

Session 4

E-Government and New Dissemination Paradigms

Introductory Remarks
Lawrence A. Greenfeld, Director
Bureau of Justice Statistics

Good afternoon and welcome to Session 4 entitled E-Government and New Dissemination Paradigms. We have some excellent speakers who are going to talk about Stats Canada and what they are doing to improve the distribution of information over the Internet and some speakers on FEDSTATS.

There are a variety of challenges confronting US Federal statistical agencies as greater centralization of computing authority and control occurs. Many of us must now work with parent agencies with CIO's and reason with them about mechanisms to consider which focus on how data can be insulated and protected, content controlled and managed by the stats office, and presentation of data offered in a manner consistent with the needs to assure privacy to respondents and to guard against pre-release. This is not easy as CIO's are now charged with taking control of entire Departments computing resources and often the budgets associated with both hardware and software acquisition. For a small stats agency in particular, protecting our core values about data and its handling for public use and our desire to insure the proper usability of what we produce is an emerging challenge. I have little doubt that soon all stats agencies will begin to face such issues.

Although this section is primarily focused on the users of statistical data, it is important to think about the use of the Internet as a data collection tool, particularly for administrative data from agencies. BJS has been gradually migrating certain collections from mail-out to web-based. This has created a set of interesting challenges with respect to respondent-identification and the ability to edit previously submitted data.

At BJS, we make all of our pubs and datasets for public use. We have nearly 4,000 staff-produced spreadsheets of data which are cross-referenced to relevant reports and datasets. Every graphic on our website easily converts to a spreadsheet for download with just a couple of clicks. In addition, we have a wide variety of datasets with which customers can directly interact to produce tabulations and cross-tabulations. Any number published by BJS should be capable of being reproduced by the public. What a dramatic change from the days when customers were bound by what was in books and limited to the use of whatever data was printed on a page of a Federal document. Having been in my field now for over 30 years, the extent to which we have liberated and democratized statistical information and the data used in computations in the last few years is absolutely awesome. It is our job to insure that those managing computing resources do not interfere with this kind of progress simply to promote uniformity within Departments. Maintaining the vitality, creativity, and exuberance in stats agencies about sharing their policy-relevant and publicly-funded information collections is our most important challenge and responsibility.

I am very appreciative for the work of Cathy Dippo from Bureau of Labor Statistics who organized this session and to our speakers and discussant.

We will begin the session with David Roy from Statistics Canada who will present his thought about “How the Internet is transforming Client and Respondent Relationships at Statistics Canada.”

How the Internet is Transforming Client and Respondent Relationships at Statistics Canada

David Roy, Director Marketing
Statistics Canada

Introduction

I'd like to begin by thanking Cathryn Diplo of the Bureau of Labour Statistics for inviting Statistics Canada to take part in a discussion of E- Government and New Dissemination Paradigms.

Like other national statistics offices, (NSO), Statistics Canada's use of the on-line channel began well before the creation of a Canadian E-Government initiative. The Internet is a natural fit for the business of a national statistical office and our user communities were among its early adopters - so we were well advanced when Canada's E-Government initiative began in 1999.

Also, for many years there has been a sharing of information on dissemination and marketing strategies among NSOs. Most recently there was an excellent meeting held in early September in Annapolis involving sixteen countries, that was organized by John Kavaliunas and Colleen Flannery of the USCB. Statistics Canada has benefited greatly from these meetings and to some extent I think there is a great commonality in the dissemination strategies – emerging paradigms - of many of the participating countries because we have been sharing information on best practices for many years.

In my presentation I'll begin by giving some context to our activities by briefly describing the Canadian E-Government initiative. Then I'll give a summary of some of our activities – in dissemination and other key services and how our client relationships are being transformed -and finally I'll provide a couple of information sources on E-Government that you might find helpful.

Canada's Government On-Line Initiative

What I'm going to be talking about is Statistics Canada's activities that are related to a program called Government On-Line, (GOL). This Federal Government-wide initiative includes the delivery of all appropriate information and services on-line as well as a Service Improvement Initiative. The latter is essentially the application of marketing principles to government activities – understanding client needs, developing appropriate products and service standards and monitoring performance and client satisfaction. In Canada the GOL initiative has been strongly client focused. That's one new paradigm in itself.

The GOL initiative was launched in late 1999 with the goal of having 'all' information and services accessible online by 2004. This target has recently been extended to 2005 - in part because investment funds have not been as available since last September – and now the target only applies to services for which there is 'sufficient demand' to warrant the development of an online delivery option.

Of course, E-Government has a broader context than the service delivery focus of GOL, and it incorporates a more fundamental re-examination of our government and democratic processes. Statistics Canada participates in such an initiative and I'll say a few words about it at the end of the presentation.

The Government On-Line, (GOL), initiative was a high priority of the government from the outset and among the benefits frequently mentioned by the Prime Minister and Cabinet Ministers were:

- Playing a leadership role in creating the infrastructure and practices to encourage a wider use of the Internet among businesses,
- More efficient service delivery - a high priority of citizens because of its potential to lead to tax reductions,
- Higher approval ratings of the Federal Government by Canadians in public opinion and satisfaction surveys, and
- National unity through the perceived high value of Federal Services

Several parallel initiatives were conducted to increase connectivity of schools and communities, establish an appropriate technical infrastructure, make available cultural content and provide an environment conducive to e-transactions.

Government On-Line Objectives

Here you see some of the same ideas expressed in the objectives that were set for the GOL initiative

- Stronger relationships with clients and better service
 - Interact with more clients where they live and work
- A catalyst for electronic commerce
 - Help meet the Prime Minister's challenge to capture a 5% share of world of e-commerce by the year 2003

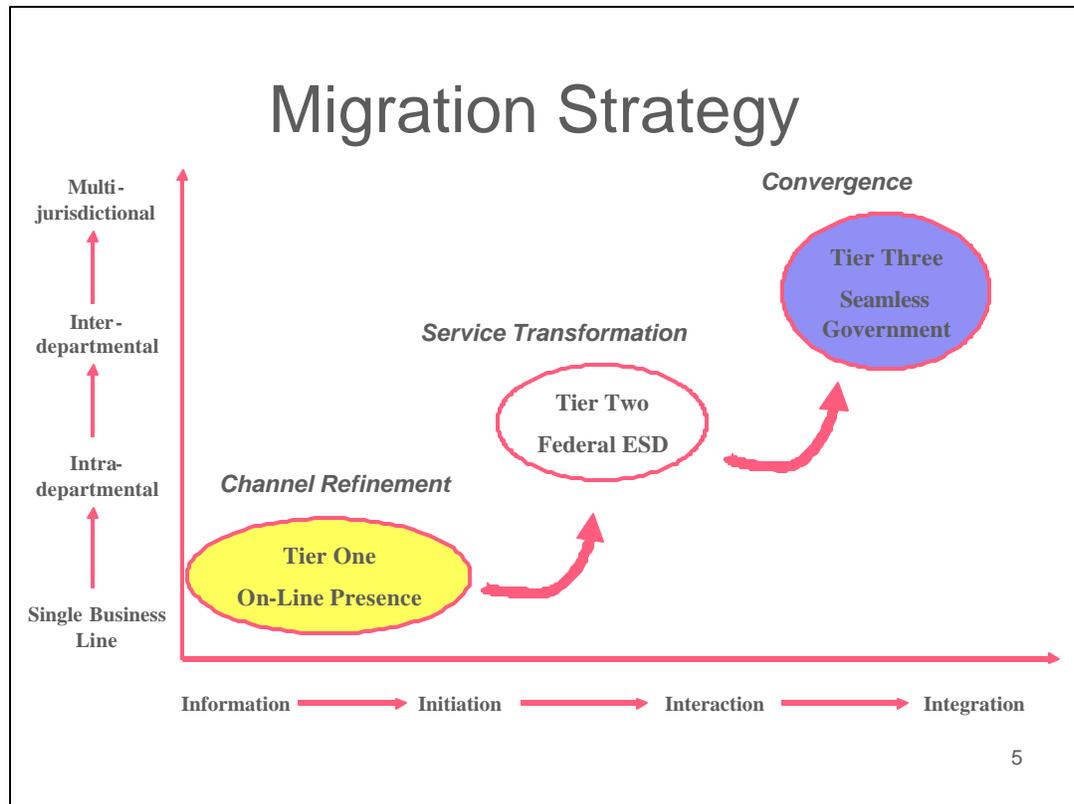
The focus on improved relationships was motivated by some early research conducted by our Treasury Board showing that Canadians' satisfaction ratings for most public sector organizations were well below the ratings of private sector services. There was concern that online government services would be judged by the service standards and client service orientation of private sector organizations in the delivery of E-services and so a very strong client orientation for GOL was adopted.

