in HTML, the website's structure was user-friendly and the pages were regularly updated. This may be one of the reasons leading to the Philippine Webby Awards⁸ selection of the NSCB website as the Best Outstanding Website in the government category in 1998.

A year later, with the assistance of the Swedish International Development Agency (SIDA) and Statistics Sweden (SCB) through the Development of the National Statistical Information Center (NSIC) System Project, the NSCB was able to acquire its own web server and leased-line connection, and consequently obtained its own domain name: nscb.gov.ph. Also through the Project, the SCB-developed PC-Axis Family Software [7] was introduced and later served as the basis in the development of the i-stats⁹ data subscription service of the NSCB.

The NSCB expanded its website by covering more sectoral time series and cross sectional data with corresponding metadata including a glossary of statistical terms and concepts and technical notes. In 1999, the NSCB website made available for access its first web-enabled database - the telephone area code of the Philippines is a searchable web database that allows users to look for the area code by entering the name of a locality. And for the second time, the NSCB website was judged as the Best Outstanding Website in the government category by the Philippine Webby Awards.

At present, the NSCB website is updated at least four times a week with two to three new web articles released every week. Statistical databases like the i-stats are updated as soon as new data are released. The StatRef (an online catalogue of the NSIC library holdings) is updated every Wednesday while the various classification systems are updated upon official release of amendments.

Usability of the website has indeed been high as shown by the increasing growth in the number of page views which numbered from less than 50,000 in 1997 to more than 2.5 million in 2003. The NSCB website as of July 2004, is comprised of around 4,500 web pages consuming around 30MB of disk space. And for seven years of providing online services, the NSCB has upgraded its bandwidth three times just to meet the information needs of its local and international clients.

B. Static or Dynamic

While it is true that, in contrast to printed and CDROM versions of publications, the marginal costs of informing an additional user through the Internet is minimal, there are significant costs in operating a website. The cost not only includes hardware and software acquisition and maintenance but also manpower skills development. In particular, as the number of pages in a website increases, the costs of maintaining and updating the website become significant. And that is notwithstanding the time and effort spent in doing repetitive and tedious manual updating of pages [1].

When the NSCB went online in 1997, the statistical data posted on its website were presented in pre-formatted statistical tables on static web pages.

⁸ Philippine Webby Awards (now the Philippine Web Awards): <http://www.philippinewebawards.com>

⁹ i-stats (Interactive Statistical Databases): <http://www.nscb.gov.ph/istats>

Despite the use of standard templates in presenting statistical tables, the process of manual updating and layouting took much of the time of the web developers. Such method was found too rigorous and inefficient particularly in updating related statistical tables.

The development of the NSCB website has gone through several phases. It started off with static web pages containing textual and statistical information that were manually updated and later included dynamic pages in producing statistical tables and interactive results based on a user's query. Further, early layouts of the web pages were created individually on a per-page basis. But with the introduction of CSS¹⁰ and DHTML¹¹, modifying the layout and design of the web pages became a straightforward job.

C. Flat or Relational

Storing data onto a web server using flat files (structured text files) is conceptually equivalent to operating a relatively simple database system in which each database is contained in a single table. In contrast, relational database¹² systems can use multiple tables to store information and each table can have a different record format. Relational systems are more suitable for large applications, but flat databases are adequate for many small applications [8].

During the implementation phase of the NSIC Project, the PC-Axis Family software was introduced to the NSCB and to other government statistical offices to address the problem of integrating statistical data with corresponding metadata in the process of data build-up and report generation. With technical assistance from Statistics Sweden, the NSCB was able to develop several databases in PC-Axis format. Prior to this, the NSCB has no organized statistical databases except those stored in MS Excel spreadsheets, MS Access databases and in MS SQL Server which were done out of necessity.

