

# Statistical Information Databases on Society and Economics on the Internet

*Shifra Har and Reuven Graf*

Central Bureau of Statistics, Government of Israel.

## 1. Background

The National Committee on Information and Computer Communications Infrastructures has selected a research proposal to conduct a sample project to make the databases of the Central Bureau of Statistics (CBS) available to the public via the information superhighway. The CBS databases contain a great deal of statistical information gathered over decades about the population, migration, tourism, the economy and economic branches, prices, labour and wages, health, and other topics. This project should assist the CBS in fulfilling its obligations as the agency responsible for publishing official statistics, while meeting world standards set by international organizations such as the International Monetary Fund.

From the beginning, the CBS management has understood the importance of the project, which will cause a leap upward in quality and a vast improvement in the timing of publication of CBS data. For this purpose, a steering team was appointed that includes a management representative, a representative of the Database Division, and a representative of the Information Services Division. These appointments made it clear that information and technology alone are not enough for creating an up-to-date, reliable site; only integration of all components, with backing and support from management, will make it possible for the project to succeed.

The team studied the issue by visiting the Web sites of statistical bureau around the world and reading large quantities of professional literature. As a result, several issues were raised that had to be resolved before the site could be established.

## 2. Budget

The CBS received a grant to establish the site, but this budget did not cover all the start-up costs. The remainder of the funds for establishment, ongoing maintenance, and day-to-day updating came from the CBS budget.

## 3. Selection of Information to Be Posted on the Web

3.1 The CBS has accumulated 40 GB of statistical data, arranged in several computer platforms.

3.2 In addition, the CBS has textual information describing the series, their magnitudes, data-collection methods, frequency of data collection, sampling methods, mention of unusual problems with the data, and more. This information should be posted together with the relevant series (metadata).

3.3 From this extensive store of information it was necessary to select the preferred data to be posted on the new medium, while preserving information security and statistical secrecy.

## **4. Selection of the Technology and the Language of the Site**

### **4.1 Technology**

Three main factors influenced the selection of the technology: the tools available to the organization, the limits posed by timetables, and information security and statistical secrecy.

The Database Division has an Informix database. For the sake of compatibility, it was decided to retain the same platform.

Information security and preservation of statistical secrecy are fundamental to the CBS, and therefore it was decided that two networks would be constructed (double infrastructure)—the Intranet and the Internet—with no physical link between the two. (The double network also serves as a backup means.)

### **4.2 Hebrew on the site**

The CBS chose the Logical technology for the site. In a discussion with the Science Ministry's steering team it was agreed to switch to Visual technology until a binding standard is set. This decision was made on the grounds that academic institutions such as the Hebrew University tend to post information using this technology.

For the most part the data are posted on the site in two languages, Hebrew and English, as in all CBS publications. This bilingual site doubles the number of pages. Each page is published in both languages and users can move from one language to the other.

## **5. Professional Project Personnel**

The steering team was assigned the task of deciding who should carry out the project. The team considered the following options:

1. Training employees from within the organization;
2. Using people from outside (outsourcing);
3. A combination of the two.

It was decided that the third option is the most appropriate for the needs of the CBS, because of the desire to be completely independent in the long run and not to be absolutely dependent on outsiders. Adoption of this track made it necessary to train CBS employees quickly and to have CBS employees work together with outside people on constructing the initial site.

## **6. Main Milestones in Setting Up the Site and the Lessons Learned**

The milestones were predetermined when the grant application was submitted.

### **6.1 Construction of the site through an outside service provider**

The goal was to set up a site quickly in order to conduct a preliminary study of the issue and to increase public awareness of the existence of a CBS Web site.

*The following problems were encountered:*

- At this point, the team learned the meaning of ‘unbearable dependency’ on outsiders. It is extremely difficult to make ad-hoc changes—both updating the pages and adding pages. For example : the CBS wants to provide special services regarding the Consumer Price Index, but this service could not be provided due to outsiders.
- Another problem was that users had trouble finding the CBS site because the address was changed in mid-project from the service provider's address to the permanent address. The service provider was not helpful in setting up a link from the temporary site to the permanent site.

*Lessons learned:*

- The decision to have an outside firm run the site should depend on the number of update pages needed on a regular basis and the frequency at which topics are added to the site.
- CBS employees should be trained in relevant issues even if the site is constructed by an outside firm.

