1. Introduction

1.1 Purpose of This Document

This document describes the requirements specifications for the American FactFinder system. The project is sponsored by the U.S. Bureau of the Census, and is the intended audience of this document.

1.2 Scope of This Document

This document is divided into seven sections. It provides an introduction and overview of the American FactFinder and Census Bureau, a general description of the system and its users, functional requirements in detail, interface requirements of how the software interfaces with users, performance requirements, design constraints, and other non-functional requirements.

The requirements set forth in this document were determined through interviews with the staff of the Census Bureau and requests from the general public user community.

1.3 Overview

The Census Bureau’s American FactFinder is a Web-based data dissemination system that accesses data from the 1990 Decennial Census, Census 2000 Dress Rehearsal, the American Community Survey, and 1997 Economic Census. Additional data sets will be added in the future. The American FactFinder will permit users to submit queries to create custom tabulations to satisfy their own objectives. In accordance with Title 13 of the U.S. Code, the Census Bureau is obligated to protect confidentiality of survey respondents. Any query results created by external users will have to pass disclosure rules developed by the Bureau to protect confidentiality.

1.4 Business Context

The Census Bureau is a government agency that generates statistical information based upon censuses and surveys of individuals and establishments in the United States. The Census Bureau is required by Title 13 of the U.S. Code to protect the
confidentiality of the respondents to these censuses and surveys. This law prohibits revealing identifiable information about any person, household or business.

The Census Bureau’s website lists the following as its mission, goal and organizational structure.

**Mission** - To be the preeminent collector and provider of timely, relevant, and quality data about the people and economy of the United States.

**Goal** - To provide the best mix of timeliness, relevancy, quality, and cost for the data we collect and services we provide.

The Census Bureau is headed by a Director, assisted by a Deputy Director and an Executive Staff composed of the associate directors. The Bureau has twelve regional offices with additional processing centers set up temporarily for the decennial censuses.

### 2. General Description

#### 2.1 Product Functions

The unique function of the American FactFinder is that users can create queries to obtain only the data they need for their specific purpose. To protect confidentiality, the American FactFinder does not allow the user to construct all possible tabulations. There will not be output of individual records and only the recoded or the modified version of sensitive variables on the files will be used in the results. This limitation is achieved through two filters, the Query Filter and the Results Filter. The query filter will restrict queries that are permitted to access the full microdata files and the results filter will restrict the detail of tables generated.

The Census data that American FactFinder will access are divided into two categories: the Decennial Public Use Microdata Samples (PUMS) and the Decennial full microdata files. The sample files were drawn from the full microdata files, and are safer, from a confidentiality perspective, because of this sampling. When queries are posted against the full microdata files, the system will use rules to determine the level of detail of the attributes chosen by the user. For example, for a given geographical area, a query including the attribute Age might return respondent ages in 5-year groups (“recoded age”) for the full microdata files, but in 1-year groups for the sample files.

The user interface will permit the user to create custom tabulations. The user interface will be designed so the user cannot design a query that would violate certain limitations such as:
Levels of Geography – the block is the lowest level of geography from 100% data and the tract is the lowest level of geography permitted from sample data. Split blocks or tracts are not permitted.

Number of Table Dimensions – The user interface permits only 3 table dimensions (page, column and row).

Total Population per Geographic Unit – Summary files will be used to determine population size. This will indicate whether a population is too small to pass the query filter.

In addition, the query filter also restricts the use of sensitive variables such as race, Hispanic origin, or group quarters, cost of electricity, gas, water, fuel, property taxes, property insurance cost, mortgage payments, condo fees/mobile home costs, gross rent, selected monthly owner cost, household/family income, individual income types.

If the query satisfies the conditions of the query filter, and if the query filter further determines that the query results stand a reasonable chance of passing the results filter, the query is sent from the external server outside the firewall to the internal server inside the firewall. The internal server will be configured to only accept requests from the American FactFinder external server and only return results to the American FactFinder external server.

The Results Filter will limit tabulations presented to the user so as to not disclose information that could identify an individual or business. The resulting tabulation must meet certain criteria to prevent the release of sparse tabulations that could lead to disclosure. If the tabulation does not meet the criteria of confidentiality, the user will receive a message stating that the request cannot be completed for confidentiality reasons.

The Census Bureau will also require users to register before using the system. Before accessing American FactFinder, users must agree not to attempt to manipulate the data to produce tables that were previously denied for confidentiality reasons. Users will also be notified that their use is being monitored and any misuse will result in terminating their registration.

2.2 Similar System Information

The American FactFinder interfaces with data from the 1990 Decennial Census, Census 2000 Dress Rehearsal, the American Community Survey, and 1997 Economic Census. It performs its own processing to determine whether a query is likely to pass the results filter tests before submitting the request. After the results are returned it verifies that the results do not violate confidentiality. It interfaces with the internal database where the data is stored and the queries are processed.

