

NATIONAL CENTER OF EXCELLENCE FOR AVIATION OPERATIONS RESEARCH
University of California at Berkeley

Development of a National Database of Air Passenger Survey Data

Conceptual Design of a National Database of Air Passenger Survey Data

Geoffrey D. Gosling

NEXTOR Working Paper
NWP-2005-001

March 2005

Aviation System Consulting, LLC
805 Colusa Avenue
Berkeley, CA 94707-1838

Preface and Acknowledgments

This report documents research undertaken by the author for the National Center of Excellence for Aviation Operations Research (NEXTOR), under Federal Aviation Administration Research Award DTFA01-01-C-00029 (DO#0018). The preparation of this report was performed under a subcontract from the University of California to Aviation System Consulting, LLC of Berkeley, California, under the direction of Prof. Mark Hansen of the University of California at Berkeley as Principal Investigator for the project. Any opinions expressed herein do not necessarily reflect those of the Federal Aviation Administration or the U.S. Department of Transportation.

Table of Contents

	Page
Preface and Acknowledgments	ii
Table of Contents	iii
Executive Summary	iv
1. Introduction	1
2. Proposed Operational Concept for a National Database	3
3. Initial Estimate of Implementation Costs	10
4. Sources of Air Passenger Survey Data	14
5. Integration of Air Passenger Survey Data from Multiple Sources	18
6. Plans for the Remainder of the Study	22
References	24

Executive Summary

The information derived from air passenger surveys forms an essential input to the airport planning process. The development of improved models of air travel demand at a market and national level, as well as many other analyses that form the basis of local, regional and state airport system planning, particularly studies of airport ground access and egress travel, require information on the characteristics of the air passenger trips that comprise the existing pattern of air travel. While detailed information on the number of these trips, together with the airlines used, fares paid and trip itineraries, is reported to the U.S. Department of Transportation by airlines, very little is known about the purpose of these trips and other characteristics of the air parties involved, such as their income, occupation, travel party size, day and time of travel, trip duration, and so forth. None of this information is currently reported by the airlines and indeed much of it is not even known by them. The only practical way to get most of this information is to ask the travelers themselves through an air passenger or household survey.

Such surveys are conducted routinely by airport authorities, regional planning agencies, and other organizations, but the results are not readily available outside the sponsoring organization. Furthermore, since these surveys are performed at different times, using different survey instruments and different sample sizes, combining the results of such surveys can be technically challenging. There is a need to improve access to the data collected in the course of previous air passenger surveys, both to support future airport planning activities as well as to facilitate research to improve air passenger survey methodology and develop improved guidance on the design and conduct of future surveys

This working paper has been prepared as part of research undertaken by the National Center of Excellence for Aviation Operations Research (NEXTOR) for the Federal Aviation Administration (FAA) to research the feasibility and potential benefits of the development of a web-based national database of air passenger travel characteristics from air passenger surveys conducted by airport authorities, regional planning agencies and others. The working paper defines the conceptual design for such a national database, as well as develops an initial estimate of implementation costs and identifies potential sources of air passenger survey data. It also discussed some of the issues involved in combining air passenger survey data from multiple

sources and reviews previous efforts to do so, as well as describes the work to be performed in the remainder of the study.

Proposed Operational Concept

The wide range of potential uses of a national database suggest that the database needs to provide access to both the survey response data itself as well as meta-data about each survey, including details of the survey methodology, sampling strategy, question wording and response coding, in addition to such information as the dates of the survey and the overall sample size. Given the widespread use of the Internet (World Wide Web) to provide access to information and distribute data, it is envisaged that the proposed database would be web-based. Web pages would provide information on the survey data available and links to the survey data files that can be downloaded from the site.

The database would need to provide four different types of information and data for each air passenger survey:

1. Background information on the survey, including the dates when it was performed, the sample size, the survey methodology, *etc.*
2. Reports by or for the survey sponsor presenting the survey results and discussing the survey methodology, or links to those reports on the survey sponsor's website.
3. Metadata describing the content of survey response data. This would include the definitions and specifications of the variables in the data file and the definitions of any codes used for each variable.
4. The survey response data, formatted as a data table with each response as a row and each variable as a column.

The planned structure of the website comprises a series of interlinked web pages that lead the user to a web page for each specific survey. The home page will provide a brief overall description of the database and navigation buttons in the left panel of each page will guide the user to the various parts of the website.

