

Internet Data Collection System Requirements Study

FINAL REPORT

This evaluation study reports the results of research and analysis undertaken by the U.S. Census Bureau. It is part of a broad program, the Census 2000 Testing, Experimentation, and Evaluation (TXE) Program, designed to assess Census 2000 and to inform 2010 Census planning. Findings from the Census 2000 TXE Program reports are integrated into topic reports that provide context and background for broader interpretation of results.

Prepared by Titan Systems
Corporation/
System Resources Division

Kevin A. Shaw, Project Manager
Planning, Research, and
Evaluation Division

U S C E N S U S B U R E A U

Helping You Make Informed Decisions

Intentionally Blank

PREFACE

Purpose of the System Requirements Study

The main objective of the System Requirements Study is to assess the efficacy of the requirements definition processes that were employed by the U.S. Census Bureau during the planning stages of the Census 2000 automated systems. Accordingly, the report's main focus is on the effectiveness of requirements methodologies, including processes for coordination, communication, and documentation, and their impact on overall system functionality. The report also addresses certain contract management issues and their effect on system development and/or operational considerations.

The System Requirements Study synthesizes the results from numerous interviews with a range of personnel--both U.S. Census Bureau staff and contractors--who were involved with the planning, development, operations, or management of Census 2000 systems. Our findings and recommendations in this report are qualitative in nature; they are based on the varied opinions and insights of those personnel who were interviewed. The intent is to use the results from this study to inform planning for similar future systems.

Intentionally Blank

CONTENTS

EXECUTIVE SUMMARY	iii
1. BACKGROUND	1
2. METHODOLOGY	1
3. LIMITS	2
4. RESULTS	3
4.1 Requirements definition	3
4.2 Requirements issues	4
4.3 Alignment with business processes	6
4.4 System deficiencies	7
4.5 Contract management practices	9
5. RECOMMENDATIONS	9
5.1 Reliance on in-house expertise	9
5.2 Public awareness	10
5.3 Time limitations	10
5.4 Requirements methodology	11
5.5 Integrating system components	11
5.6 System development life-cycle (SDLC) methodology	12
5.7 Project oversight	12
5.8 Form replication	13
5.9 System sizing	13
5.10 Census 2000 Dress Rehearsal	13
5.11 Security	14
References	15
Participants	16

Intentionally Blank

EXECUTIVE SUMMARY

Internet Data Collection and its companion system, Internet Questionnaire Assistance, are Internet-based systems that were developed by the same team. While there is a technical linkage from an operational standpoint, for purposes of this report they will be treated as separate systems. This study presents information based on debriefings with Internet Data Collection system designers. A separate customer service survey evaluation will provide the user perspective of this system.

U.S. Census Bureau personnel assessed Internet Data Collection as being very successful in spite of limited usage by the public. The overall objective of the system was to provide census respondents with a highly secure Internet filing option to the paper-based short form questionnaire. Internet Data Collection succeeded in replicating the form's key features. Software limitations prevented precise replication of the paper form in every detail, however, this did not impact the form's usefulness.

One of the goals of the Internet Data Collection system was to ease the form capturing burden on the U.S. Census Bureau Data Capture System 2000; however, it did not meet this expectation as only about 66,000 forms were submitted through Internet Data Collection in March/April 2000. During this time, system utilization was characterized by two very brief periods of heavy access. Internet Data Collection did not achieve consistent and widespread usage owing to a lack of publicity that stemmed from Internet related security concerns at senior management levels. The system was developed under a very aggressive schedule and deployed within 18 months. Major results of the study include:

- Right system for the job. Internet Data Collection was positively perceived by the personnel interviewed as “the right system for the job.” It provided an effective means of capturing Census 2000 information through the submission of short form questionnaires over the Internet. The few system shortcomings were all minor. However, there were some usability problems that were not resolved.
- Security was essential. One major requirement and design issue was the need for the system to provide strong levels of security against unauthorized access; Internet Data Collection succeeded in providing this protection. This requirement interfered with including members of the public in usability testing.
- Dedicated personnel were a key factor. The success of the system was largely due to the involvement of a few highly talented and dedicated Census personnel. Contractor support (with the exception of penetration testing) did not play a role in the planning or development of the system.