### **6.2 Construction of the site on CBS premises**

*Goals:*

- Setting up a site on CBS premises with appropriate technology, including static pages only.
- Creating a template of the contents and form of the Web pages. The team intended to create a balance between the aesthetic form and a limit to the time it takes for pages to come up.

*Problems encountered:*

- Selection and integration of the hardware and software.
- Lack of information and lack of practical experience in the problematic field of acquiring truly professional knowledge from outside sources.
- Lack of knowledge and experience in the CBS in professional graphic design in the Internet environment.
- Conversion of data to HTML. The organization has made a strategic decision to use Office and SAS software as the work tools. The need was stressed to find tools for automatic conversion between the organization’s work tools and HTML. So far no satisfactory solution has been found for converting Excel tables containing Hebrew to HTML.

*Lessons and solutions:*

- Outside firms should be used for graphic design.
- Most CBS publications are prepared in Quattro (for DOS), with a tendency to switch to programming in Excel (Windows). In the absence of tools for converting to HTML, an attempt was made to introduce a system for converting files as needed. From time to time programs that supposedly perform the required actions are considered and evaluated.

### **6.3 Development of the static site into a dynamic site**

*Goals:*

Expanding the site to the dynamic format: The CBS intends to publish tens of thousands of time series, and it is not feasible to create HTML pages for each one. HTML pages have to be planned in such a way that they will serve as a uniform template for all the series. This expansion includes the introduction of a database and a search engine that allow for queries in various forms in real time. We have also begun publishing press releases, monthly bulletins, and annual reports.

*Problems:*

- Constructing the interface (CGI) between the database and the HTML pages.
- Difficulty integrating text, tables, and graphs in press releases.
- Due to the lack of standards for submitting the material for publication, each publication must be adapted manually. The standards for publication in the new medium are different from the standards for publication on paper. From now on, precise definitions are needed for preparing files while adhering to the rest of the rules used for print publications.

*Lessons and solutions:*

- Assistance from outside firms for constructing the interface.
- Training employees in the relevant programs.
- Writing procedures for submission of material for posting on the Internet, in addition to editing rules.

## **7. Changing the Work Culture**

Setting up the CBS site has required a change in the CBS work culture. The organization's site is not supposed to depend on the whims of a small group of workers who deal only with the Internet. It is a system that involves all CBS employees, from the managers, who are in charge of data quality and the publication schedule, to the last worker whose task it is to gather and process the data. As a result, a slow process of adaptation to the new requirements resulting from the innovative technology is beginning.

## **8. The Present Situation**

The CBS maintains the site independently in terms of operating systems, regular updates, and development of new topics. Two CBS committees have been formed to improve the site, make it more efficient, and most importantly, to maintain the high level of motivation and consideration of regular activities.

- CBS Internet Committee: This committee determines what topics will be posted on the Web site and monitors actual implementation.
- Subcommittee of the CBS Internet Committee: The subcommittee is supposed to recommend the graphic design of the site, the form in which the data are presented, and a general redesign of the site.

## **9. Information Security**

The CBS has a committee to set information-security policy. This committee has determined that CBS employees should not be permitted to hook up to the Internet through the computers that are connected to the Intranet. Consequently, a separate physical network has been established for feeding data to the Internet. In addition, every division has at least one *designated* computer for Internet hook-ups, which connect via modem to the company that provides Internet services.

## **10. Assimilation**

Publishing via the superhighway is totally different from publishing through hard copy, therefore all the CBS employees have to learn the new medium and use it properly and freely. The first group of employees has received several hours' guidance from an outside firm. This guidance was not adapted to the CBS's needs, and therefore two CBS employees who specialize in software assimilation were trained in appropriate courses. A syllabus for an internal course was also designed.

The CBS intends to declare 1998 the year of Internet assimilation. This year most of the employees will take a 16-hour course consisting of four four-hour sessions, with specific topics covered in each session. Each employee will choose the most convenient of three class times: 8 a.m.–12 noon, 12 noon–4 p.m., or 4–8 p.m.

## **11. Success Factors**

Among the factors that have contributed to the advancement of the project are the following:

- Support from management, which put all its weight behind the project;

- The full cooperation of all the members of the professional team;
- Correct planning of the site so that it would present the large amount of diverse statistical information available at the CBS.

## **12. Conclusion**

About 80,000 hits per month have been counted on the CBS site. In addition, we receive dozens of thank-you letters and compliments from users. Some of the letters include recommendations for improving the site. We try to relate to these letters positively.

All this seems to indicate that the project is a success.