Currently, census data from the Bureau can only be obtained using pre-defined fixed queries. Commercial applications that permit importing Census data and defining
queries do exist. However, these applications are not web based. To reach a large portion of the public, the Census Bureau requires that American FactFinder be available through the Internet.

The nations of Iceland and the Netherlands have similar systems, but on a much smaller scale. This effort appears to be the first that will make available large datasets containing sensitive personal information.

### 2.3 User Characteristics

American FactFinder will be used by a wide spectrum of both internal and external users. These are the probable users of American FactFinder, in decreasing order of expertise.

#### Professional Users

These are users who are technically savvy and often use Census data. They are familiar with using Census data to perform statistical analysis. Users such as Census Bureau staff and researchers fall into this category.

#### Power Users

These are users who are technically savvy and somewhat familiar with Census data. Examples could include data analysts.

#### Average Users

These are users who are computer literate but not that familiar with Census data. Students enrolled in higher education programs would typically fall into this category.

#### Casual Users

These are users who range from being somewhat familiar with a computer to extremely computer literate, but they are not familiar with Census data. A large portion of the general public falls into this category.

All users will have access to the Internet and have a computer capable of using a web browser.

### 2.4 User Problem Statement

The Census Bureau compiles large amounts of data from its surveys and census. The data is delivered to the public as tables in printed reports or through data sets purchased from the Bureau. These tables are pre-defined, forcing the user to further manipulate the data to achieve the results needed for their use. Often users did not
understand the data presented to them or how to use it. The Bureau needs to improve access to their data and make it easy to use. Users need a way to easily define a custom data set and quickly retrieve the results.

2.5 User Objectives

The American FactFinder will protect security and confidentiality required by Title 13.

The American FactFinder will improve access to statistical data and make the data more accessible.

The American FactFinder will use the Internet to reach the user community.

The American FactFinder will support a wide range of internal and external users, from novice to expert, with minimal restrictions on end-user hardware platform. Users should be able to access the system using a Netscape 4.0/Internet Explorer 4.0 or higher compatible browser.

The interface of the American FactFinder will be easy to use and understand. The interface will be easy to navigate. It should be obvious to users how to create a query by simply looking at the interface or after reading a brief online tutorial.

The interface will be accessible by disabled users by providing capabilities for technologies such as screen readers.

The American FactFinder will allow users to create custom queries and receive results in an agreed amount of time depending upon the complexity of the query. The results can be displayed for printing or downloaded in several data formats.

American FactFinder will be scalable to allow for future enhancements, additional data sources, and increasing numbers of users.

The results from the custom defined queries will be accurate.

2.6 General Constraints

The American FactFinder will not violate confidentiality requirements set forth by Title 13 of the U.S. Code.

The American FactFinder will run on Web application servers that are outside the firewall. It will send requests to the internal server inside the firewall. The internal server will be configured to only accept requests from the American FactFinder external server and only return results to the American FactFinder external server.
The American FactFinder must be accessible by end-users using different combinations of operating system platforms and browsers. This includes Windows, Macintosh, Unix, and Linux.

The American FactFinder must support the HTTP and File Transfer (FTP) protocols.

The time from when the query is submitted to the response is returned shall not exceed twenty seconds to 7 minutes, depending upon the complexity of the query. The Census Bureau and design staff will mutually agree upon exact guidelines determining the time limits for each complexity level. Complexity levels will be determined by number of joins required, number of variables, and size of tables accessed in the query. If the time estimated to complete the query would exceed the limits, the user will be notified and given the opportunity to cancel or modify the request.

3. Functional Requirements

This section lists the functional requirements in ranked order.

1. Create or modify a custom query.
   a. Description

   Users define a custom query through the browser. The user interface will be designed, so far as possible, so the user cannot design a query that would be denied for confidentiality reasons. If the user creates a query that is syntactically incorrect, a message is displayed directing the user to the Help section on creating queries. If the query is syntactically correct, it is sent to the query filter. If it passes the query filter, it is sent from the web server to the internal server; if not, a message is returned to the user indicating that the query could not be answered for confidentiality reasons. Before the query is sent to the internal server, the processing time is estimated. If it is believed the query will take longer than the agreed upon thresholds, the user will be warned and given the option to modify or cancel the request.

   b. Criticality

   This is the primary purpose of the American FactFinder; therefore it is critical to the system’s success.

   c. Technical Issues

   The user interface must be designed to limit the possibility of the user creating a query that could compromise confidentiality. The query filter must be able to detect whether the query is in correct form before submission. Execution time must be computed before the query is submitted.
2. Run a query.
   a. Description