While it is expected that much of the information in the database will be publicly available, there may be a need to restrict access to some of the data to certain users. Owners of the data from a given survey (generally the organization sponsoring the survey) may wish to

restrict access to some or all of the data to approved users. This suggests that data access restrictions could take two forms:

1. Limiting access to the entire data for a given survey to approved users;
2. Limiting access to selected fields of a given survey to approved users.

Both requirements could be met by limiting access to selected fields for a given user. In the first case, an approved user would simply be given access to all fields.

In addition to providing access to the source data for each survey, it would be desirable for the database to also provide value-added enhancements, such as support for integrating information from multiple surveys. Some thought also needs to be given to simplifying the process of updating and maintaining the database. While it would be possible to have the web pages for each air passenger survey written using standard web publishing software, it would be more efficient to have the relevant information stored in a data table for each survey and have the web page generated from the information in the data table. This not only would reduce the amount of work involved in adding a new survey to the website, but offers the possibility of developing utility software that would allow the organization responsible for the survey to enter all the relevant data using the website.

Implementation Cost

An initial estimate of the likely resources required to implement a national database of air passenger survey data was developed based on information provided by staff at the National Aviation Safety Data Analysis Center (NASDAC), a web-based aviation safety database maintained by the FAA Office of System Safety. The current phase of the project will result in a prototype website hosted on a web server at the University of California. However, to implement a national database on an on-going basis will require additional resources to update and maintain the database, to transition the database to a production location and provide on-going technical support, and to provide user support. It is estimated that the database transition would require about three person-months of contractor effort while subsequent database maintenance and support could require about 4 person-days per month of FAA or contractor staff time, or about 20 percent of a full-time staff position.

Sources of Air Passenger Survey Data

There are three principal sources of air passenger survey data that could be included in a national database:

1. Surveys performed at specific airports by the airport authority
2. Surveys performed at specific airports by state or regional transportation agencies or other organizations
3. Household travel surveys that include questions about air travel.

While air passenger surveys are typically performed at airports, more general household travel surveys that ask about the respondents' recent travel behavior are undertaken by various agencies, and in some cases trips by household members reported in these surveys include air trips. The data for these air trips can be extracted from the survey data for all types of trip and presented in a comparable format to air passenger surveys performed at airports.

Integration of Survey Data from Multiple Sources

The full value of having access to data from multiple air passenger surveys lies in the ability to integrate the information from more than one survey. This is particularly important in performing analysis of air passenger travel patterns in multiple airport regions, or performing corridor studies, where air passenger ground travel patterns are needed at both ends of the market. However, the integration of information from multiple air passenger surveys involves a number of difficult issues that need to be carefully addressed if the resulting data is not to become misleading or invalid. These include inconsistency in question wording and coding and making adjustments for different sampling rates.

In spite of the importance of being able to combine information from multiple surveys, there appears to be very limited discussion in the literature of the issues involved and how this can best be done. One recent report that specifically addresses the issues involved in combining information from multiple surveys has been prepared as part of a larger model development study for the Southern California Association of Governments (SCAG). Similar issues have been addressed in an on-going regional airport planning study for New England. In the case of the New England study, the difficulties involved in combining information from multiple existing air passenger surveys at airports in the region proved so intractable, that it was decided to undertake a new survey using a consistent methodology and questionnaire. However, in the

case of the Southern California study, it proved feasible to combine the survey information from multiple airports to provide data on total air party trips originating in an area that accounted for a little under half of the air passenger trips in the SCAG region in order to support the development of a regional air passenger trip generation model.

Plans for the Remainder of the Study

The remainder of the study consists of three activities:

- Perform a proof of concept study that will assemble air passenger survey information from a sample of past surveys, incorporate these into a prototype web-based database, and perform an analysis of the resulting survey data to demonstrate the utility of having access to the results of multiple surveys.
- Undertake a survey of potential users of the proposed national database of air passenger survey information to assess the support for and perceived usefulness of the proposed database.
- Perform the necessary research to define the necessary technical and organizational procedures that would be needed to implement an archive of air passenger survey data on a national level, and develop detailed estimates of the resources and costs involved in doing so.

The planned proof of concept study serves two purposes. The first is to demonstrate the utility of providing web-based access to air passenger survey data by allowing potential users to access data for a sample of recent surveys and by conducting a number of illustrative analyses of air passenger information obtained from multiple surveys. The second is to acquire practical experience in resolving the issues involved in implementing such a database and develop the understanding necessary to prepare a detailed implementation plan for a national database.