The PC-Axis Family Software consists of several applications in the conversion, editing, and generation of statistical data and tables. The strong point of the software is its seamless integration of metadata with statistical data which allows specifying notes down to the cell value level. It also provides a wide array of functions like pivoting, grouping, graphing, and mapping (a la GIS¹³). Likewise, it can export to various file formats and can connect to an SQL¹⁴ database through the PC-Axis common macro metadata model. But the "killer" application

¹⁰ CSS - Cascading Style Sheets - a Hypertext Markup Language (HTML) specification developed by The World Wide Web Consortium (W3C) that allows authors of HTML documents and users to attach style sheets to HTML documents. The style sheets include typographical information on how the page should appear, such as the font of the text in the page.

¹¹ DHTML - Dynamic HTML - a technology designed to add richness, interactivity, and graphical interest to Web pages by providing those pages with the ability to change and update themselves dynamically—that is, in response to user actions, without the need for repeated downloads from a server. This is done by enabling the interaction of HTML, cascading style sheets (CSS), and JavaScript.

¹² Relational Database (RDB) - a database or database management system that stores information in tables—rows and columns of data—and conducts searches by using data in specified columns of one table to find additional data in another table.

¹³ GIS - Geographic Information System - an application or suite of applications for viewing and creating maps. Generally, geographic information systems contain a viewing system (sometimes allowing users to view maps with a Web browser), an environment for creating maps, and a server for managing maps and data for real-time online viewing

¹⁴ SQL- Structured Query Language - A database sublanguage used in querying, updating, and managing relational databases—the de facto standard for database products.

included in the suite is the PX-Web component which is a middleware between the client and the server allowing interactive queries among users in the production of user-customized statistical tables [7].

The metadata-rich feature of a PC-axis file is attributed to its being a flat file. And being a flat file, it is compatible with most applications and smaller in file size. PC-Axis files store statistical data in delimited format that are queried by a custom web application (PX-Web) designed to show two-dimensional data tables together with corresponding metadata on a user's browser.

Presently, the NSCB maintains almost 200 PC-Axis files containing time series and cross-sectional economic, social and environment statistics with some indicators having reference years dating back to 1945. These "databases" are accessible by all NSCB staff and are the sources in the generation of statistical tables for NSCB publications like the annual Philippine Statistical Yearbook (PSY) and the monthly Economic Indicators (EI). Most importantly, the PC-Axis database system is the backend of the successful i-stats data service of the NSCB which it offers on a paid subscription basis on its website.

PC-Axis is superb in handling metadata and in generating cross tabulations of statistical data; unfortunately, it is not a relational database. Data in one PC-Axis file could not be related to other data on other PC-Axis files. Further, metadata are integrated in each file, hence there is a need to edit each PC-Axis file when a metadata is revised. Although the SQL component of PC-Axis supports a call from a single metadata database, there is still a need to regenerate the PC-Axis files to reflect the updated metadata.

In a relational database, data are stored in tables which are related to other data in other tables through a common identifier. This system differs from flat-file databases wherein each database is self-contained in a single table (file) with no relationships with other tables (or databases). Relational databases are nowadays the most common way of storing and managing data and most relational database management systems provide robust, secure and efficient data management systems which are needed for data-intensive work like those dealing with statistical information [9]. Because of this, the NSCB, in addition to having PC-Axis files, thought it practical to maintain its data holdings on relational databases, such as MS Access and MS SQL Server as complement to PC-Axis databases.

The NSCB currently operates an MS SQL Server where most of the time series data and classification systems are stored. The different divisions either have their sectoral statistical databases in MS Access or Excel format. Conversions of data files from the SQL Server to PC-Axis formats are done manually since the NSCB has not implemented the PC-Axis SQL macro metadata model yet.

V. NSCB WEB-ENABLED DATABASES

The nature and content of the databases maintained by the NSCB are based on its mandates and functions. As the policy-making and coordinating body

on statistical matters in the country, the NSCB is tasked, among others: to allocate statistical responsibilities among government agencies by designating the statistics to be collected by them; review and clear, all funds for statistical operations; develop, prescribe, and maintain appropriate framework for statistical coordination; and prescribe uniform standards and classification systems in government statistics. [10]

The NSCB likewise releases official statistics on national, regional and inter-industry accounts, poverty, human development index, foreign direct investments, composite leading economic indicators and economic indices, among others. The NSCB also maintains various statistical frameworks that involve the compilation of various social, environmental and economic indicators.