These and other findings have led to the following key recommendations:

- **System development methodology - establish agency-wide guidance.** There were no formal (i.e., agency endorsed) guidelines to indicate how the requirements definition process should be conducted. The requirements definition phase is an especially critical step in all generally accepted system development life-cycle methodologies in that it establishes the foundation for a system. Therefore, future systems development efforts would benefit from an agency-wide set of guidelines which outline the steps in the requirements definition process including methods for change control. It is recommended that the agency implement and adhere to such guidelines well in advance of future application development efforts. These guidelines must be flexible in order to minimize any adverse impacts on the design and development effort.
- **Time limitations - begin early planning for 2010.** Internet Data Collection was essentially over-engineered for capacity and redundancy because the anticipated traffic levels did not materialize due to a lack of publicity. Widespread Internet usage will place significant demands on the next generation data collection system in 2010. Therefore, it is recommended that planning commence early and reflect an expectation of radically different capacity and redundancy requirements.
- **Reliance on in-house expertise - assess staffing risks.** Historically, the agency has relied on in-house expertise to develop decennial systems. In the case of Internet Questionnaire Assistance, a single individual was the driving force behind the development effort. Given the nature, scope, and complexity of future systems, there are risks inherent in relying solely on in-house staff without supplementing these resources with external support.

1. BACKGROUND

The Titan Systems Corporation, System Resources Division (Titan/SRD) was tasked by the Planning, Research, and Evaluation Division (PRED) of the U.S. Census Bureau to conduct system requirements studies for 12 automated systems used in the decennial census. This report is a study of the Internet Data Collection (IDC) system. It addresses the extent to which the requirements definition process was successful in identifying the needed system functionality and offers one of several evaluation approaches for examining these automated systems. The report results are intended to assist in the planning of similar systems for the 2010 Census.

IDC is an application that allowed respondents to complete the Census 2000 short form questionnaire on-line through the Internet. The long form was not incorporated into the system because of technical complexities such as extensive conditional branching and the time required to display the form on the local browser. Another consideration was the relatively small universe of long form recipients.

An interdivisional team, consisting of representatives from eight divisions, had responsibility for developing IDC. To prevent unauthorized access, respondents provided a 22-digit housing unit identification number (obtained from the Census 2000 short form) when completing this form. Processing procedures ensured that duplicate responses and questionnaires were not accepted as part of Census 2000 results. The IDC operated between March 3, 2000 and April 18, 2000. Census Bureau extended the end date to coincide with the cutoff date for late mail returns.

A potential benefit of the IDC system was to ease the form capturing burden on the Census Bureau Data Capture System 2000 (DCS 2000). This benefit was not realized however, due to a rescoping of the IDC project. A total of approximately 66,000 forms were submitted through IDC in March/April 2000. During this time, system utilization was characterized by two brief periods of heavy access; this includes two “spikes” of form submissions, one of 7,539 on March 13 and another of 3,159 on April 1. These dates correspond to the day the United States Postal Service (USPS) delivered the forms and Census Day, respectively.

In addition to easing the data capture burden, IDC offered the public an additional forms submission option as a means to encourage response.

2. METHODOLOGY

The Titan/SRD Team interviewed key personnel for each of the Census 2000 automated systems using a structured approach centered around four fundamental areas. A set of questions under each of those areas was designed to explore: (1) the effectiveness of the requirements definition process; (2) how well the systems were aligned with business processes; (3) identification of any deficiencies in functionality or performance relative

to actual operational needs; and (4) how effective the agency contract management activities were in regards to contractor performance. Contractors were not involved in the development of IDC.

The purpose of the system requirements study is to summarize the results of interviews with key personnel by system. A variety of related system documentation was reviewed in connection with the interviews. The assessments provided in Section 4., Results, reflect the opinions and insights of key personnel who were interviewed by the Titan/SRD Team in September 2000. Those personnel had varying levels of knowledge about the IDC system based on their involvement with system planning, development, implementation, or operational issues. Section 5., Recommendations, provides value added perspectives from the Titan/SRD Team that seek to illuminate issues for management consideration in the planning of future systems.

Quality assurance procedures were applied to the design, implementation, analysis, and preparation of this report. The procedures encompassed methodology, specification of project procedures and software, computer system design and review, development of clerical and computer procedures, and data analysis and report writing. A description of the procedures used is provided in the “Census 2000 Evaluation Program Quality Assurance Process.”