The survey of potential users of the proposed database will include a representative sample of airport authorities, state and regional planning agencies involved in airport or aviation system planning, aviation or airport planning consultants, relevant research organizations, and other potential users of the information. In addition to identifying the type of information most likely to be of use to different organizations, and their anticipated willingness to contribute air

passenger survey data to the proposed archive, the survey will also develop better information on the frequency with which air passenger surveys are conducted by various types of organization.

In addition to preparing a detailed implementation plan, the third activity will include the development of conceptual functional designs and detailed functional specifications for utility routines to access and combine information from multiple surveys and to facilitate uploading survey data to the archive in standard format.

Conclusion

This working paper has described a conceptual design for a national database of air passenger survey data and has examined many of the issues involved in implementing such a database. The remaining research being performed by NEXTOR as part of the current project will refine these ideas and implement a prototype database that can be used to demonstrate the functionality of such a database and the value of greatly improving the ease of access to this type of data.

1. Introduction

The development of improved models of air travel demand at a market and national level, as well as many other analyses that form the basis of local, regional and state airport system planning, particularly studies of airport ground access and egress travel, require information on the characteristics of the air passenger trips that comprise the existing pattern of air travel. While detailed information on the number of these trips, together with the airlines used, fares paid and trip itineraries, is available from the data reported to the U.S. Department of Transportation by airlines, very little is known about the purpose of these trips and other characteristics of the air parties involved, such as their income, occupation, travel party size, day and time of travel, trip duration, access/egress mode use, and so forth. None of this information is currently reported by the airlines and indeed much of it is not even known by them. The only practical way to get most of this information is to ask the travelers themselves through an air passenger or household survey.

Such surveys are conducted routinely by airport authorities, regional planning agencies, and other organizations, but the results are not readily available at a national level. Furthermore, since these surveys are performed at different times, using different survey instruments and different sample sizes, combining the results of such surveys is technically challenging, even after the hurdle of actually locating and obtaining the data has been overcome. Some of the challenges involved in undertaking air passenger surveys and making use of the results of such surveys were discussed at a recent workshop on Air Passenger Survey Methodology, which was held as part of the Transportation Research Board Annual Meeting in January 2004. Among the issues addressed at the workshop were the need to be able to combine the results of surveys at different airports in order to support regional aviation system planning and a proposed research agenda in air passenger survey methodology in order to develop improved guidance on the design and conduct of future surveys.

In order to address these needs, access to, and the use of, air passenger data would be greatly facilitated by a national archive of air passenger survey results, together with appropriate information on the survey methodology used. The ability to compare the results of local surveys with those conducted elsewhere would provide some incentive to airport authorities and other

agencies to contribute their survey results to such an archive. They would also benefit from the results of research into survey methodology and other airport planning issues that would be enabled by such an archive.

For ease of access to the resulting data, it is envisaged that a national archive would be implemented as a web-based database. Password protection could be provided to control access to selected parts of the database, if this is found to be desirable. However, there are many technical, logistical and cost questions that would need to be answered before such an archive could be effectively implemented. In order to answer those questions, as well as explore the practicalities and usefulness of such an archive, the Airport Technology Branch of the Federal Aviation Administration (FAA) Technical Center, at the request of the FAA Office of Airport Planning and Programming (APP-400), has funded the National Center of Excellence for Aviation Operations Research (NEXTOR) to research the feasibility and potential benefits of the development of a web-based national database of air passenger travel characteristics from air passenger surveys conducted by airport authorities, regional planning agencies and others.

This working paper has been prepared as part of this research. It aims to define the conceptual design for such a national database, as well as to develop an initial estimate of implementation costs and identify potential sources of air passenger survey data. The working paper also discussed some of the issues involved in combining air passenger survey data from multiple sources and reviews previous efforts to do so.

The remainder of this working paper consists of five chapters. Chapter 2 presents the proposed operational concept for the national database. Chapter 3 provides an initial estimate of implementation costs. Chapter 4 discusses potential sources of air passenger survey data that could be included in the national database. Chapter 5 addresses the issues involved in combining air passenger survey data from multiple sources and discusses some previous efforts to do this. Finally, Chapter 6 presents the planned work for the remainder of the study.