Looking at the volume and diversity of the NSCB outputs, it is hard to imagine how all of these can be integrated in a manner that the dissemination activities of the NSCB will be effective and focused in meeting the information needs of its target users. The NSCB has adopted several channels of distributing its products and services. It started with printouts and later released full publications. Then these publications were converted to electronic versions that were distributed via CD-ROMs. And with the advent of the Internet, the electronic versions of NSCB publications were delivered through e-mails and most of the textual and statistical databases were web-enabled.

NSCB's decision to web-enable its databases was mainly because of dissemination issues. Studies show that databases distributed in CDROMs were found limited in terms of usability. Considering that a large number of statistical data are updated at least on a monthly basis, the CDROM-based databases become obsolete after one month. This is one reason why NSCB has limited its sale and distribution of databases in CDROMs and instead has been heavily promoting its i-stats data subscription service. Through the i-stats, data users can simply download the latest releases and add or merge them to their existing databases.

There are currently eight web-enabled databases in the NSCB website that can be accessed via a single entry page. The NSCB collectively calls them ActiveStats (<http://www.nscb.gov.ph/activestats/>). Table 1 shows the eight web-enabled databases with their corresponding descriptions.

**Table 1. NSCB Web-Enabled Databases
(as of August 2004)**

Name/ URL/ Description/ Features	Scripts / Standards	Architecture (Type / Backend / Components)	Frequency of Updating	Data Sets / Records
i-stats http://www.nscb.gov.ph/istats/ (interactive statistical databases) a paid subscription service for viewing and downloading datasets. The datasets cover time series social and economic indicators produced by the different government agencies in the Philippine Statistical System (PSS). Free sample datasets. Paid subscription.				
Inherent to PC-Axis features (downloadable tables in various file formats, graphing, pivoting, metadata)	VBScript, HTML, JavaScript /	Statistical / PC-Axis for statistical data; MS Access	Monthly, quarterly, annually	182 data sets

support); search capability; user authentication; sessions to maintain state; linear	CSS	for user authentication / PX-Web		
Countryside in Figures (CIF) http://www.nscb.gov.ph/countryside/ an integrated database of provincial level statistics which covers seven sectors on the 79 provinces in the Philippines which are ranked based on their performance. Free access but limited to a list of 10 provinces. Paid subscription allows viewing of all provinces				
metadata; user authentication; sessions to maintain state; cyclical	VBScript, HTML, Javascript / CSS	Statistical / MS SQL / ASP (ADO, OLE DB, SQL)	annually	90 data sets
StatRef http://www.nscb.gov.ph/statref online statistical reference system of library holdings available at the National Statistical Information Center (NSIC). Free unlimited access.				
Search-based; linear	VBScript, HTML, Javascript / CSS	Textual / MS SQL / ASP (ADO, OLE DB, SQL)	Every Wednesday	Around 10,000 records
Philippine Central Product Classification (PCPC) http://www.nscb.gov.ph/activestats/pcpc a searchable database of standard classification of goods and services in the Philippines based on the full document of the 2002 PCPC. Free access limited to 5 searches. Paid subscription for full access.				
Search-based; linear; linked with PSIC	VBScript, HTML, Javascript / CSS	Textual / MS SQL / ASP (ADO, OLE DB, SQL)	Upon release of amendments	Around 7,000 records
Philippine Standard Industrial Classification (PSIC) http://www.nscb.gov.ph/activestats/psic a searchable online database of statistical classification of economic activities obtaining in the country based on the full document of the 1997 PSIC. Free unlimited access.				
Search-based; linear	VBScript, HTML, Javascript / CSS	Textual / MS SQL / ASP (ADO, OLE DB, SQL)	Upon release of amendments	Around 1,500 records
Philippine Standard Geographic Code (PSGC) http://www.nscb.gov.ph/activestats/psgc a searchable database on geographical classification which serves as a tool for ensuring the comparability of statistics relating to the geographical areas of the country. Free unlimited access.				
Search-based; cyclical; linked with CIF	VBScript, HTML, Javascript / CSS	Textual/Statistical / MS SQL / ASP (ADO, OLE DB, SQL)	quarterly	Around 42,000 records
Area Codes of the Philippines http://www.nscb.gov.ph/activestats/areacode searchable database of telephone area codes of the Philippines. Free unlimited access.				
Search-based; linear	VBScript, HTML, Javascript / CSS	Textual / MS Access / ASP (ADO, OLE DB, SQL)	Upon release of updates	Around 400 records
Zip Codes of the Philippines http://www.nscb.gov.ph/activestats/zipcode searchable database of zip codes for areas and postal offices. Free unlimited access.				
Search-based; linear	VBScript, HTML, Javascript / CSS	Textual / MS Access / ASP (ADO, OLE DB, SQL)	Upon release of updates	Around 2,000 records