Other research among business showed that Canadian businesses rated the Internet far lower as a priority than US businesses and there was a concern that Canada would not get the share of global e-commerce that would ensure the competitiveness of our economy in world markets.

One of the most significant findings of this research for Statistics Canada was that Canadians placed ‘completing government surveys and questionnaires’ as their second most important use of the online channel after tax filing.

Phases of Government On-Line Implementation

This graphic illustrates a planned phased approach to the GOL initiative which would take advantage of lessons learned along the way and apply them to subsequent activities.



The horizontal axis denotes the type of on-line interaction and the vertical axis denotes integration among service delivery agents.

Tier One was meant to establish the federal government’s on-line presence by putting key departmental and program information on-line and making it accessible either directly through a department site or through a revamped Government of Canada Portal. The target for this phase was December 31, 2000 and generally it was met. Statistics Canada had achieved this target about two years before that date.

Tier Two represents a significant step-up from Tier One. This second tier is essentially the delivery of end-to-end secure ‘transactions’ for all key programs and services by December 2005. For Statistics Canada, transactions also include data collection activities. The words

‘service transformation’ characterize this stage – the fundamental redesign of service delivery from a client needs perspective to capitalize on the inherent benefits of the Internet.

Tier Three involves inter-jurisdictional service delivery and a variety of pilot projects are already underway to foster partnerships and the cross-jurisdictional integration of services from different levels of government - another new paradigm. I’ll be mentioning some pilot projects that Statistics Canada has been involved in.

Departmental ‘Key Services’

Each Department/Agency was required to developed a GOL plan for each of its ‘Key Services’ and for Statistics Canada these are the three key Service we identified.

Collection: Collecting data from individual citizens, households, institutions and businesses as part of census and survey programs undertaken by Statistics Canada.

Communications & Dissemination: Serving information users via the news media, with standard products, the Internet, custom services and our distributor network with outputs of statistical programs.

Stakeholder Relationships: Managing relationships with key interest groups and constituencies with whom Statistics Canada has strategic alliances, e.g. associations, provincial agencies, education, data researchers.

You’ll notice that we did not identify programs such as ‘Census’ or ‘National Accounts’ as key services. The functional approach we chose provides both a highly simplified way of describing all of the Agency’s client relationships and also an effective way to plan and implement our online activities in an integrated way.

While only one of these key services has information dissemination as its principal focus, the other two – collection and stakeholder relationships – have strong dissemination components as well.

As a starting point we developed a strategy paper for each key service on the opportunities that the on-line channel presented for each service’s constituency. These formed the basis for a corporate plan that we produced for the Treasury Board and which we continue to update. The corporate plan is a template based document which allows Treasury Board to compile an overall government plan.

The balance of my presentation will be about the ‘service transformations’ occurring in each of these key services and how they are fundamentally changing the relationships between Statistics Canada and its clients. In the process, a number of new paradigms should become obvious.

The ‘Communications and Dissemination’ Key Service

Our Communications and Dissemination key service has already achieved the GOL Tier Two 2005 goal of service transformation – a fundamental re-engineering of our dissemination services from a client perspective. The key elements of this transformation include:

A Corporate Data Warehouse: At the heart of our dissemination strategy is a corporate data warehouse - CANSIM II - which includes virtually all of Statistics Canada’s published information and is the source from which much of the other content of our web site is dynamically updated. Since its launch two years ago the number of time series has grown from 800,000 to approximately 11 million.

All Publications Available On-Line: With a small number of exceptions all tabular and analytical publications, methodology papers, user guides and research papers are available on-line – primarily in PDF.

Official Release On-Line: The DAILY, our official release publication for data and products, has over 7,000 subscribers and in the near future subscriptions to it will be available for 28 ‘themes’ – health, employment etc, - as a first step towards more specific personalization.

Daily Updates: Over 450 National/Provincial tables in the Canadian Statistics module are updated on a daily basis and most are linked to Statistical Data Documentation, (Meta Data).

Community Profiles: Profiles of 6,000 Canadian communities now include Census and Health information and other social data will be added.

On-Line Catalogue: There is a comprehensive Online Catalogue and products descriptions are linked to our Integrated Meta Data Base which describes the statistical survey where the information originates and the underlying concepts.

E-Commerce: The site has included E-Commerce since 1997 and total revenues in 2001-02 were approximately equal to the cost of maintaining the site.

Integration of Service Delivery Channels: A ‘Contact Us’ button is included on almost all site pages which provides users with a range of access options including toll free telephone and email. The latter are received by our Advisory Services group and answered directly or routed to the appropriate subject matter or other contact for direct response. Last year over 30,000 email messages were answered and are themselves an excellent source of client research on information needs and navigational issues. Standards of service for all service channels including custom services are published on the site.

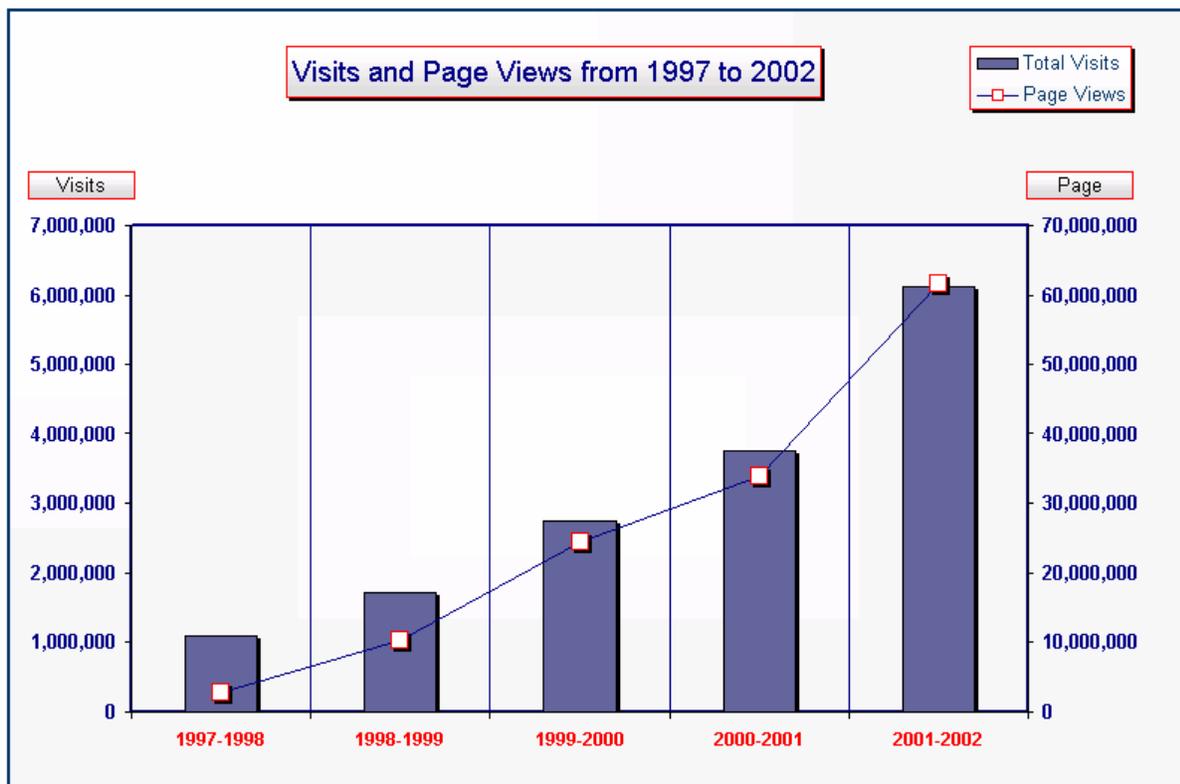
Common Look and Feel: All Federal Government sites follow a set of strict guidelines that give them a common look and feel. This benefits users who develop a familiarity with the type of information located in each area of pages on Federal Government sites, (common toolbars, navigation features, etc), and contributes to ease of use for visitors. Although many sites initially

resented the limitations these standards placed on creativity and their ability to have a unique look, most would agree there is still sufficient latitude for individual ‘branding’ and the users benefit from the common design elements of government sites and within sites.

Client Focused Site Development: The development of our site has been guided by research with visitors since its inception. We have conducted a number of online surveys with site visitors, focus testing, observational research and testing of particular products by closed user groups.