   Once a valid query is submitted, American FactFinder must be able to execute the query. The query is first sent to the database, and the results are sent to a results filter. The Results Filter will limit tabulations presented to the user so as to not disclose information that could identify an individual or business. The resulting tabulation must meet certain criteria to prevent the release of sparse tabulations that could lead to disclosure. If the tabulation does not meet the criteria of confidentiality, the user will receive a message stating that the request cannot be completed for confidentiality reasons. If the results filter is satisfied, the query result is returned to the user.

   b. Criticality

   This also is the primary purpose of the American FactFinder; therefore it is critical to the system’s success.

   c. Technical Issues

   Rules to determine whether query results will be returned to the user must be designed and tested to ensure confidentiality.

   d. Risks

   Maintaining confidentiality creates a high risk for this requirement. Techniques used by the Bureau in the past, such as data swapping and recoding of sensitive variables that could disclose confidential data have proved successful in delivering data to the public. However, these were pre-defined tables that could be scrutinized by humans before distribution. The online system analyzes each query individually. Users could create multiple queries and subtract tabulations to obtain confidential information. Logging
the SQL generated for each request will aid in monitoring and analyzing the
effectiveness of the rules the results filter utilizes.

e. Dependencies with other requirements

This requirement depends upon the ability of the user to create a query and the
ability to log requests.

3. Display the results.

a. Description

The result of the executed query must be displayed to the user. If the query
did not execute successfully, a message indicating the reason for failure is
displayed to the user. If user intervention can correct the error, the user is
directed the proper Help section.

If the query executed successfully the results are displayed on screen as a
table. Users are presented with options to print, download, or modify this
query.

b. Criticality

It is critical that results be displayed to the user.

c. Technical Issues

The original query must be saved in order to permit the user to refine or
modify it.

d. Risks

No risks are involved in displaying the results.

e. Dependencies with other requirements

The results are received from the executed query. This requirement also
interacts with the Help Function in the case of an error, ability to create a file
to be exported, and creating and modifying query requirement.

4. Export the results as a file that can be downloaded by the user.

a. Description

Upon successful execution of a query the user should be able to download a
data file. File formats should include comma or tab delimited ASCII text,
Excel spreadsheet, SAS data set, and SPSS data set. Before downloading a file, users should be informed of the file size.

b. Criticality

Exporting the results to a file is critical to provide the intended functionality of the system.

c. Technical Issues

The system must support the File Transfer Protocol (FTP).

d. Risks

File size could become large in complex queries. A compression utility could be used to reduce file size and download times.

e. Dependencies with other requirements

This requirement is dependent upon a query successfully executing.

5. Log queries and response times.

a. Description

Each query processed will be logged and stored in a database for security and monitoring reasons. Information logged will be

   i. The SQL generated
   ii. Whether the query executed successfully
   iii. Response time
   iv. Size of output
   v. Number of active users at the time of the query
   vi. The ID of the user submitting the query.

b. Criticality

This requirement is critical to the security of the application to monitor what data returned to the user.

Technical Issues

Logging could affect the responsiveness of the system.

c. Risks
Logging is a common function in applications. This requirement poses little to no risk.

**d. Dependencies with other requirements**

This requirement does not depend upon other requirements. However, it is a tool in monitoring not only the security of the results filter but also the non-functional requirement of responsiveness.

6. **Provide online help and tutorial.**
   
   a. **Description**

   Extensive on-line help will be provided to the user. This will consist of explanatory text near each field prompting for user input, a FAQ section, searchable and indexed help topics, and short tutorials. The Help section will cover topics explaining the functionality of the system, constructing queries and their syntax, how to use Census data, and a troubleshooting section. The Help topics will be indexed to provide context sensitive help from error messages received in other parts of the application or by clicking the Help button on a screen. Short tutorials that can be completed in under 10 minutes each will cover topics such as, Getting Started, How to Create a Query, and Using Advanced Features. Tutorials can be completed online or downloaded in Adobe Acrobat (PDF) format.

   b. **Criticality**

   While this function is not required for the system to function, it is important to increase usability of the system. It is also customary for sophisticated applications to provide a help facility.

   c. **Technical Issues**

   GUI elements in the user interface must be linked to specific help sections to provide context sensitive help to the user.

   d. **Risks**

   There is no risk associated with this requirement.

   e. **Dependencies with other requirements**

   This requirement interacts with all other aspects of the system to provide assistance to the user.

7. **Provide error messages alerting the user when something has gone wrong and provide the user with optional course of action.**
a. Description

When the American FactFinder cannot return the results of the query, the user will be presented with an error message and options of other courses of action. Examples of messages include:

**Tabulation cannot be released for confidentiality reasons.**
If the results filter determines the tabulation would disclose personal information the user will be notified and presented with the option of canceling or modifying the query.