A. Features

Except for the i-stats whose features are inherent to PC-Axis and the CIF which is primarily a menu-based database, the other web-enabled databases are mostly search-based where users enter a keyword or phrase to look for a particular code, classification, or reference material. The resulting page of the query shows a list of matched records from the database.

Six web-enabled databases are linear in terms of page flow. That is, a series of queries and interactions lead to a final page with no more links as compared to other databases that employs cyclical navigation where the user is provided an endless series of related hyperlinks. The CIF and the PSGC feature cyclical links.

Although the PCPC and PSIC are separate databases, they are linked together by a common attribute, the PSIC code. By clicking on the PSIC code in the PCPC page, the user will be brought to the PSIC page showing further information about the code. The PSGC and the CIF are likewise linked via the Province code. Users of the PSGC database can click on a province to view the time series data on a selected indicator for that province. The StatRef on one hand is a simple web-enabled database which is similar to an electronic card catalogue. The StatRef enables users to look for reference materials and have them reserved for their use prior to their visit to the NSIC library.

The i-stats, CIF, and PCPC require a paid subscription account for unlimited browsing. Because an active user account is required, cookies are utilized to authenticate and maintain session states of logged in users. Sessions expire once the user closes the browser.

B. Scripts and Standards

The NSCB uses Active Server Pages (ASP)¹⁵ technology as the server-side scripting environment. ASP's ability to combine HTML, scripts and components makes it an ideal programming environment in developing web applications. The choice of ASP over other programming technologies was not planned during the time when the NSCB decided to develop web-enabled databases. It is more of that fact that most of the NSCB IT staff are familiar with Visual Basic, making it easy for them to adopt VBScript¹⁶ which ASP supports. Further, since the IT environment of NSCB was Microsoft-centric, it was a matter of using what is known and what is available.

To a shorter extent, the NSCB employs JavaScript¹⁷ mostly used to validate a user's inputs in filling out forms for subscription, feedback, data request or queries on databases. If a user submits a form that is not completely filled,

¹⁵ ASP - Active Server Pages - a Web-oriented technology developed by Microsoft that is designed to enable server-side (as opposed to client-side) scripting. ASPs are text files that can contain not only text and HTML tags as in standard Web documents, but also commands written in a scripting language (such as VBScript or JavaScript) that can be carried out on the server.

¹⁶ VBScript - Visual Basic Scripting Edition - a subset of the Visual Basic programming language, optimized for Web-related programming. As with JavaScript, code for Visual Basic, Scripting Edition is embedded in HTML documents

¹⁷ JavaScript - a scripting language developed by Netscape Communications and Sun Microsystems that is loosely related to Java. JavaScript, however, is not a true object-oriented language, and it is limited in performance compared with Java because it is not compiled.

he/she is alerted to fill in the lacking information, otherwise he/she is not allowed to proceed to succeeding pages. JavaScript in addition to CSS is used in the generation of the navigational menu of the site and in some sections are used to randomly show images in a way similar to banner advertisements.

The NSCB website also takes advantage of Cascading Style Sheets (CSS) to maintain a site-wide consistency in the design and layout of web pages.