Study participants reviewed the results of this system requirements study. Comments have been incorporated to the fullest possible extent.

3. LIMITS

The following limits may apply to this system requirements study:

- The perception of those persons participating in the interview process can significantly influence the quality of information gathered. For instance, if there is a lack of communication about the purpose of the review, less than optimal results will be obtained and the findings may lack depth. Each interview was prefaced with an explanation about its purpose in order to gain user understanding and commitment.
- In some cases, interviews were conducted several months, even years, after the participant had been involved in system

development activities. This extended timeframe may cause certain issues to be overlooked or expressed in a different fashion (i.e., more positive or negative) than if the interviews had occurred just after system deployment.

- Each interview was completed within a one to two hour period, with some telephone followup to solicit clarification on interview results. Although a detailed questionnaire was devised to guide each interview and gather sufficient information for the study, it is not possible to review each aspect of a multi-year development cycle given the limited time available with each participant. Although this is a limitation, it is the opinion of the evaluators that sufficient information was gathered to support the objectives of the study.
- Every effort was made to identify key personnel and operational customers who actively participated in development efforts. In the case of IDC, all government personnel who participated in the study are still with the Census Bureau. Since contractors were not involved in the development of the system, no contractor representatives were interviewed for this study.

4. RESULTS

This section contains findings that relate to the effectiveness of the requirements definition process used during the development of IDC. The requirements process establishes the foundation for a system and, as such, must be designed to thoroughly consider all technical and functional aspects of development and operation of the system.

4.1 Requirements definition

Early efforts leading to the development of IDC can be traced back to 1996 when the Commerce Department was investigating Internet reporting and assistance options. By 1997, a group known as the Census 2000 Internet Questionnaire Team (C2IQT), was actively looking into the feasibility of using Internet technologies to support innovative business processes. However, in 1998, top level management concerns over the potential for security weaknesses led to a cessation of planning activities. Work resumed in October 1998, due to emphasis on e-government initiatives at the Commerce Department. The C2IQT was subsequently reformed as an interdivisional group known as the Internet Questionnaire Team (IQT).

The IDC requirements process benefited significantly from the involvement of a few talented people with technical skills and organizational knowledge as well as Census Bureau's previous experience with web-based applications. Because time constraints precluded a traditional development cycle, an interdivisional team was put in place by the Decennial Management Division (DMD) to accomplish the project. The team consisted of representatives from the following divisions:

- Computer Assisted Survey Research Office (CASRO)
- Decennial Management Division (DMD)
- Decennial Systems and Contracts Management Office (DSCMO)
- Decennial Statistical Studies Division (DSSD)

- Housing and Household Economic Statistics Division (HHES)
- Population Division (POP)
- Planning, Research, and Evaluation Division (PRED)
- Systems Support Division (SSD)

The team met on a frequent basis for one year to monitor status and progress and to address issues. Initially, it was thought IDC would be a major decennial system; however, owing to a lack of publicity, the scope of the system was diminished.

Requirements were continuously being “fed” to the developer right up to system deployment. The process lacked formal change control procedures. The developer was an in-house Information Technology (IT) specialist who also participated in the requirements definition process. There was no formal (i.e., agency endorsed) requirements process or change control process in place to guide the team. The system requirements definition and development processes were, by default, essentially a Rapid Applications Development (RAD) approach¹. Time constraints and a lack of any systematized development methodology led to an accelerated, but effective, system development effort that succeeded in bringing IDC (and IQA) on-line in time for Census 2000 despite the break in continuity resulting from the cessation of early planning efforts.

4.2 Requirements issues

4.2.1 System design supports maximum accessibility

Access to IDC was available on-line through the Census Bureau’s Census 2000 Internet site and via the Census Bureau’s main Internet site. The Universal Resource Locators (URLs) are www.Census2000.gov and www.Census.gov, respectively. A requirement to maximize access to the website by the public was successfully met through a design which supported the use of several types and versions of Internet browser software. This goal was made difficult by the existence of many different types of web browsers--Microsoft Internet Explorer; Netscape Navigator; WEB TV, America Online--and compounded by the lack of any industrywide browser standard. The Microsoft Internet Explorer was particularly problematic for the developer. Overall however, there appeared to be very few submission failures. An operational analysis is being conducted on IDC that may provide insight into these issues; A.2.b, Census 2000 Operational Summary of Internet Data Collection (IDC).