Growth in Internet Traffic

Site traffic has grown steadily – by over 50 percent in 2001-02 – with over 6 million visits last year. The following chart illustrates the pattern of growth in visits and page views we’ve experienced. In part the growth can be explained by the general increase in Internet use among information users but there are a number of other reasons. We’ve promoted the republishing of content from our site with the condition that those doing so provide a link back to www.statcan.ca. Today there are over 10,000 pages from 3,000 sites indexed in AltaVista that link to our site. We’ve also invested heavily in registering our pages with the most widely used search engines so we come up high in search results. And we do a significant number of other awareness creating activities as well.



'E-Clusters' – A citizen Centered Approach to Government Services

E-Clusters are one of the core elements of the Government On-Line initiative as they allow citizens to find information and services without having to understand the structure of government. E-Clusters are single entry points to information and services on a common theme which are provided by a number of Departments and Agencies and they are accessed through the Home Page of the Canada site, www.canada.ca.

Statistics Canada participated in the development of the E-Cluster concept, particularly in the market research to determine the categories of information and services sought by three major client groups; Canadian Citizens, Businesses and International Visitors. The Canada site with these three 'gateways' was launched in January 2001 with 35 E-Clusters.

The following graphic illustrates the concept.



Statistics Canada will play the lead role in developing two E-Clusters:

- **Economy** – which involves three partner departments and the Bank of Canada provides information on Canada's economy in relation to other countries and is designed for citizens rather than specialists in this area, and
- **Business Information and Statistics** – which involved ten partner- departments and is aimed at small and medium size enterprises to improve the success rate of new start-ups and enhance the international competitiveness of Canadian Business.

E-Clusters in Action

The home page of the Canada site is found at www.canada.gc.ca which provides access to thirty five E-Clusters grouped in three categories or Gateways:

- Services for Canadians
- Services for Non-Canadians
- Services for Canadian Business

Today approximately 6 percent of our site traffic comes through these portals. This compares with 34 percent from search engines.

If you click on [Services for Canadians](#) it will bring you to a listing of topics organized by Subjects and Audiences. Because of the range of information Statistics Canada provides, we expect that eventually almost all E-Clusters will have links to our site which will provide many more pathways to our content.

If you click on [Economy](#) from the Subjects list you reach the home page which Statistics Canada created in partnership with four other Agencies, Industry Canada, Foreign Affairs and International Trade, Agriculture Canada, Finance Canada and the Bank of Canada.

If you visit this site you'll notice that the information created for this site is designed to inform the average Canadian about the performance of the economy. More typically our users are economists and policy planners but this site is targeted to a broader audience and provides a great deal of information on economic concepts as well as a time line of key economic events.

The home page of the Economy E-Cluster includes a number of key economic indicators. These indicators are updated dynamically from Statistics Canada's corporate data warehouse, CANSIM, whenever it is updated. This is the first table that is dynamically updated outside Statistics Canada's web site but many more are anticipated.

If you click on [Current Economy](#) you'll get an idea of the range of information available from the partner Agencies. Among other information it includes:

- A Quarterly newsletter from Finance Canada, 'The Economy in Brief'.
- Monthly Analysis from Industry Canada which provides more detail on trends within industries,
- The Statistics Canada Daily links to the home page of www.statcan.ca
- Other headings such as Families and Workers have links to Canadian Statistics tables on www.statcan.ca.

We will not promote the site until we have completed the first visitors research study. We have, however, registered it with most of the major search engines and it is listed first when the search term 'Canadian Economy' is used in Google.

The site is expected to get a great deal of visibility when major economic announcements are made such as a Federal budget or at pivotal points in the performance of the economy, e.g. entering a recession or a recovery and will provide information on these topics for the average citizen.

A Business Portal within Our Site

Research with visitors to www.statcan.ca over time has shown some under-representation of business information users. Focus testing has revealed a preference by businesses for a focal point or portal providing links to information of interest to them on the Statistics Canada site. The GOL initiative presented Statistics Canada with the opportunity to create a Business Data portal which can be accessed from the left side tool bar on our home page or from the Business Gateway on the Canada site.

Further focus testing during the development of the [Business Data](#) page revealed that users wanted both a thematic access to information, Browse our Comprehensive List of Business Topics, as well as organization of content around key business activities, e.g. Obtain Trade Data for Canada and Abroad.

The page was launched in October 2001 and has surpassed the traffic forecast. We have conducted some research with site visitors and they have given the concept favourable ratings but want more content added to the site, particularly organization of information by industry so they can compare their firm's performance to their industry and geographic comparisons.

The page also includes the top 10 business information products as well as the same key indicators that appear on the homepage of www.statcan.ca. Visitors also indicate they would like to see indicators more directly related to business activity in this area.

We will continue to develop Business Data with additional content and will consider adding links to sites of other Federal Agencies with relevant content and possibly to provincial sites.

The Data Collection Key Service

Our Data Collection key service is at a much more preliminary stage of development than Communications and Dissemination. While there have been some early business survey experiments they were not truly online activities and required downloading of an application or questionnaire, completion off-line, encryption and then transmission. In general take-up rates were low.

There was also a small test conducted in two municipalities within the 2001 Census and take up was also low for similar reasons.

Statistics Canada's approach can best be described as cautious because of the many unknowns associated with electronic data reporting. Certainly, in the initial stages, it will be an additional channel creating the uncertain impact of mixed methodologies on data quality.

We have obtained funding from our Treasury Board to create an online response option for 60 business surveys and one household survey by 2005. The surveys selected are mainly monthly

and quarterly surveys with relatively few questions and respondent communities that are highly connected to the internet therefore offering the greatest potential to maximize take-up rates. The Electronic Data Reporting project will also create a Personalized Reporting site for a small number of very large businesses to provide them with information on the surveys they will be asked to complete, assist them with managing online reporting and provide a focal point for respondent support.

The 2006 Census will draw on lessons learned from these initiatives and will be implementing an online response option throughout Canada. Census management have set an operational target of 25 percent response for online response.

Respondent Research

We recently conducted a study of households and business, which had just completed a Statistics Canada survey, to better understand respondents' readiness and willingness to use online response. Combined, about 85% of respondents had Internet access at the location where they completed their survey. About 80% of those who had Internet access said they 'definitely' or 'probably' would have used an online option to complete their most recent survey if it had been available to them.

They would only have used online, however, if it had been more convenient, more efficient and they were assured that there could be no unauthorized access to their information. Security of their information was the most important decision factor.

Online response is not a question of 'if' but 'when'. Certainly businesses who are using the Internet to manage supply chains for reasons of efficiency and who are able to do e-filing of tax returns will have growing expectations that survey questionnaires can be completed online as well. Households will value both the convenience dimension of online as well as the improved security online should eventually offer.

Earlier studies of factors that would motivate respondents to participate in surveys - particularly businesses - included access to the survey results. This expectation is expected to increase with the use of online reporting. Providing a business with a profile of how the firm compares to its industry and with access to other relevant data useful to its decision-making will not only motivate participation in surveys but should also improve the quality of response. Other timeliness and quality improvements are possible if respondents can link survey templates on personalized web pages with their own electronic information systems.

These features of online data reporting should present the opportunity for Statistics Canada to transform survey participation from an onerous activity - based on legal obligation - to one that is advantageous for respondents. Our goal must be to find that new 'value proposition'.

The Stakeholder Relations Key Service

The Stakeholder Relations key service could be included as part of our other two key services, Communication and Dissemination or Data Collection. However, we decided it would be useful to identify a category of activities that we conduct in order to improve relationships with the interest groups and constituencies with whom Statistics Canada has strategic relationships, e.g. business associations, provincial agencies, the education and researcher sectors. The following are brief descriptions of some typical initiatives.

Education Community Liaison Program: Statistics Canada has made a strong commitment to the use of Canadian information and data in Canadian classrooms and academic research. The Educational Community Liaison Program includes the development of a Learning Resources module on www.statcan.ca and the creation of an Education Account Executive position in each of our regional offices. These resources work with teachers and schools, educational publishers, faculties of education that train teachers, and with school boards and ministries of education to encourage the use of statistics Canada data in teaching activities. The majority of this information, including teacher developed lesson plans and curriculum guides, is provided via the Learning Resources module on our site.

The Data Liberation Initiative: The Data Liberation Initiative was created to provide access to all of Statistics Canada's published electronic databases and public use micro data files for research and teaching purposes in Canadian universities. All have now joined the program at a fee which covers its cost. Electronic files are distributed to data librarians via the Internet and a very active user community has evolved sharing information on the holdings, again via list serves and other Internet communications.