C. Architecture

Except for the i-stats, the rest of the NSCB web-enabled databases are deployed in either MS Access or MS SQL Server which act as backends for producing statistical and textual information via the web. For the i-stats, tables from MS SQL Server are converted to PC-Axis formats and deployed in the web server. Textual databases like the PCPC, PSGC, PSIC and the StatRef are also stored in MS SQL Server while the databases for telephone area codes and zip codes are in MS Access.

As ASP is the programming environment that runs on the NSCB's web server, several ASP components are used to provide connections between user interfaces to NSCB databases. ADO¹⁸ components are utilized to serve as data access interfaces to communicate with OLE DB¹⁹ compliant data sources such as MS Access and MS SQL Server. For users to connect to the databases, ADO program commands are embedded in ASPs. Further, SQL statements are included to allow query operations on the databases..

At present, the NSCB operates a MS SQL Server (6.5) which runs on a different machine but is connected to the IIS (4.0)²⁰ web server through a trusted connection. The NSCB recognizes the need to upgrade its server not only because it is obsolete but also to accommodate the large number of users simultaneously accessing the website and the databases.

VI. CHALLENGES AND FUTURE DIRECTIONS

In pursuit of an effective statistical dissemination policy, the NSCB took different approaches to address the varying data needs of its clients, not only in the forms or types by which statistics are delivered but also the means and channels through which statistics are disseminated. However, the demand for new statistics and the continuing emergence of new technologies pose a bigger challenge for the NSCB to adopt innovative data dissemination strategies.

Further, the architecture of the existing web-enabled databases of the NSCB still needs improvement not only in terms of structure and schema but most importantly, its scalability and interoperability. Cognizant of the fast pace of

¹⁸ ADO - Active Data Objects - an application programming interface (API) developed by Microsoft for applications that access databases.

¹⁹ OLE DB - Object Linking and Embedding DataBase - An application programming interface developed by Microsoft for accessing databases. OLE Database is an open specification that can interface with all types of data files on a computer network.

²⁰ IIS - Internet Information Server - Microsoft's brand of Web server software, utilizing HTTP (Hypertext Transfer Protocol) to deliver World Wide Web documents. It incorporates various functions for security, allows for CGI programs, and also provides for Gopher and FTP services.

which technology evolves, the NSCB should find measures on how to keep its IT hardware and software recent. And more importantly, NSCB should continue to explore innovative ways of disseminating statistics that will enhance further the usability of the NSCBs web-enabled databases in disseminating statistics.

A. Emerging Internet Technologies

1. Extensible Markup Language (XML)

Although XML²¹ technology became a W3C²² recommendation in 1998, it was in late 2000 that it became popular largely because of the proliferation of RSS²³ feeds on the Internet. As defined by the W3C, XML is a simple and very flexible text format derived from SGML²⁴ (ISO 8879) which is originally designed to meet the challenges of large-scale electronic publishing. XML is also playing an increasingly important role in the exchange of a wide variety of data on the Web and elsewhere. According to Microsoft's Bill Gates, XML is a panacea for the IT industry that will be key to the development of web services [11].

With XML, developers can define their own tags and attributes and their own document structure for use on the Internet. XML is likewise extensible, platform-independent, and fully Unicode compliant. This and other characteristics of XML have caused it to be widely adopted as a data representation format mainly because of its ability to easily represent both tabular data (such as relational data from a database or spreadsheets) and semi-structured data (such as a Web page or business document). This led to the widespread adoption of XML as the lingua franca of information interchange [12].

For purposes of statistical information dissemination, the use of XML for the exchange of data requires a certain degree of standardization of metadata to be able to use the real advantages of such an open format. Downloading of statistical data in XML format will probably be the most widely used format in the near future [7].

The current RSS approach of syndicating news and information through XML feeds can be used as a practical implementation of XML in disseminating statistics. In addition, the ability of embedding XML pages from other websites and include them in pages of another website presents a huge potential of data sharing among statistical offices.