¹ The RAD concept involves developing faster and higher quality applications through: (1) requirements gathering using workshops or focus groups; (2) prototyping and early, reiterative user testing of designs; and (3) a compressed development schedule that defers design improvements to the next product version. It also calls for reduced formality in reviews and other team communication. With IDC development revolving mostly around a single person, it is quite accurate to conclude that a RAD approach was applied by definition since resources and time were very limited.

It was estimated that over 80 percent of the population (approximately 100 million households) would be eligible to use the short form--if they had a valid ID number, access to the Internet, and any one of several popular Internet browsers in use since 1997. According to the Department of Commerce, about 26 percent of U.S. households had Internet access at the time of Census 2000. Therefore, the short form universe for IDC was approximately 26 million households, recognizing however that some of these households would have received the long form or had access to the Internet outside the household. Accordingly, the requirements phase assumed that access would be heavy and planned on a “worst case” basis considering peak times and workloads.

4.2.2 Development lacks formal change control process

There was no formalized change control process to manage the changing features and functions of IDC. Changes recommended by the usability team and the wording sub-group were not always incorporated into the system and there was often no accounting of their disposition. Some recommendations that were included in the system were modified from the original request (e.g., rewording of instructions). However, every effort was made to implement as many recommendations as possible, but some could not be implemented because of software limitations. These modifications may have reduced the effectiveness of the user interface.

4.2.3 Penetration testing used to address security concerns

The security requirement for IDC was paramount from the outset given prior management level concerns about this issue. This was true for IQA as well, since from a security standpoint, both systems were essentially treated as one by decennial management. Accordingly, penetration testing was conducted by an outside contractor. The test plan was approved by the Census Bureau security office and designed to identify weaknesses in security that could result in unauthorized access to IDC. The system withstood all attempts to breach security.

4.2.4 User interface designed to replicate paper form

The user interface on IDC was dictated by a need to replicate the short paper form. A collateral benefit of re-creating the form on-screen as faithfully as possible was that it would reassure the public that it was, in fact, entering data into an official Census Bureau website. This requirement was understood from the outset and drove development efforts. Usability testing was also conducted in the Center for Survey Methods Research (CSMR) Usability Laboratory. Constructive feedback from CSMR led to design changes in IDC. There also was an agency-wide usability test in which over 260 Census Bureau employees participated. This test included a Usability Evaluation Questionnaire, which most participants completed. An evaluation is being conducted on Census 2000 Internet usage that may provide insight into customer reaction to the IDC user interface; A.2.c, Census 2000 Internet Web Site and Questionnaire Customer Satisfaction Surveys.

4.2.5 On-line help assists respondents

IDC addressed a requirement to provide on-line contextual help to respondents who needed assistance completing the web-based short form.

4.2.6 System scoped for high demand

System capacity and redundancy needs were greatly oversized due to expectations of high traffic volumes at the site. As previously mentioned in this section, this expectation was based on a very large potential user base (over 80 percent of Internet users). The system was designed to capture ten million forms. Since widespread utilization never materialized due to a lack of publicity about the system, IDC proved to be over-engineered for the task. The IQA/IDC system was designed with six redundant servers. The original design called for four, but because of changes in the load balancing scheme, six servers were used for both IQA and IDC. A seventh server was used for processing IDC returns. Due to the small volume of form submissions, the IDC servers were running at less than two percent of their capacity. Server redundancy “roll-overs” performed as designed.

4.2.7 Housing unit ID required to access forms

The housing unit ID was the key to accessing the web-based questionnaire and prevented unauthorized access to the system. Without this number, every household address would have to be verified--a very tedious and time-consuming process. Requirements relating to the issue of how to handle multiple entries of the same 22-digit housing unit ID number were not fully developed. An informal policy was put into place to handle these situations like the paper form.

4.3 Alignment with business processes

This section contains findings that relate to how well IDC supported the specific business processes that were associated with the Census Bureau's goal to offer some segments of the public the option of responding to Census 2000 through submission of data via the Internet. Designing IDC to fully support this objective, built upon Census Bureau's previous experience with web-based applications. The development was still a challenge however because of the requirement for capturing the submission of decennial census data through on-line forms.