Pilot Inter-jurisdictional Projects: Several pilot projects were funded through the GOL initiative which have been completed and are now being evaluated to assess the potential to apply lessons learned in other program areas.

- In conjunction with Health Canada, online training materials were developed for local health professionals to support the use of data for local decision making;
- Synthetic micro data files of education data were made available to researchers via the Internet which allowed them to specify tabulation requests from unpublished data, to submit them and have confidentiality screened results returned online in order to minimize the normal time requirement; and
- A secure communications channel was established to collect justice information and enable pre-release reviews by the justice community including local police departments.

The findings of the pilot studies will be available in the Fall of 2002.

Increasingly the Internet will be used to manage relationships with key stakeholders.

Census Consultations: For the 2006 Census we have planned a two-stage process to simultaneously discuss Census content and outputs with data users. In the first phase we will provide information materials through traditional channels and offer a range of options to

provide in depth recommendations and feedback as a second phase. There will be at least one pilot test of obtaining this input online through a 2006 Census consultation web site.

On-Line Advisory Committees: Statistics Canada has 22 subject matter Advisory Committees and the National Statistic Council which guide our programs. Later this Fall Statistics Canada's GOL working group we will contact the secretaries and chairpersons of these committees to identify a small number to test online consultative processes using extranets and closed user groups. Again, the GOL initiative has developed some standardized approaches and tools for these types of consultative activities and we will use these in the test.

Respondent Relations and Research: A critical element of the success of the electronic data reporting project, (EDR), will be the provision of information to prospective respondents related to their key concerns such as security and confidentiality, and the convenience and efficiency of the process. As well respondents must have an online single point of access to support, links to survey results and other data related to their interests. Research on respondent relations in support of EDR will be conducted in conjunction with the 11 surveys which will begin to offer an online response option later this Fall.

Dynamic Updating of Other Sites: there has already been a large increase in the number of organizations wanting to republish Statistics Canada data on their sites and it will continue to grow. Tables on our site are dynamically updated whenever the CANSIM II database is updated. The Key Indicators table on the Economy ECluster is the first instance of this process being used for a table on another site. This process will be actively promoted as it ensures that wherever STC data appear they are consistent and will be accompanied by a link back to www.statcan.ca.

Recruitment: An Employment Opportunities module has been added to www.statcan.ca to provide information on the full range of recruitment initiatives which generate the majority of our new professional, technical and social science support staff. This module will evolve to provide more of the primary screening of applicants to streamline the process.

New Data Dissemination Paradigms

What are the new paradigms for National Statistical Offices in an E-Government world?

One Stop Data Shopping: First our web sites must be comprehensive repositories - enabling information users to access all of our published data online - and our research shows that effective search is the critical factor in successful access to content and finding the information sought is the key determinant of visitor satisfaction with their site experience. As well all information must be linked to the underlying meta information for users to fully understand the concepts and the processes through which it was created to use it effectively.

Dynamic 'Database Publishing': Because of the huge amounts of information available from our sites their overall integrity must be ensured by updating processes that, to the extent possible, minimize human intervention. Otherwise the cost of maintaining a comprehensive site is prohibitive. Today most NSO sites are driven by linked databases, (multi-dimensional tables, meta information, analytical text, catalogues, etc), which allow data to be presented in a variety

of formats. When a database is updated it automatically triggers the updating of information throughout the site so there is consistency.

Personalization: To build effective relationships with site visitors, they need to be able to identify which topics are of interest to them and be notified of the availability of newly released information related to their interests or have it automatically sent to them. Demand for personalized services will grow quickly.

Single Points of Entry: Information users expect Government Portals or Gateways to provide access to information and services from many sources without having to understand the structure of government. The E-Clusters do this effectively across Departments, and within departmental sites users expect to be able to search thematically and to have other integrative mechanisms such as our Community Profiles and our Business Data modules to integrate content across statistical programs.

Branding: Our data will be republished, so we must provide the tools for other sites to provide appropriate sourcing information and to create links back to our sites which are more comprehensive and current. If we are not identified as the source of our data, respondents will not see the value in participating in our surveys. We also need to do more public opinion research to understand more about how households and business perceive our brand – to know more about our ‘brand equity’ to help us develop more effective communications programs.

Respondents are Clients Too: We need to use information outputs to create a new ‘value proposition’ for survey respondents to motivate them to provide high quality input to our surveys. We must apply the same marketing principles to electronic data reporting we have applied to our dissemination activities so we re-engineer them from a client perspective.

Online Partnerships: Build online partnerships with key stakeholders - groups that play roles that sustain our core mission. Closed user groups, Extranets and online consultative processes help to build relationships.

Apply Marketing Principles: E-Government is a client focused process. Know your clients, listen to their messages and act on them.

Don’t Re-Invent the Wheel: And finally, build relationships with your international colleagues and share best practices. There are likely many people who are also working on ‘your great idea’. Develop a network – you may even get to travel.

E-Government Information Sources

Although the title of this session and my paper refer to E-Government, most of what I have talked about is really the use of the Internet to deliver our organization’s information and services – what we call Government On-Line. This is occurring in all of the developed countries in the world and I’m pleased that we’re sharing our experiences much as we have with output databases.

E-Government is a much broader concept that is also being studied in democratic countries around the world. New communications and information technologies make many of our existing institutions and their focus irrelevant as connectivity erases organizational boundaries and even national boundaries. It also permits a much broader participation in policy development processes and increased transparency and accountability in government.

Statistics Canada has participated in funding an initiative called Crossing Boundaries which explores these opportunities, in part because of the key role played by information in the policy development process and because there is a growing perception that information is an essential public resource in this new paradigm.

We have had one presentation for our senior management community on the first report, 'Realigning Governance: From E-Government to E-Democracy. If this is a topic of interest, you can register to receive their newsletter URL and any of their reports at www.crossingboundaries.ca .

Finally, I want to mention a report that was prepared by Andersen Consulting called the Accenture Report. It is their third annual assessment of E-Government in 23 countries. Their assessment model includes ratings for 'service maturity' and 'customer relations management' which are combined to give an overall rating for each country. Service maturity measures the breadth of services available online plus degree of completeness. Customer relationship management measures the level of service sophistication. An electronic copy of the report can be obtained at www.accenture.com

Our Treasury Board has adopted this model to assess the performance of Canadian Federal departments and agencies in E-Government.

Thank you for the opportunity to participate in the conference. Please contact me by email if you would like further information on any of the topics in this presentation.

**FedStats—Statistical Information Dissemination in the 21st Century—
The Next Generation**

Valerie Gregg
FedStats Interagency Task Force

And

Marshall DeBerry
FedStats Program Manager

Preface

Citizen Access to Federal Statistics Scenario 2020⁵

Individuals want access to federal statistical data. They wish to learn, for example, the demographics of different areas (e.g., information about schools, cost of living, recreation), what is going on in business and agriculture, what is driving prices in a particular area, or what to expect with regard to inflation and interest rates.

How far have we come today toward realizing this vision? FedStats provides a single portal for federally collected data sets and for documents based on that data. Data sources and documents are organized topically and geographically across all the federal statistics agencies. In many cases, the available data are constrained, owing to confidentiality protection, but summary information and reports may be available. Still, one cannot make such queries as, How many people will be displaced if an evacuation at the 100-year flood line for Manhattan, Kansas, is required? Or, what would be the economic impact of locating a particular new business in my town?

Imagine asking FedStats the latter question in 2020. This might trigger a series of questions back to the user not only to acquire more details about that business but to learn more about that user: his or her quantitative/scientific literacy and visual/verbal/textual/cognitive abilities. Then, the relevant data, complemented by additional data sources where needed, would be "crunched" with the aid of models and simulations. A response containing the requested information both fully and in user-friendly form would quickly be returned to the individual making the query.

To realize this requires IT innovation on several fronts, such as representation of information, archiving and searching, modeling and simulation, and information integration. Subtle but important issues, such as the underlying integrity of responses, will also become key. For example, when people of varying degrees of quantitative sophistication ask the same basic question, answers must be consistent.

Taking the scenario one step further: imagine being able to get a second opinion. The local chamber of commerce has contracted with a small economic modeling company to give you access to a model that uses a different set of assumptions. Running this model using a portal to the company offered by the chamber, the model accesses the same underlying census and economic data that were used in the government's model. The

⁵ This scenario is paraphrased from Appendix A, "E-Government Scenarios" of the National Academies of Science, National Research Council's Computer Science and Telecommunications Board (CSTB) May 2002 report entitled *Information Technology Research, Innovation and E-Government*. The full report is available on-line at http://books.nap.edu/html/itr_e_gov/.

modeling company's software is able to access the underlying government databases directly, using an application programming interface offered by the government to allow non-government computer programs to analyze the data in new or different ways.