2. Web-based Geographic Information System (GIS)

²¹ XML - eXtensible Markup Language - a condensed form of SGML (Standard Generalized Markup Language). XML lets Web developers and designers create customized tags that offer greater flexibility in organizing and presenting information than is possible with the older HTML document coding system.

²² W3C - a standards body based in the United States, Europe, and Japan. The W3C is dedicated (in part) to encouraging the development of open Web standards, such as the HTML and XML document markup languages, to promote interoperability and assist the Web in achieving its potential.

²³ RSS - Really Simple Syndication/Rich Site Summary, a lightweight XML format designed for sharing headlines and other Web content.

²⁴ SGML - An information management standard adopted by the International Organization for Standardization (ISO) in 1986 as a means of providing platform- and application-independent documents that retain formatting, indexing, and linked information.

The geographical composition of the Philippines which consists of regions, provinces, cities, municipalities and barangays poses a big challenge for the NSCB not only in the generation of statistics but also in terms of presentation and dissemination. Government programs that require monitoring for sound distribution policies essentially need local-level statistics while businesses need regional statistics in order to formulate their marketing strategy. Thus, there is a complex user situation wherein users want different types of statistics for different types of geographical breakdown.

Where geography is a factor in disseminating statistics, GIS is the only technology which can present statistics along and across geographical boundaries clearer and more understandable. The advent of easier to use, affordable desktop GIS has provided users with a new and flexible means of analyzing information and developing policy to address local and regional issues. GIS output provides a visual representation in a geographic context recognizable by a broad range of stakeholders in the policy development and planning process.

The NSCB has taken initial steps in the use of GIS in disseminating statistics needed for policy making. In 2002, under the Project on Re-engineering the Government Statistical Services, the NSCB produced prototype poverty maps for the provinces of Laguna and Albay. Using GIS software, municipal indexes were computed and plotted on the maps which were made for the individual indicators; the component indexes; and the composite poverty indexes. The maps serve as very useful tools to local government units as well as local chief executives in the identification of priority areas that should benefit from anti-poverty programs [13].

B. FUTURE DIRECTIONS

The heterogeneous nature of the databases maintained by the NSCB is a problem that needs to be addressed. Integration of these databases such that more and varied information can be disseminated must be done. Likewise, other online dissemination strategies should be explored using available technologies. On one hand, shortage on manpower, limited fund support and slow upgrading of IT resources should be the problems to be solved first.

1. Integrated Classification Systems

Although the PCPC is itself an integrated database of different product classification systems, its current usage is limited to look-ups of classification codes of products and the industries that produce these products. To take full advantage of the PCPC database is to link the codes to actual statistics on exports, imports, production and even tariff rates.

Producers of statistics which utilized the PCPC or the PSIC codes can link to the database to obtain the latest classification codes. But at the current state of the database's architecture, this can not be done remotely and the only way is for the data producers to have a physical copy of the database in their own servers.

This is where XML can come in. Unfortunately, the steep learning curve of XML towards this endeavor and its “novelty” still need to be addressed.

2. Web-based GIS

As local-level statistics are always in demand, the NSCB should not only work towards the data production process but also in the dissemination aspect. The Internet provides an enormous opportunity to disseminate statistics in any form to a large clientele. With the development of GIS tools for the Internet, statistical agencies are able to create new mapping applications and data products. And by bringing GIS on the web, users can access these applications at a very low cost and with minimal computer knowledge.

There are currently a number of available mapping software applications that provide web applications for GIS. The PX-Map component of the PC-Axis Family Software can be explored by the NSCB to deploy a web-based GIS.

3. Statistical Portal of the Philippines

In its 17 years of government service, the NSCB has been recognized locally and internationally of implementing significant programs towards the maintenance of a well-coordinated statistical system. In terms of statistical dissemination activities, the NSCB through the NSIC has initiated various activities in responding to the data needs of its clients. The NSIC serves as a one-stop center of government statistical products and services.

The challenge now for the NSCB is to bring statistics from the various government data producers on the web that can be accessed in a single website and not just through hyperlinks to the websites of other GSOs. This can be done through the establishment of a Statistical Portal of the Philippines ²⁵.