2. Proposed Operational Concept for a National Database

Design of a national database of air passenger survey information should take into account the potential users and the uses to which they might want to put the survey data. These potential users include agency staff and their consultants undertaking the following types of study:

- Airport authorities needing information on passenger characteristics for airport planning studies could expand the size of the sample obtained from surveys at their own airport by examining data for passengers traveling to their airport from other airports for which survey data is available;
- Airport authorities interested in the share of a regional market captured by their airport could examine the characteristics of passengers using the different airports in the region;
- Regional and state planning agencies performing airport system planning studies could obtain information on air passenger travel characteristics in the region or state for use in the studies, including trip origins and destinations, airport choice, and airport ground transportation mode use;
- The FAA itself and other federal agencies involved in forecasting the demand for air travel, analyzing air travel patterns, or planning the future development of the national airport system;
- The research community could utilize the air passenger travel characteristics from multiple airports to develop better air travel forecasting models, intercity travel demand analysis models, and airport choice and airport ground access/egress mode choice models;
- The research community could use comparative survey results to undertake research into air passenger survey methodology and question wording to improve the quality of the information obtained from these surveys.

The range of these potential uses of a national database suggest that the database needs to provide access to both the survey response data itself as well as meta-data about the survey, including details of the survey methodology, sampling strategy, question wording and response coding, in addition to such information as the dates of the survey and the overall sample size. Although it would be useful to perform some standardization of question response coding to facilitate comparison of survey results from different survey (for example using the same codes for trip purpose), the database should provide access to the original coding so that this is available for research purposes or to enable users to recode data using a different coding convention.

Since different users are likely to be interested in different surveys or survey data for different airports, it would seem appropriate to maintain the data for each survey as a separate file, so that users could combine data for different surveys as needed. This would also have advantages in terms of managing access to the database, particularly if the data from some surveys are to be restricted to certain users.

Internet Access

Given the widespread use of the Internet (World Wide Web) to provide access to information and distribute data, particularly by the FAA on its various websites, it would seem logical that any national database of air passenger survey information would be based on this technology. Web browsers have the capability to both display web pages and download data files, as well as to support security protocols for exchanging sensitive information. It is therefore envisaged that the proposed database would be web-based. Web pages would provide information on the survey data available and links to the survey data files that can be downloaded from the site.

It is not envisaged that the web server would perform very extensive data manipulation, apart from selection of certain subsets of data, as discussed below. Any more extensive analysis would be undertaken by the users on their local machines after downloading the source data.

Available Data

The database would need to provide four different types of information and data for each air passenger survey:

1. Background information on the survey, including the dates when it was performed, the sample size, the survey methodology, *etc.* This would be presented most easily as a text file.
2. Reports by or for the survey sponsor presenting the survey results and discussing the survey methodology, or links to those reports on the survey sponsor's website.
3. Metadata describing the content of survey response data. This would include the definitions and specifications of the variables in the data file and the definitions of any codes used for each variable. While this could also be a text file, it might be more convenient for subsequent analysis to format it as a set of data tables that could be loaded into statistical or other analysis software.
4. The survey response data, formatted as a data table with the response from each respondent as a row and each variable as a column. This is a standard form in which survey data of this type is generally stored by statistical and database software packages.

Since users of the survey data in the database will typically want to perform their own analysis on the data, it is critical to provide access to the raw survey response data, so that this can be tabulated and analyzed as desired, and not limited by any previous aggregation of the results or manipulation of the results. Incomplete survey responses should be included to the extent that they are available, although flagged as such, since they may contain useful information in the questions that were answered. However, the source data should reflect any data cleaning performed by the sponsoring organization to fix obvious typographical or data entry errors. Where the value for incomplete fields can be inferred from the answers to other questions, or apparently erroneous responses can be corrected from other information, the survey data should ideally include separate variables for both the corrected responses and the original response. This allows a user of the survey data to decide whether to accept the correction.

Format of Data Tables

While it is fairly common for the results of air passenger surveys to be stored using standard statistical software, since these packages often provide the capability to include information on variable definitions and coding within the same file, it would be desirable for a national database to use a more universally available format to download data tables. The most flexible and widely used format for this type of data is a text file using Comma Separated Values (CSV) format for each record. Most spreadsheet, database management, and statistical software can directly import such data into their own internal formats with little or no intervention by in a database management system format, such as Microsoft Access. For those users who intend to store the data on their local machines in the same format as the web server, it may be convenient to provide the option of downloading the files directly in that format.