This paper, prepared for the Federal Committee on Statistical Methodology's Statistical Policy Seminar Challenges to the Federal Statistical System in Fostering Access to Statistics "FedStats—Statistical Information Dissemination in the 21st Century—The Next Generation" will provide one perspective on bringing this scenario to fruition.

Introduction

FedStats is a major success story and an exemplar for interagency, multi-sector partnerships. The award-winning website not only exceeds the initial objective as defined in 1995 by the Interagency Council on Statistical Policy (ICSP), it is now rapidly becoming a demonstration environment for new technologies that will enable the entire Federal statistical community, as well as individual agencies, to become a leader in "Electronic-Government", or "E-Gov" implementation.

This paper provides an historical perspective on how FedStats evolved, how FedStats will continue to evolve within the E-Gov context, and the role FedStats will play in near, mid- and long-term statistical information dissemination in the 21st century. FedStats will help lead statistical agency dissemination efforts towards realizing the 2020 Scenario described in the preface.

Background

The United States Federal statistical system is decentralized, with individual agencies having statutory responsibility and authority for statistical activities. Hence, it is difficult for the general public, and even frequent data users such as social science researchers, to know about and to access the entire wealth of information produced by the Federal statistical system. To address these organizational barriers to accessing Federal data, the ICSP (consisting of the agency heads of the 14 largest U.S. statistical agencies), under the leadership of the Chief Statistician of the United States, Katherine K. Wallman, launched *FedStats* in May 1997. Prior to the public launch, the FedStats Interagency Task Force had been working together since the fall of 1995 to design and develop a "One-Stop Shopping" or "Virtual Statistical Agency" for Federal Statistics Website.

This interagency web site <http://www.fedstats.gov/> now serves as the Internet gateway to the full range of official Federal statistical information available to the public from more than 100 U.S. Federal agencies. FedStats provides a centralized set of links to the Internet sites and the subject-matter data that individual agencies maintain and update. The site's primary objective is to help users find the information they need without having to know and understand in advance how the decentralized U.S. Federal statistical system is organized or which agency or agencies may produce the data they are seeking.

From June 1997 through August of 2002 there have been nearly 8 million user visits to the FedStats site, which represents nearly 25.5 million pages served to visitors. User traffic has increased by approximately 60 percent from 2001 to 2002. The user profile represents a wide spectrum of visitors, ranging from private citizens, academic users, the media, policy makers, and visitors from countries outside the United States. Frequently visited sections of the site include the “Topic Links A to Z” section and the MapStats section, which provides a simple “drill down” capability to retrieve statistical information at various levels of United States geography.

The Task Force reports to the ICSP on an annual basis, providing an annual assessment of the previous year, a set of recommended projects for the coming year and a set of resource requirements. Starting in Fiscal Year 1998, the U.S. Bureau of the Census, via interagency agreements with each of the ICSP agencies, is reimbursed annually for supporting the technical design, development, and maintenance of FedStats. The agreement covers the costs of the FedStats Chief Architect, an additional technical FTE and hardware and software. Until this year, the total cost was \$285,000/year (5 largest agencies paying \$30,000 each and the 9 smaller agencies paying \$15,000 each).

The Interagency Task Force continues to upgrade and expand FedStats coverage and access to Federal statistical sources. Additionally, the Federal statistical community is exploring new information technologies and undertaking research projects in collaboration with the National Science Foundation’s (NSF) Digital Government (DG) Research Program to achieve a much broader vision for the future (discussed in more detail in a further section). New technologies and methods being developed as a result of more than 14 NSF DG research grants are helping to guide design and development of the Next Generation of FedStats. (For more information on the DG-FedStats research projects see Appendix II; for more information on the DG Research Program see <http://www.diggov.org>)

Current Features and Capabilities

Over the past five years, FedStats has become “The gateway to statistics from over 100 U.S. Federal agencies”. The current features and capabilities include the following:

Links to statistics

- **Topic links A to Z**—Direct access to statistical data on topics of your choice.
- **MapStats**—Statistical profiles of States, counties, Congressional Districts, and Federal judicial districts (drop down list of states)
- **Statistics by geography from U.S. agencies**—International comparisons, national, State, county, and local.
- **Statistical reference shelf**—Published collections of statistics available online including the Statistical Abstract of the United States.
- **Search**—across agency websites.

Links to statistical agencies

- **Agencies listed alphabetically**—with descriptions of the statistics they provide and links to their websites, contact information, and key statistics.
- **Agencies by subject**—select a subject (drop down list of key subjects)
- **Press Releases**—The latest news and announcements from individual agencies.
- **Kids' pages**—on agency websites.
- **Data access tools**—Selected agency online databases.

Other features

- **Additional Links**—to other statistical sites and general government locator sites.
- **About FedStats**
- **Feedback**
- **Federal statistical policy**—Budget documents, working papers, and Federal Register notices.
- **Site privacy policy**
- **Site document accessibility**

Many of these features and capabilities offered at the FedStats including the design of the homepage have evolved over time as a result of usability testing and research on information seeking behaviors. For example, Topic links A to Z, three different experimental versions in addition to the active version on the FedStats website were tested to help determine the best way to present an index of topics. The results of the usability testing helped guide the current design of the index.⁶

Site Architecture

The FedStats site is designed to be robust and flexible in terms of data access and display. Web pages are designed to meet the Federal government requirements for access by the disabled (Section 508 of the Rehabilitation Act) as well as being accessible to the wide variety of web browsers available on personal computers and mobile devices, such as cell phones. Computer hardware that uses the Unix operating system is used for the public portion of the site, and development work is done on computers that use the Linux operating system. Open Source software has been used extensively on the site because it is robust, scalable and a very usable utility for web development. Open Source software is software that is available for use without the payment of royalties or fees to an organization, and may be inspected and further modified as needed by other programmers. A variety of Open Source software tools are used extensively in developing the FedStats site, and including the Linux operating system, the Apache web server, the MySQL database server, and Perl and PHP software code for the development of web pages.

⁶ Hert, C.A., Jacob, E.; Dawson, P. (2000). A Usability Assessment of Online Indexing Structures in the Networked Environment. Journal of the American Society for Information Science 51(11): 971-988. The technical report is available at <http://istweb.syr.edu/~hert/BLSPphase2.html>

Ongoing Projects

FedStats Section 508 Accessibility Workshop

Section 508 of the Rehabilitation Act requires Federal agencies to meet specific requirements in making their websites accessible to people with disabilities. Several of the requirements are particularly problematic for the Federal statistical community as they affect tables, statistical graphics, and formulas. However, little attention has been paid to the accessibility of these elements in a statistical context. Given the enormous volume of tables, formulas, and statistical graphics on Federal statistical agency sites, FedStats Interagency Task Force decided to sponsor a 508/Accessibility Workshop on June 24, 2002, to focus on ways that statistical agencies can meet the new accessibility requirements and make their Web content accessible to people with disabilities. The workshop brought together about 150 participants including Webmasters and content managers from statistical as well as other federal agencies, researchers, vendors (assistive technology, Web editors and validators, and authoring tools), standards organizations, and the disability community. Forty Federal agencies were represented. Presentations and related materials from the workshop are available at <http://workshops.fedstats.gov>.

As a result of the workshop, Interagency Task Force plans to release three papers in the newly established FedStats Working Paper series. The first paper will summarize the workshop proceedings--highlighting the areas in which additional research and work needs to be done. The second paper will offer a recommend implementation of the Section 508 guidelines for tables as a short-term solution to the problems many agencies are facing. And the third paper will propose ways in which the current standards could be changed to better facilitate the accessibility and usability of complex statistical tables.

MapStats for Kids

In August of 2001, the FedStats Taskforce received a \$90,000 cash award through a competitive selection process from the e-Government Committee of the Federal CIO Council for the development of a *MapStats For Kids* section of the site. The *MapStats for Kids* project is focused on making Federal statistical information interesting and meaningful to younger citizens and thereby foster the development of statistical literacy. Statistical literacy can be viewed as the ability to interpret, critically evaluate, and communicate about statistical information, conveyed either through numbers or graphics. The GeoVISTA Center and Geography department of Penn State University was selected to work on developing a prototype for a *MapStats for Kids* section of the site based on their past work in geospatial displays of quantitative information. A target audience of fourth to eighth graders was selected as being age-appropriate in the development of the prototype, and the software tool Macromedia *Flash* was chosen to create interactive web applications that would engage the target audience. By presenting young citizens with statistical data and information in an engaging manner, these visitors to the site would be stimulated to further explore and ask questions about the various data series collected and disseminated by the various Federal statistical agencies.