4.3.1 System perceived as effective by study participants

IDC was perceived by those interviewed as being successful with respect to its ability to support census business processes and as the “right system for the job.” It was not able, however, to provide full accessibility to the disabled due to problems encountered while testing the IDC instrument with screen reading software programs that were available at the National Federation of the Blind (NFB). Disabled users include both the sight impaired who must use screen reading technology and the mobility impaired who must use assistive navigation devices. In-house testing verified that the form itself exceeded all industry accessibility standards in place at the time the form was deployed. While the form was usable, screen reader technology for web browsers had not matured enough to fully utilize these industry accessibility standards. It should be noted that the paper form also was not designed for the sight impaired.

4.3.2 Form replication limited by technology

The business process for IDC was modeled around the “short form.” IDC did not exactly replicate the paper form due to technical limitations in software, but it provided the same questions and key features in the questionnaires.

4.3.3 Security was a major consideration

Security of IDC was a major design consideration. It was achieved in the following manner: users accessed an initial page and entered their 22-digit ID number. Once this number was verified, the appropriate form for that user appeared on the screen. When the respondent finished filling out the form, it was sent back to Census Bureau host machines using data encryption technology.

4.4 System deficiencies

This section contains findings that relate to any specific shortcomings that were identified with respect to the system's ability to accomplish what it was supposed to do. Recognizing that 100 percent success is rarely achievable, especially in the case of a completely new system, it is still worthwhile to assess deficiencies in the spirit of constructively identifying “lessons learned.” Such insights can greatly contribute to improvements in future system development activities.

4.4.1 Management metrics provided although not initially defined

The need for collecting system utilization metrics for management purposes (e.g., hit counts) was “conceptualized” but not fully defined up front. Nonetheless, system usage statistics were available on a daily basis and submitted to the Executive Information System (EIS) for management use. These statistics provided extensive data for analytical purposes.

4.4.2 Edit checks were not used

A major design goal was to replicate the key features of the short form. However, some type of edit checks (or reasonableness criteria) may have been beneficial if not for time and technology limitations. For example, edit checks would have required the use of JavaScript². At the time of deployment, as many as 21 percent of Internet users had serious issues with JavaScript not working properly or had JavaScript disabled³. Since many users “turn off” JavaScript due to security and other reasons, this would have likely created access difficulties for those users who were not using JavaScript. In addition, the implementation of edit checks would have required expanded usability testing to evaluate the respondent’s reaction to the design.

4.4.3 Screen navigation difficult for some usability testers

Some aspects of “tabbing down” from field to field (i.e., moving from one answer block to another) in the IDC instrument were seen as awkward and caused a problem for some of the usability testers. This was entirely due to the lack of industrywide browser standards and was well beyond the control of the developer.

4.4.4 Lack of IDC in Dress Rehearsal reduces management confidence

Beta testing was conducted on the IDC instrument and the Systems Support Division (SSD) conducted security testing. Although testing and internal/external usability trials were performed on IDC, it was not included in the Census 2000 Dress Rehearsal. The testing was beneficial from a user interface standpoint, but the lack of a dress rehearsal contributed to a reduced confidence in the systems' ability to meet security requirements.

4.4.5 Submission failures caused by a number of factors

Entry of an ID number was required to access a form; IDC had to evaluate the ID number in order to return the appropriate form to the user. If the number entered was invalid, the system would alert the user to verify the ID number. About 17 percent of all housing unit ID submission attempts resulted in failures. This rough figure is an estimate because it cannot be determined the number of forms that were successfully resubmitted after the ID was rekeyed.

² JavaScript is an interpreted programming language/code that is typically embedded into an HTML document to enable web authors to design interactive sites with dynamic content. There are several versions of JavaScript supported by certain browsers and browser versions, which can lead to problematic software incompatibilities.

³ Source: <http://www.TheCounter.com/stats/2000/March/javas.html>

Failures were due to a number of factors such as mis-keying the number, entering a number from an unsupported form type, and using a browser with insufficient encryption. Another factor regarding the use of IDs was security. It was essential to control housing unit IDs in order to prevent individuals from making up non-existent households for the purposes of fraud. Some of the failed attempts may have been a result of these activities. Submission failures are being researched in Evaluation A.2.b, Internet Data Collection (IDC) Operational Analysis. This research will help verify the actual percentage of submission failures.