Restrictions on Data Access

While it is expected that much of the information in the database will be publicly available, there may be a need to restrict access to some of the data to certain users. Owners of the data from a given survey (generally the organization sponsoring the survey) may wish to restrict access to some or all of the data to approved users. For example, an airport authority may have included questions in an air passenger survey on passenger satisfaction with airport facilities and services and not wish this information to be publicly available. Other sponsors of air passenger surveys may wish to limit access to the survey data to approved users.

This suggests that data access restrictions could take two forms:

1. Limiting access to the entire data for a given survey to approved users;
2. Limiting access to selected fields of a given survey to approved users.

Both requirements could be met by limiting access to selected fields for a given user. In the first case, an approved user would simply be given access to all fields.

Implementing access restrictions of this sort has a number of implications for the design of the database. In the first place, it will be necessary for the software controlling the downloading of data files to be able to selectively download approved fields. This in turn means that information will need to be stored on which users have access to which fields in which datasets. In order to avoid excessive computational overhead in those situations where access to the data is unrestricted, it is proposed that data files be classified in two categories. The first

category, unrestricted data, would be available for viewing or download from the web server using standard web browser software in the usual way. Restricted files would appear on the website as a link to a secure page that would prompt the user for a login name and password. This would be checked against the approved access to that dataset, and a temporary file created containing only the approved data. This file name would appear on the web page and could be opened or downloaded in the usual way.

Establishing appropriate security measures and maintaining the information on authorized users could require a significant level of effort, although this may be necessary in order to avoid unduly restricting the content of the database. However, these features might not be implemented in initial versions of the database, but could be added later as the requirements become clearer and resources permit. In any event, security measures for restricted data would need to be tailored to the capabilities of the system supporting the production website.

Structure of Website

The planned structure of the website comprises a series of interlinked web pages that lead the user to a web page for each specific survey. The home page will provide a brief overall description of the database and navigation buttons in the left panel of each page will guide the user to the various parts of the website.

As discussed in more detail later in this working paper, there are two broad categories of air passenger survey data that need to be addressed in the website structure: airport based surveys and household travel surveys. While it is logical to organize airport based survey information by the airport where the survey was performed, household travel survey data will include air travel utilizing multiple airports and is most logically organized by the geographical scope of the survey.

It is therefore envisaged that there will be five major sections in the database:

- Overview of the database and procedures for use. This section will include summary statistics about the surveys included in the database, such as the number of surveys by year, the number of airports represented in the database, and the total number of survey responses by year.
- A listing of all airports for which airport survey data is available in the database, showing the date of each survey for each airport. The airport list

can be sorted by airport name, by three-letter code, or by city. A subset of the list can be displayed for each state or each metropolitan region. These lists can also be sorted.

- Selecting a specific survey from the airport list will display a web page providing description information about the survey in a standard format and links to download survey data and supporting documents.
- A list of household travel surveys for which information is available on the website. This list will be divided into national and regional surveys. The regional survey list can be sorted by region name or by state, or a subset of the list can be displayed for each state.
- Selecting a specific travel survey from the household travel survey list will display a web page providing descriptive information about the survey in standard format and links for additional information. Because the data for the national household travel surveys are available from the BTS website, and air travel data in regional household travel surveys is highly variable, it is not envisaged that the national air passenger survey database would provide direct access to the actual data for the national and regional household surveys, although this could be the focus of future development if it appears to be a useful addition to the database.

Design of preliminary versions of the various web pages will be undertaken as part of the development of a prototype web-based database in the subsequent research task.

Data Integration and Value-Added Enhancement

Since one of the objectives of establishing a national database of air passenger survey data is to facilitate the integration of information from multiple surveys, it would be desirable for the website to provide some support features to simplify this process or otherwise increase the utility of the survey data. One such feature would be a consistent response weight variable across all surveys that would expand survey response counts to equivalent annual passenger counts. This would enable results from two surveys to be directly compared or for responses from two or more surveys using the same variable definitions to be combined.

Another value-added feature would be a set of standard variables and variable response codes, so that data from multiple surveys could be analyzed without needing to recode the responses into a common convention. Since the ability to do this may be limited by the nature of the source data, it may not be possible to do this for all surveys, while it may only be possible for other surveys by using highly aggregated categories.