To date, several prototypes have been developed which work towards developing three sets of skills that are central to statistical data analysis: logico-mathematical skills, representational skills, and spatial skills. Logico-mathematical skills can be related to the concept of geo-coding, that, understands the unique representation of units within a hierarchical framework, such as countries, states and counties. Representational skills can be represented by the concepts of understanding symbols on a map—blue for water, black for roads—and the context in which they represent. For example, a black line may represent a road, but due to its small representational size on a map, younger children may view it as not representative of their real-world experiences of what constitutes a road. Spatial skills can be thought of as representational objects, such as the outlines of state boundaries or three dimensional shaded relief projections on a map, and mentally “mapping” them into a context that conveys the underlying meaning. All of these skills are important in the process of manipulating and understanding statistical data. For example, young children may be presented with the current rankings of sports teams located throughout the United States, and using these three skill sets could gain a better understanding of the concepts of averages, regional variations, and the concept of place among various geographic boundaries. As the project progresses, the FedStats Taskforce will continue to evaluate and suggest different strategies that can be utilized in developing these skill set areas, with the goal to have a fully functional *MapStats for Kids* section on the site with the resultant software code available for use by other interested agencies.⁷ (For more information about this project see: <http://www.geovista.psu.edu/grants/MapStatsKids/index.html>)

Outreach and Promotion

The FedStats Interagency Task Force recognizes the need to systematically undertake outreach and promotion activities. While some efforts have included working with the Interagency Public Information Officers, others have included printing brochures and flash cards for distribution by individual member agencies at their respective outreach events. Still other efforts have included contacting members of the news media to feature new FedStats capabilities.

FedStats is represented on the Cross Agency e-Gov Solutions Working Group that is a part of the Government Services Administration’ Office of Citizen Services and Communications which is responsible for the First Gov web portal. As a member of the working group, FedStats seeks to share best practices with other portal projects across the spectrum of Federal agencies.

FedStats has garnered the interest of non-Statistical agencies like the Department of Housing and Urban Development and the U. S. Geologic Survey and has collaborated on several projects related to geospatial representations of agency information with FedStats data and applications. Both of these agencies have discussed becoming official members of the Interagency Task Force, and have in the past, contributed towards the design, development and implementation of the MapStats project.

⁷ Paraphrased from the MapStats for Kids - Phase I Report; PI Alan M. MacEachren et. al; GeoVista Center and Geography, Pennsylvania State University, July 30, 2002, page 3.

Another manner in which FedStats promotes itself is to enter competitions for recognition, some of which award funds to the winners. For example, FedStats was awarded \$90,000 by the CIO Council for development of the “MapStats for Kids” project.

New Project

Improving Automated Access to Statistical Databases

Most federal statistical agencies provide user access to electronic databases and data files through their Internet websites. This is a valuable service that users of statistical data rely on and use routinely. There are, however, many users for whom the web browser interface to federal statistics does not fully support their data access needs. These “power” users are those who: (1) regularly download many databases and data files; (2) regularly download data from several agencies; (3) need downloads of entire databases; or (4) need to maintain timely subject-area databases using the most current statistical releases from one or more agencies. Ironically, this user community includes many federal agencies that use federal statistics as input to their own programs (e.g. economic analysis).

Existing technologies are available to provide power users with automated, computer-to-computer, data exchange through the Internet, but there are several roadblocks to their implementation that the Interagency Task Force is in a unique position to resolve. Among these obstacles is the lack of a standard protocol for automated data exchange. The Interagency Task Force is forming a working group to begin addressing this problem and plans to draft a protocol for exchange of non-confidential data for prototyping and testing. This protocol will be based, in part, on the method used to maintain the White House Federal Statistics Briefing Room. An additional obstacle, when a standard protocol is available, is the need for a registry of statistical agencies that support the protocol and the data they make available through it--a role parallel to the role that <http://www.fedstats.gov/> now plays for statistical agency websites.

FedStats Within The E-Government (E-Gov) Context

During the past several years, as new information technologies have proliferated and been applied to government operations and services, the public’s expectations for ease of access and use of government information and services has increased. “E-Gov” initiatives have assumed a much higher profile within the Federal Government. While agencies have increasing E-Gov demands, there are little or no new resources to implement E-Gov applications. However, for a rather small investment, the Federal statistical community is well positioned to continue building valuable E-Gov services by leveraging the various research and development collaborations being undertaken by the FedStats agencies and their public and private partners. These types of collaborations save individual agency from having to do E-Gov all by themselves. FedStats has often been cited as an exemplar for providing valuable E-Gov information services to the public.

In 2001, the Interagency Task Force conducted an intensive strategic assessment and planning process, taking into account various E-Gov Directives and initiatives issued during the Clinton Administration. The outcome was a newer, more comprehensive strategic plan with a mission,

vision, and strategic goals that would enable FedStats to move well beyond a simple, yet highly acclaimed, award-winning portal web site towards the Next Generation FedStats. A year later the mission, vision and goals remain entirely consistent with the more detailed E-Government vision outlined by the Bush administration.

“Expanding E-Government” Initiative

Mark E. Forman, the Office of Management and Budget’s Associate Director for Information Technology and E-Government issued on February 27, 2002, his E-gov strategy report entitled Implementing the President ’s Management Agenda for E-Government—Simplified Delivery of Services to Citizens. Information on this E-government effort may be found on the Internet at, <http://www.firstgov.gov>.

In the report, several key goals and strategies that are most relevant to the FedStats mission include (emphasis added)--

“Among the primary goals in the President’s “Expanding E-Government” initiative are to **make it easy for citizens to obtain service and interact with the federal government**; improve government efficiency and effectiveness; and to improve government ’s responsiveness to citizens.”

“Effective E-Gov strategies will result in significant improvements by, among other things “simplifying delivery of services to citizens; **making it possible for citizens, businesses, other levels of government and federal employees to easily find information and get service from the federal government**; and by simplifying agencies' business processes and reducing costs through integrating and eliminating redundant systems.”

And, on providing opportunities to transform delivery of government services, the report provides the following guidance:

“Build easy to find, easy to use, one-stop points-of-service that make it easy for citizens to access high-quality government services.”

The report concludes that the E-Gov pay-off will not result from automating current processes, but rather through the:

“...transformation of how the government interacts with its citizens and customers. Only through changing how we do business internally —that is, streamlining work processes to take advantage of modern IT systems —will citizens experience the transformation envisioned.”

FedStats is entirely consistent with Forman’s E-government strategy and is clearly evident in the FedStats mission, vision and strategic goals:

Mission Statement

To provide effective, efficient, and timely access to, and use of, the full range of Federal statistical information needed for informed decision-making.

Vision

Informed decision-making starts with the information and knowledge available through FedStats.

Strategic Goals

- To provide Federal statistical information/knowledge effectively, efficiently, and in a timely manner.
- To enhance the effective use of statistical information.
- To provide an organizational framework and resource base in order to achieve the FedStats' mission.
- To foster broad collaboration that can strengthen the statistical system.
- To be widely recognized as an essential resource and knowledge base for informed decision-making.

To effectively accomplish the mission, vision, and goals, the Interagency Task Force and the FedStats website will have to continually evolve. While the Interagency Task Force remains a collection of involved and committed agency representatives meeting on a monthly basis, the actual infrastructure is becoming more substantial and agile because of several factors noted below.

The Interagency Task Force recognized that ICSP agencies needed assistance in leveraging and/or making operational, in a more systematic and beneficial manner, best practices and approaches for statistical information dissemination, methods, and new technologies. These might be developed within the FedStats environment, or might be those innovations being developed in individual agencies and/or by academic researchers collaborating with statistical agencies via NSF's DG Research Program.

In September 2002, the ICSP agreed with the Interagency Task Force's recommendation to hire a full-time program manager for FedStats and to fund the position by increasing individual agency contributions. to cover the costs of a full-time FedStats Program Manager. The total FedStats budget in FY 2003 will be \$470,000 (5 largest agencies contributing \$50,000 each and the 9 smaller agencies contributing \$25,000 each).

Next Generation FedStats

FedStats will continue to be a premier E-government portal. So what is the Next Generation of FedStats and how might it differ from the current portal?