4.5 Contract management practices

This section contains findings that relate to the effectiveness of contract administration activities. All IQA/IDC system development and operations were performed “in-house” by Census Bureau personnel. Outsourcing was considered, but deemed to be too risky with respect to protecting Title 13 data. The possibility of jeopardizing the integrity of the Census Bureau “in the eyes of the public” posed an enormous risk. Contractors were only used to conduct penetration testing. Some limited contracts were in place with a systems manager and with Internet Service Providers (ISP).

5. RECOMMENDATIONS

This section synthesizes the findings from above and highlights opportunities for improvement that may apply to the Census Bureau's future system development activities. The recommendations reflect insights from Titan/SRD analysts as well as opinions regarding “lessons learned” and internal “best practices” that were conveyed by Census Bureau personnel during interviews.

5.1 Reliance on in-house expertise

Historically, the Census Bureau has relied on in-house expertise to develop decennial systems. The success of IDC was due in large part to the expertise and dedication of a single person who had extensive knowledge of computer technology and census business processes. This individual wore several “hats” (designer/developer and content coordinator/programmer) and worked extremely long hours to support both the development and operation of the system. Because of time constraints imposed on the project, several other Census Bureau personnel were assigned to supporting roles as part of the overall management structure.

Recommendation: Extensive reliance on one individual, while it may have been very cost-effective and expedient, introduced significant risks for the Census Bureau. Had that person's availability come into question, IDC might not have been fully developed and could have led to a situation that posed negative publicity ramifications for the agency. It is recommended that the Census Bureau give careful consideration to contingency planning when selecting personnel for high profile system development and operational activities. This should include designation of back-up personnel for critical positions and cross-training of team members. Given the nature, scope, and complexity of future systems, it is recommended that the Census Bureau augment in-house experts with external resources. In addition, usability specialists and technical writer/editors should be included on the design team.

5.2 Public awareness

IDC was not publicized, consequently, the system was underutilized. Only about 66,000 Internet forms were submitted through IDC. These forms represent 66,000 households covering approximately 180,000 persons. Given today's emphasis on e-government Internet-based applications and services, an effective publicity campaign could have led to much greater system usage and IDC could have been a significant Census Bureau "success story."

Recommendation: In spite of being underutilized, IDC clearly demonstrated its capability to provide the public with a web-based questionnaire. This presents a major opportunity for the Census Bureau to automate future data collection activities on a large scale. It is recommended that the Census Bureau publicize the next generation system to ensure greater utilization by the public. Accordingly, system requirements should reflect the need to interact with the American public via Internet technologies.

5.3 Time limitations

As previously discussed, the time constraints impacting this project stemmed from unusual circumstances that were beyond the control of the development team. There were only 18 months to develop and test the system, thereby necessitating deployment of IDC without the benefit of a formal dress rehearsal. This compressed development time essentially ruled out the possibility of acquiring contractor services and led to reliance on in-house resources for both development and operation of the system.

Recommendation: To the extent possible, compressed development schedules should be avoided in that they introduce additional technical, cost, and schedule risks for the Census Bureau. It is recommended that the Census Bureau initiate system planning and requirements definition efforts early on to allow sufficient time for application development, testing cycles, dress rehearsal, program documentation, and user training.

5.4 Requirements methodology

Due to the cessation of early system planning efforts, valuable time and continuity were lost with respect to system planning efforts. Nonetheless, the requirements team did an effective job in defining what the system needed to do. Essential functionality was provided and the user interface was designed to be compatible with various types of Internet browsers. However, the requirements were not prepared using any type of standardized methodology and were not “frozen” to allow programming efforts to proceed without interruption. Typically, constantly changing requirements greatly increase the risk to a project; however, in the case of IDC this risk was reduced because of flexibility built into the design of the code.

Recommendation: Institute an agency-wide set of guidelines for the requirements definition process. This process is central to establishing a sound foundation for any system and would be a critical success factor with respect to any future Internet-based applications. Additionally, requirements should also be “frozen” at some point to allow developers to concentrate on writing code. Constant changes could have easily introduced software “bugs” that may have created an unstable application. Attempts at fixing bugs may inadvertently give rise to logic problems in other areas of the application that are connected through tie ins or “hooks.” Such connections can introduce instability into related modules in the application, which might not be immediately apparent without extensive regression testing.