The Portal will serve the statistical information needs of data users by providing them the latest key statistics, statistical news and developments, time series and cross-sectional data, and everything about statistics. But, the portal should be personalized. That is, the user can select his/her own sets of statistics that every time he/she visits the portal, only those statistics and related information are shown. With this approach, the user is given a personalized website where he/she finds only the statistics he/she needs.

VI. CONCLUSION

Internet has changed the way GSOs disseminate their statistics. The introduction of new technologies has further opened up various approaches by which statistics can be disseminated. The integration of databases and the web is a paradigm shift for GSOs that are accustomed to having databases as data storage systems and the web as a “window to the world.”

²⁵ A portal is a collection of links, content, and services designed to guide users to information they are likely to find interesting—news, weather, entertainment, commerce sites, chat rooms, and so on. Yahoo!, Excite, MSN.com, and Netscape NetCenter are examples of portals.

For a start, common software applications like MS Access can be used as a relational database system. By deploying it to a web server and developing user interfaces, a simple web-enabled database is ready to serve its purpose.

The advantages of web-enabled databases over traditional database systems are countless, especially when it comes to disseminating statistics. The time has passed not only from using databases as mere data storehouses but also from developing websites as sheer window shops of an organization. Now is the time to web-enable statistical databases as tools in disseminating statistics.

In developing web-enabled databases, the approach of making use of what is available and what is known proved effective for the NSCB. When other government offices were still developing their websites, the NSCB has already operationalized web-enabled databases. But this is not to say that the NSCB has stopped taking proactive steps in exploring other technologies to better serve the data needs of its stakeholders.

The Internet will continue to evolve into mainstream. As a result, the amount of information on the Web will continue to grow. Now is the time for GSOs to start disseminating statistics online so as not to be left behind.

REFERENCES

- [1] Louis Boucher, "Laying the foundation for electronic dissemination: Statistic's Canada's output database, CANSIM II," Paper presented as part of the Federal Committee on Statistical Methodology Research Conference, Washington DC, 14-16 December 2001.
- [2] United Nations (1994), Fundamental Principles of Official Statistics.
- [3] National Statistical Coordination Board Resolution Numbered 10, "Adoption and Implementation of the Government Statistics Accessibility Program (G-SAP) ," 29 October 1997.
- [4] National Statistical Coordination Board Resolution Numbered 8, "Adoption of the General Standards on Statistical Information Dissemination," 18 March 1999.
- [5] LJ Haravu (1999), "Web-enabled databases: an overview of technology and problems," Paper presented at ITT 1999: Towards Information Content for Global Competitiveness, Hyderabad, India, 16-19 November 1999.
- [6] United Nations Statistical Commission/United Nations Economic Commission for Europe, "Best Practices in Designing Websites for Dissemination of Statistics," Conference of European Statisticians on Methodological Material, Geneva, 2001.
- [7] Statistics Sweden, PC-Axis Family Software: <http://www.pc-axis.scb.se>. (02 Aug. 2004).
- [8] Peter Go, "Adding a database to your website," CGI City Articles, <http://www.ichthus.net/CGI-City/articles/Adding.A.Database.shtml>, (31 July. 2004).
- [9] "The role of relational database systems in survey and statistical processing," <http://www.sasc.co.uk/Guides/srdbms.htm>. (28 Jul 2004).
- [10] Executive Order numbered 121, Signed by President Corazon C. Aquino on January 30, 1987.
- [11] Neil Sutton (2002), "Bill Gates ties Microsoft's future to the XML bandwagon," Computer Dealer News, <http://www.itbusiness.ca/index.asp?theaction=61&sid=49873>, (03 Aug 2004)
- [12] Dare Obasanjo (2003), "Understanding XML," <http://msdn.microsoft.com/XML/Understanding/default.aspx?pull=/library/en-us/dnxml/html/understxml.asp>. (03 Sept. 2004).
- [13] National Statistical Coordination Board (2004). "NSCB maps the country's poorest provinces," <http://www.nscb.gov.ph/poverty/2000/mapping/default.asp>. (04 Sept. 2004).