The Next Generation FedStats will be a national distributed statistical digital library with tools for information finding, for information extraction and reuse, information visualization, and for transforming knowledge into intelligence while maintaining the privacy and confidentiality of respondents. To achieve this vision, FedStats will require common user interfaces, data access and searching tools usable by persons with different levels of computer and statistical literacy, which enables appropriate uses of the data with analysis within and between databases.

The current decentralized, independent sources of statistical information have few commonalities in terms of concepts and definitions; system architectures, software, and hardware; measurement methods; interfaces; or dissemination and presentation modalities. Interoperability is a major hurdle in a variety of areas. Data integration issues abound. Significant challenges in high-end computing and computation and large-scale networking exist for the making the Next Generation FedStats vision a reality.

Computer and information scientists will solve some of these challenges, while others will require a more multidisciplinary, multi-sector approach. For example, involving mathematical statisticians with expertise in creating estimates from complex sample surveys, building small area estimation models, and estimating measures of error for the resulting estimates that incorporate all sources of error, including those due to sampling and nonsampling errors.

If the metadata needed to interpret and use statistical information are to be made available and integrated with the data, the processes and procedures for collecting and compiling statistical information must also be the focus of information technologies research and development efforts.

As one of the first set of Federal agency partners with the NSF in its Digital Government program over four years ago, the statistical agencies have improved upon their historical tradition of being in the forefront in exploring new and novel ways to better handle the ever-increasing volume of data that flow from the varied statistical programs of the U.S. government. In turn, the NSF and the research community have recognized that the Federal statistical agencies have a unique challenge in ensuring that statistical information is collected and provided to the public in as robust and reliable manner as possible, while ensuring that cost-efficiencies are achieved.

Digital Government Research Projects

Over the past five years, the statistical community has taken the “longer-view” on how to improve the Federal Statistical community’s data and information dissemination programs. The NSF’s Digital Government Research Program is providing government agencies with unique opportunities to better understand what new information technologies are being developed in university research labs and to participate in test bed applications development along side the researchers funded being funded by NSF. Many opportunities exist for leveraging these research efforts (and those yet to be defined) that could lead to radical improvements in agency business practices as well as improving government information services.

In February 1999, the National Academies of Science’s Committee on Computing and Communications Research to Enable Better Use of Information Technology in Government, chartered by the National Research Council’s Computer Science and Telecommunications Board (CSTB) and the Committee on National Statistics (CNSTAT) held the second of two workshops as part of a larger study being undertaken at the request of the DG research program. The workshop focused on the Federal statistics application area.

“Underlying the presentations and discussions at the workshop was a desire to tap IT innovations in order to realize a vision for the federal statistical agencies. A prominent theme in the discussions was how to address the decentralized nature of the US national statistical system through virtual mechanisms. The look-up facilities provided by the FedStats Web site are a first step toward addressing this challenge. Other related challenges cited by workshop participants include finding ways for users to conduct queries across data sets from multiple surveys, including queries across data developed by more than one agency—a hard problem given that each survey has its own set of objectives and definitions associated with the information it provides.”⁸

Further, the workshop identified a broad range of IT issues for engaging the information technology research and federal statistics communities in research activities of mutual interest. These include human computer interaction, database system, data mining, metadata, information integration, and information security. Two other challenges of particular interest include survey instruments and the need to limit disclosure of confidential information.⁹

In the convening years, the NSF’s DG program has funded more than fourteen FedStats research projects. These projects examine such topics as privacy and confidentiality issues in microdata files, new ways to display information contained in statistical tables, tools and methods for automatically building metadata, testbeds for distributed architectures that enable data integration, data collection technologies such as those involved in the use of handheld devices and wireless data transmission, data visualization and validation technologies, etc. Now the challenge is transferring knowledge and/or technologies from the research labs to production systems. The FedStats environment now that there will be a more permanent infrastructure can help with the transition of results to the FedStats website and or to individual statistical agencies. (See Appendix II for more details)

NSF’s digital government grantees have received over \$10 million to focus on FedStats-related research. The ICSP agencies have already augmented the NSF awards by approximately an additional \$2 million.

⁸ *Summary of a Workshop on Information Technology Research for Federal Statistics*; Computer Science and Telecommunications Board and the Committee on National Statistics; National Research Council. The full report can be found at http://books.nap.edu/html/itr_federal_stats/

⁹ *Summary of a Workshop on Information Technology Research for Federal Statistics*; Computer Science and Telecommunications Board and the Committee on National Statistics; National Research Council. The full report can be found at http://books.nap.edu/html/itr_federal_stats/

In the DG Research Program's recent announcement, proposals are being accepted for two classes—

- 1) Multi-disciplinary and multi-sector partnerships of researchers in information technologies and government agencies at all levels in order to foster collaboration among societal sectors, and
- 2) Research on the relationships between the design and use of information technologies on: i) forms, processes, and outcomes of democracy, ii) government organizational forms, learning, and adaptation, iii) new forms of government-government collaboration, iv) citizen/government interaction, and v) other social and political science research related to IT and government.

This second class of proposals, in addition to the first class, which FedStats has leveraged quite well, will enable scientists to better identify and understand the government and citizen user needs for the Next Generation of FedStats. This is an untapped opportunity ripe for further exploration by the Federal Statistical community.

FedStats Interagency Research and Development (R&D) Working Group

In addition to the FedStats Interagency Task Force, in 1997 the ICSP authorized a FedStats interagency R &D working group. As a first step, the working group identified common challenges facing many statistical agencies that could potentially be overcome by applying cutting-edge information technologies. The FedStats R&D working group coordinates the Federal agency responsibilities and activities (along with the academic researchers) as outlined in each DG research proposal. The FedStats R&D working group also is fostering new and/or modified FedStats R&D partnerships that will continue to develop research proposals for submission to the wide array of NSF and other Federal agency research programs.

Conclusions

In seven years much has been accomplished for laying the frameworks for statistical data dissemination in the 21st century, both within individual agencies and by interagency efforts such as FedStats. However, to realize the Next Generation of FedStats much remains to be done.

As noted in the May 2002 CSTB report *Information Technology Research, Innovation and E-Government*¹⁰

“A number of these portals represent a fairly mature realization of present-day information-access technology, but considerable scope for improvement remains.”

¹⁰ National Academies of Science, National Research Council's Telecommunications Board (CSTB) May 2002 report entitled *Information Technology Research, Innovation and E-Government*. The full report is available on-line at http://books.nap.edu/html/itr_e_gov/.

“At present, much of the thinking about e-government focuses on what can be delivered with today’s technology...But it is also essential that, in looking ahead, planners contemplate how both technology and user expectations will evolve.”

The ICSP, the FedStats Interagency Task Force, and the FedStats R& D Working Group are looking ahead, trying to bring the Next Generation FedStats to fruition. Partnering with the NSF academic community is one way in which strategic understanding of new technologies can most effectively be put to use in bringing the best tools, technologies, and policies into practice. Technology transition will remain a challenge, but the FedStats environment, with a solid infrastructure in place, is well positioned to make the vision “Citizen Access to Federal Statistics” in the 21st century a reality.¹¹

¹¹ National Academies of Science, National Research Council’s Telecommunications Board (CSTB) May 2002 report entitled *Information Technology Research, Innovation and E-Government*. The full report is available on-line at http://books.nap.edu/html/itr_e_gov/.

Appendix I

Task Force Agency Liaisons

Valerie Gregg, Co-Chair, Census/NSF
Marshall DeBerry, FedStats Program Manager, BJS
Cathryn S. Diplo, BLS and Chair of the FedStats R & D Working Group
Michael Moore, BEA
John Bosley, Rick Devens, BLS
Marianne Zawitz, BJS
Jeff Butler, BTS
EPA
Rachael Taylor, David Raszewski, Census
John Weiner, Colleen Blessing, William Jeffers, EIA
Jim Horsfield, ERS
George Patton, NASS
Bruce Taylor, NCES
Rob Weinzimer, NCHS
John Gawalt, NSF
William Wong, SOI
Laurie Brown, SSA
David Chase, John Sperling, HUD
Bill Tolar, USGS

Appendix II

FedStats/NSF Digital Government Research Projects

A list of the research project titles, the principle investigators (PI), and their academic institutions follows. In addition to these fourteen research grants, the DG Research Program has awarded another 9 FedStats-related grants including one workshop, four planning, and four small grants for exploratory research.