Additionally, in view of the increasing acceptance of the Internet as a secure medium, it is highly probable that much data collection will be via the Internet in the next decennial census. Accordingly, it would be appropriate to initiate system planning and requirements definition efforts several years before dress rehearsal. For example, current 2010 decennial planning calls for a full systems test by 2006. Any next generation Internet-based systems should be ready for that event. Special emphasis should be placed on ways to efficiently integrate these data with that collected through paper and telephone mediums.

5.5 Integrating system components

The integration of multiple components in support of system design will ensure the successful development of the next generation decennial census applications. These components include subject matter requirements, software development, Internet capabilities, and usability engineering; all crucial to the application development process.

Recommendation: Future system designs should fully integrate user-centered and technology-centered models. This can be accomplished by increasing attention to human factors and ergonomics during the design and development process, which will result in improving the overall experience of the user (e.g., questionnaire respondent). As an example, future development efforts should include usability specialists and technical writers in the planning, design, development, and testing of new applications.

5.6 System development life-cycle (SDLC) methodology

Due to a compressed development schedule, there was little time to formally address the steps in a typical SDLC. Many agencies have promulgated their own SDLC models to provide structure to the system development process.

Recommendation: It is recommended that the Census Bureau implement and promulgate an SDLC model through an agency directive. This model should be flexible enough so that it guides, but does not hamper, development efforts within the dynamic decennial environment. It is also suggested that staff receive training in the application and benefits of the SDLC methodology. This should contribute to a greater awareness of the importance of systematizing the system development process and help to avoid after-the-fact criticism from governmental oversight bodies.

5.7 Project oversight

IDC was a major undertaking for the Census Bureau that utilized unfamiliar technology. Since it was intended for public use, the success of this system was paramount. Moreover, the web-enabled system posed security issues associated with the access. An established group known as the Census Operational Managers (COM) worked in association with the IQT. The COM reviews many system development efforts and evaluates specific issues and/or recommendations that might impact census operations.

Recommendation: The COM worked together with the DMD to monitor the progress of IDC development efforts. In cases where major IT initiatives are undertaken, it is recommended that both DMD and the COM become involved early in the project life-cycle to ensure that the requirements definition process is performed according to a standardized methodology. This oversight will contribute to the development of a solid foundation for the system. Given the unprecedented nature of the IDC and the potential high risks associated with deploying new technologies, high level oversight and support was essential to address budgetary, contractual, and technical issues that may weigh on system development efforts. In general, the use of an oversight body is considered a “best practice;” however, reporting and administrative requirements for each program should be structured to minimize any adverse impacts on the design and development effort. Any project control structure should be flexible enough to encourage technical innovation.

5.8 Form replication

A key requirement for IDC was that the short form displayed on screen replicate the actual paper form as closely as possible. Although technical limitations precluded an exact replica of the paper form, the design team needed to ensure that visitors to the web site would be reassured that all features, forms, and resources were official Census Bureau instruments. This approach likely engendered user confidence. In this regard, the most frequently accessed topic on IDC was “Protecting Personal Information,” which is a strong indication that privacy concerns were a very real factor for users.

Recommendation: Future designs for any web-based applications should consider the confidence factor and establish requirements to make the site and resources appear as official as possible, without compromising usability and ease of navigation.

5.9 System sizing

Since there was no precedent for an Internet-based system of this size, IDC was essentially over-engineered for capacity and redundancy. From the start, it was designed to be “rapidly expandable” (i.e., on-the-fly, without downtime) to meet volume requirements. Due primarily to the lack of publicity about the system, the anticipated traffic levels did not materialize for Census 2000. Quite the opposite could be true for the next decennial census.

Recommendation: Given the strong likelihood of a ubiquitous computing environment during this decade, it seems reasonable to assume that a significantly larger volume of forms will be submitted via the Internet through an IDC-like system in the 2010 decennial census. Volume will, in part, be impacted by the Census Bureau’s public awareness campaign (Refer to 5.2). Accordingly, capacity and redundancy requirements will likely increase greatly, conceivably into the millions of forms, and the Census Bureau will have to base the sizing of the next generation system from an entirely different set of capacity and redundancy metrics.

Before system sizing is determined, the issue of overall Internet usage by the public should be addressed. Factors such as response capability, technical approach, and privacy issues must be evaluated. Currently, the Computer Assisted Survey Research Office (CASRO) has studies in progress to examine this perspective.