**Tabulation will exceed typical time requirements.**
If the query filter determines the tabulation will take an extraordinary amount of time to process the user will be given the option to proceed with the query, cancel the query, or modify the query, before the query is submitted for processing.

If user interaction can correct an error, the user will be prompted for action. If the error is non-recoverable, such as the database is unavailable, the user will be presented with a message communicating what has happened.

b. Criticality

Although this is important, it is not a critical requirement.

c. Technical Issues

The user-defined query must be saved before issuing the error message. This permits the user to modify or submit the query if they choose without needing to start over.

d. Risks

Most error conditions can be caught in order to provide the user with feedback. There is little to no risk fulfilling this requirement.

e. Dependencies with other requirements

This requirement does not depend upon any other requirements.

4. Interface Requirements

This section describes how the software interfaces with users for input or output.

4.1 User Interface
The user interface is a Graphical User Interface (GUI). The interface will be easy to navigate and pleasing to the eye. All pages will have a common look. Major elements will be consistently placed on each page. Frequently accessed functions will be conveniently placed in a sidebar on the left side of the menu on each page. Navigation buttons to the main page, search, help and FAQs will be placed at the top of each page.

Simple queries can be constructed using drop down menus. Top-level menus will allow the user to select from broad categories of information. Subsequent menus allow the user to narrow the selection, such as limiting the information to the state or county level. This section will be labeled “Start”. An example screen is provided below.

Below the “Start” section, is the American FactFinder Census 2000 Query Builder. More complex and customized queries can be constructed using the Query Builder. The user is presented with a list of tables, pre-defined queries, and reports in the left box. Selecting a table will display a list of variables available in the middle box. Users can build-up a query by selecting tables, variables, and using the logic operator buttons above the boxes. Users can also select from a list of pre-defined queries and reports. A Help button is available to provide online help for using the Query Builder and on the syntax of a query. The Query Builder is shown in the graphic above, as well as a detail view below.
4.1.1 Diagnostics or ROM

The tools provided in the IDE of the programming language will be used for diagnostics.

4.2 Hardware Interfaces

The American FactFinder resides on an external web server outside the firewall. It interfaces with the internal server, which contains the actual census data inside the firewall. The internal server will be configured to only accept requests from the American FactFinder external server and only return results to the American FactFinder external server.

The end-user application will run on any platform that a web browser runs on.

5. Performance Requirements

Query response times will vary based upon the complexity of the query and will take no more than twenty seconds to 7 minutes, depending upon the complexity of the query. The Census Bureau and design staff will mutually agree upon exact guidelines determining the time limits for each complexity level. Complexity levels will be determined by number of joins required, number of variables, and size of tables accessed in the query. If a query is expected to exceed the time limit, the user will be informed before the query is executed.

The system will support access by 100 simultaneous users without a degradation of speed or accuracy.

6. Design Constraints
6.1 Standards Compliance

The system will be designed utilizing industry standards in compliance with agreed upon federal design standards. GUI interfaces will be designed utilizing standard Windowing environments and standards. Data export formats will utilize standard file outputs, such as comma delimited text, tab-delimited text, SAS data sets, and SPSS data sets.

6.2 Hardware Limitations

The system will operate on the current server hardware utilizing the current networking configuration.

6.3 Security Constraints

The American FactFinder will run on Web application servers that are outside the firewall. It will send requests to the internal server that processes the queries and interacts with the actual database inside the firewall. The internal server will be configured to only accept requests from the American FactFinder external server and only return results to the American FactFinder external server.

Data shall not be stored outside the Census Bureau’s firewall.

7. Other Non-Functional Attributes

7.1 Security

Confidentiality requirements are a top priority for the American FactFinder. Tabulations that could violate confidentiality will not be presented to the external end user. Accesses to confidentially sensitive will be written to a security audit log.

7.2 Reliability

The system will be designed for availability and be able to recover from system failure in a realistic agreed upon time. The system will be available to users 24 hours a day, 7 days a week.

7.3 Maintainability

The system will be built using modular components to allow easy modifications as additional requirements are identified.

7.4 Portability

The system shall function identically on all user platforms that a web browser runs on.
7.5 Extensibility

Additional data sets should be supported as required.

7.6 Usability

The system must be easy to use with minimum training. It will be easy to navigate, create and run queries. User Interface will be accessible by disabled persons.

7.7 Responsiveness

Query response times shall not exceed a mutually agreed to limit of twenty seconds to 7 minutes, depending upon the complexity of the query. The Census Bureau and design staff will mutually agree upon exact guidelines determining the time limits for each complexity. Complexity levels will be determined by number of joins required, number of variables, and size of tables accessed in the query. If a query is expected to exceed the time limit, the user will be informed before the query is executed. Users will be given the option to cancel, modify, or run the query.

7.8 Accuracy

The data returned must be accurate.