1. [A Web-Based Query System for Disclosure-Limited Statistical Analysis of Confidential Data](#); PI Alan Karr; National Institute of Statistical Sciences

Working with several Federal statistical agencies, this grant will address an important topic for Federal statistical agencies. As part of their missions, these agencies collect a great deal of microdata (data related to an individual or particular business); this data must remain confidential. Thus, only aggregated microdata is provided publicly. However, the aggregation process reduces much of the value of the microdata for deriving knowledge to be used in research, policy and commercial purposes, so there is a balancing need to provide as much data as possible. What is proposed here is a large-scale system which tracks the history of provision of derived data and which "understands" and can quantify the potential for working backward from the derived data.

2. [Data Confidentiality, Data Quality, and Data Integration for Federal Databases: Foundations to Software Prototypes](#); PI Alan Karr; National Institute of Statistical Sciences

This award will support research in data confidentiality, data quality, and data integration. Prototypes will be built which can scale to operate on large sets of federally held data. Researchers will partner with several large Federal Government statistical agencies. This topic is of particular importance given the balance these agencies must strive for, in terms of their dual missions to collect and keep private confidential data, while at the same time making that data accessible for research and policy issues. This grant will support a multi-disciplinary multi-institution team, with participants from five universities, one non-profit, and one national laboratory. The disciplines represented include computer science, statistical science, and systems engineering.

3. [Adaptive Interfaces for Collection Survey Data From Users](#); PI Michael Schober; New School University

The objective of this research is to determine how best to design computer systems for collecting data from (rather than providing data to) users. Government agencies might use such systems to gather the factual data used to calculate the unemployment rate or the Consumer Price Index. Three sets of laboratory experiments focus on actual and simulated desktop (i.e., keyboard and mouse entry) and speech survey interviewing systems. The first

set of studies examines response accuracy and user satisfaction with systems that monitor users' speed of responding and speech patterns in order to diagnose when users misinterpret concepts in the survey questions and could use additional clarification. The second set of studies examines user response accuracy and satisfaction with interfaces that do (or do not) tailor this clarification through dialogue. The third set of studies contrasts interfaces that require users to educate themselves about how the questions should be interpreted with interfaces that engage users in dialogue to figure out the correct answer. The project uses the methods of experimental psychology to provide guidelines for future development of interfaces that collect information from users. This research could significantly improve the accuracy of data collected online by government agencies and others.

4. [Citizen Access to Government Statistical Data](#); PI Gary Marchionini, University of North Carolina

This proposal will conduct research to improve the location/retrieval, reading, navigation and manipulation of tabular statistical data from Federal agencies. These data cover many different domains (e.g., health, labor, transportation), of interest to professionals in the field and to citizens. This work will be accomplished through collaboration with the Bureau of Labor Statistics, the Energy Information Agency, and the National Center for Health Statistics and the Bureau of Census.

5. [Collaborative Research: Integration of Data and Interfaces To Enhance Human Understanding of Government Statistics—Toward the National Statistical Knowledge Network](#); Co-PI Gary Marchionini, University of North Carolina and Catherine Plaisant, University of Maryland

This award will support collaborative research with several Federal statistical agencies to develop better statistical data models, to explore the use of XML, to develop better map-querying tools and to integrate other available tools for manipulating, browsing, and visualizing tabular data. The goal is to develop better human/computer interfaces for expert users to novices, to increase general statistical literacy, and to provide seamless access to data held by multiple Federal agencies and agencies at other levels of government, in particular state and local data.

6. [Quality Graphics for Federal Statistical Summaries](#); PI Alan MacEachren; Penn State University

This award will support collaborative research with several Federal statistical agencies to develop better statistical data models, to explore the use of XML, to develop better map-querying tools and to integrate other available tools for manipulating, browsing, and visualizing tabular data. The goal is to develop better human/computer interfaces for expert users to novices, to increase general statistical literacy, and to provide seamless access to data held by multiple Federal agencies and agencies at other levels of government, in particular state and local data.

7. [Quality Graphics for Federal Statistical Summaries](#); PI Dan Carr; George Mason University—See MacEachren

8. [Quality Graphics for Federal Statistical Summaries](#); PI David Scott; Rice University—See MacEachren

9. [Collecting and Using Geospatial Data in the Field: An Extensible Framework and Testbed](#); PI, Sarah Nusser, Iowa State University

This work will conceive, develop, and test an extensible framework to support the collection and use of geospatial data in the field. Partner Federal agencies include the Bureau of the Census, the US Geological Survey, and several agencies of the US Department of Agriculture. The proposed activities are designed to meet five key objectives:

1. Develop a model documenting and formalizing the infrastructure, tools, and key capabilities required to support a flexible and extensible field data collection system.
2. Conduct research on computer science tools and associated information technologies required to fully integrate digital geospatial data into the collection process.
3. Conduct research on infrastructure components that are needed to implement the system in a manner that limits the complexity of the system from the vantage point of the user in the field.
4. Investigate emerging field data collection technologies to determine how the usage of geospatial data is transformed by these new interfaces.
5. Explore the framework model and research developments in an application environment by developing prototype components and testbeds that correspond to agency data collection settings.

Six developments will be needed to address the research objectives: 1. A user-driven framework model, 2. A conceptual framework for conflation of heterogeneous geospatial data for field use, 3. A multi-agent system to support tools required using and collecting geospatial data in the field, 4. Interoperable searching and discovery mechanism for prepared, existing, and potentially unknown sources of data, 5. Object-oriented warehouse designs for the field data collection environment, and 6. Evaluations of emerging field technologies and their impact on user activities.

10. [Digital Government Research Center: Energy Data Collection](#); PI Yigal Arens; Information Sciences Institute, University of Southern California

This proposal will create an Energy Data Collection to support real-time integrated viewing, interaction, and manipulation of the Department of Energy's gasoline-related data collection, through a partnership with the Energy Information Agency. The proposed research will cover automated ontology development and distributed information integration across data held by multiple Federal agencies.

11. [I2T: An Information Integration Testbed for Digital Government](#); PI, Chaitan Baru; San Diego Super Computing Center

This project will address one of the major problems in government information systems, the inability to integrated information from various heterogeneous data sources. Usually these data are collected and managed by different agencies at different levels of government, providing more impediments to integration. Partners from the Bureau of the Census, National Archives and Records Administration, US Geological Survey, the State of Pennsylvania, and the San Diego Association of Governments will work with researchers from the San Diego Supercomputer Center, the University of California at San Diego, the University of Michigan, and the University of Pennsylvania. Building upon the initial work of the Mediation of Information using XML (MIX) project, this grant has four major technical thrusts: 1. Allow for an extension of MIX's wrapper technology to the domain of geospatial information, 2. Develop data transfer protocols for lightweight network-based agents, 3. Investigate new interfaces to the data, and 4. Build wrapper toolkits for geospatial and statistical survey metadata.

12. [Survey Authoring and Administration Testbed](#); PI, Robert Balzer; Information Sciences Institute, University of Southern California

This grant will address an important problem area for the US Bureau of the Census, and through them, the various Federal agencies who commission the Bureau to conduct statistical survey, i.e., the specification and creation of complex survey instruments. At present Census is using a very old proprietary system, which occupies nearly 100 Census staff. The PI will use commercially-available software as an infrastructure upon which will be created a research prototype of a modern Web/relational database system, using modern software engineering techniques, to allow graphically-specific surveys with built-in error-checking and administration.

13. [Digital Government Research Center \(DGRC\): Bringing Complex Data to Users](#); Co-PI Judith Klavans, Columbia University and Ed Hovy, Information Sciences Institute, University of Southern California

In partnership with the Federal Energy Information Agency on the topic of trade data, Columbia University and the Information Sciences Institute of the University of Southern California will work in three areas of relevance to the Agency mission: 1. Main memory query processing, which provides extremely fast querying of multiple statistical data sets, an area of concern to all statistical agencies which must provide aggregated data which maintains the confidentiality of the citizens and businesses which contributed the data; 2. Multilingual question and answering, which will explore the possibility of providing automated translation and querying from English to Spanish and Chinese, and perhaps one other language. As the US population becomes increasingly multi-lingual, natural language processing as a service of gov't web sites will become more and more expected. 3. Usability testing of components developed in this and in another grant to this team under the Digital Government program.

14. Digital Government: Improving Statistical Literacy Through FedStats; PI Bill Smith;
American Statistical Association

This grant will support a planning process to develop concepts for research in user interfaces and forms of on-line learning and analysis to improve statistical literacy for the citizen. The proposer will work with an existing group of collaborating Federal statistical agencies, know as “FedStats”. FedStats has an award-winning web site, and collaborates with several other Digital Government award recipients.