5.10 Census 2000 Dress Rehearsal

Usability testing had been performed on IDC as well as external reviews by the National Security Agency (NSA) and the National Institute for Standards and Technology (NIST). However, given management’s concerns over security issues, a formal dress rehearsal would have served to further validate the adequacy of the system’s security provisions.

Recommendation: Dress rehearsal is an opportunity to simulate the operational environment for an application and to identify/correct any shortcomings. In order to maximize system readiness, the application(s) must be in a near-finished state. It is recommended that, to the extent possible, Census Bureau schedule development activities so that ample time is allowed for dress rehearsal.

5.11 Security

The IDC instrument required a minimum level of encryption for encoding data on a user's browser, and was used in conjunction with https/secure html protocol⁴. The system checked to see if a minimum level of encryption was being used on a user's machine, and if not, encouraged the user to download a browser version that supported the recommended level of encryption before proceeding to fill out the form. The user also received a warning that, if the user opted to use a security level lower than 128-bit encryption, they would do so at their own risk. The user was also advised about the option of submitting a paper form. In short, the instrument automatically tested the user's encryption level and made appropriate security warnings and recommendations.

Recommendation: IDC's ability to evaluate browser security levels and alert users to the hazards of insufficient security was a “best practice” and clearly showed Census Bureau's commitment to protecting Title 13 data. This commitment probably served to reassure site users that their personal data were being protected. It is recommended that the Census Bureau continue the practice of automatically testing browsers in this fashion for security compliance.

⁴ Secure Hypertext Transfer Protocol encrypts and decrypts user page requests as well as the pages that are returned by the web server.

References

Census 2000 Systems Architecture, Version 2.0, September 2000.
Section 7, pages 7-1 to 7-7.

Unrestricted Summary of Program Master Plan (PMP) for Internet Questionnaire Assistance Data Collection Operations, October 28, 1999, pages 1-3.
Author(s): David Coon.

Draft Operational Assessment of Census 2000 Internet Questionnaire Assistance and Data Collection Operation, May 11, 2000.
Author(s): David Coon.

Considered Alternatives to the Currently Planned Methodology of Collecting Census 2000 Forms on the Internet, June 11, 1999. Author(s): Cary Bean.

Program Master Plan (PMP) for Internet Questionnaire Assistance Data Collection Operations, August 23, 1999. Author(s): David Coon.

2000 Decennial Census Internet Data Collection (IDC) Beta Test Summary Report, March 29, 2000. Author(s): Frank Bush.

Study Plan for A.2.b; Census 2000 Operational Summary of Internet Data Collection (IDC), June 12, 2001. Author(s): Erin Whitworth.

Study Plan for A.2.c; Census 2000 Internet Web Site and Questionnaire Customer Satisfaction Surveys, June 9, 2001. Author(s): Courtney Stapleton.

Participants

Edward 'Cary' Bean Jr	Systems Support Division 3-1222, +1.301.457.1926 CBean@info.census.gov
David Coon	Computer Assisted Survey Research Office Suitland House, +1.301.457.1415 David.A.Coon@census.gov
Erin Whitworth	Decennial Statistical Studies Division 2-2228, +1.301.457.8024 Erin.M.Whitworth@census.gov
Rose Cowan ^{5,6}	Population Division 3-2372, +1.301.457.2408 Rosemarie.C.Cowan@census.gov
Elizabeth Murphy ^{5,6}	Statistical Research Division 4-3224, +1.301.457.4988 Elizabeth.D.Murphy@census.gov
Danny Burkhead ^{5,6}	Decennial Systems and Contracts Management Office 2-2321, +1.301.457.4133 Danny.L.Burkhead@census.gov
Courtney Stapleton ⁵	Planning, Research, and Evaluation Division Mod B, Building 2, +1.301.457.4142 Courtney.N.Stapleton@census.gov
Ed Gore (DMD Program Support)	Decennial Management Division 2-2012, +1.301.457.3998 Edison.Gore@census.gov
Agnes S Kee (DMD Program Support)	Decennial Management Division 2-1422, +1.301.457.4223 Agnes.S.Kee@census.gov

⁵ Individual not interviewed for this study.

⁶ Individual identified as a participant after development of the Study Plan.