Database Update and Maintenance

While it would be possible to have the web pages for each air passenger survey written directly as HTML (hypertext markup language) files using standard web publishing software, it would be more efficient to have the relevant information stored in a data table for each survey and have the web page generated from the information in the data table. This not only would reduce the amount of work involved in adding a new survey to the website, but offers the possibility of developing utility software that allows the organization responsible for the survey to enter all the relevant data using the website.

The webmaster would assign a login and password to the appropriate staff at the organization responsible for the survey, who would then enter the relevant data for a new survey on a secure data entry page and upload the associated files. The webmaster would review the information for completeness and any obvious data entry errors, run various data acceptance tests, and after resolving any outstanding issues with the appropriate contact at the organization responsible for the survey, activate the survey data on the website using a custom designed database administration utility program.

3. Initial Estimate of Implementation Costs

The current phase of the project will result in a prototype website hosted on a web server at the University of California. However, to implement a national database on an on-going basis will require additional resources to update and maintain the database, to transition the database to a production location on a server maintained by the Federal Aviation Administration (FAA) or one of its contractors and provide on-going technical support, and to provide user support. This chapter provides an initial estimate of the resources required to do this. This estimate will be refined during the course of the project, as experience is gained in implementing the prototype database.

The estimate of the likely resources required was based on information provided by staff at the National Aviation Safety Data Analysis Center (NASDAC), a web-based aviation safety database maintained by the FAA Office of System Safety, and accessible to the public at <http://nasdac.faa.gov>. The NASDAC data is considerably more extensive than envisaged for the proposed National Database of Air Passenger Survey Data, involving some 22 distinct data sources. However, it involves many of the same issues that will have to be faced by the proposed air passenger survey database, including the need to integrate data from multiple sources and providing support for user analysis of the data.

The NASDAC Data/Information Portal uses four quad Compaq servers running the following three Oracle software components:

- *Oracle 9i Application Server* provides the development tools used to develop the web-based data/information portal. This also provides communications software and provides users with On-line Analytical Processing (OLAP) capabilities using the Oracle Discoverer tool.
- *Oracle Data Base Management System* provides database management functions, advanced text search capabilities, and security functions. It also supports an integrated thesaurus tool that can be customized to use by specialized user groups.
- *Oracle Warehouse Builder* features an integrated metadata repository that provides users with a broad range of information about source data

systems and NASDAC generated data products, as well as providing tools that support automated data extraction, transformation, loading, and data integration.

These software components support a data management strategy referred to as the Advanced Data Architecture (ADA) that provides the conceptual framework and processes required to utilize and maintain the data and information resources that comprise the NASDAC.

Staff support for the NASDAC includes both FAA personnel and contractor support. The annual budget for NASDAC development and operations is about \$4 million per year, divided about equally between FAA Operations (OPS) accounts and Facilities and Equipment (F&E) accounts. Initial implementation of the ADA information technology framework including establishing the hardware and software architecture and importing the various source data involved a one-time cost of about \$1.2 million.

As noted, the NASDAC data sources and activities are considerably more complex than envisaged for the proposed air passenger survey database. Therefore these costs need to be adjusted to reflect the lower levels of effort likely to be required to establish and maintain the air passenger survey database. In addition, hardware and software implementation costs will depend on whether the proposed database is hosted on an existing web server platform or on a new platform developed for the purpose. Since the FAA already maintains multiple websites, including several addressing airport planning data needs, it is assumed that the proposed database will be hosted on one of these platforms. Furthermore, the current research will develop a prototype website. Therefore the implementation costs will only involve the incremental staff support needed to transition the prototype website to the new server environment, to update and maintain the new database, and to provide on-going user support.

Transition Costs

The extent of the costs involved in transitioning the prototype database to a production server will depend on three factors:

1. The extent to which it is desired to modify the web page formats and data presentation to achieve consistency with other FAA web services, in particular those on the production server.

2. The need to enhance the security features of the database to comply with FAA or contractor requirements for information security.
3. The extent to which the underlying air passenger survey data need to be restructured to be compatible with the software architecture of the production server.

These issues will be explored in more detail in later stages of the project. However, for the purposes of the initial estimate of resource requirements, it has been assumed that the database transition would require about three person-months of contractor effort.

Database Maintenance and Support

The tasks involved in on-going maintenance and support of the proposed database comprise four main activities:

1. Adding new data sources to the database as they become available
2. Technical support for the website and supporting software
3. User support
4. Future enhancements and development of the database and website.

The work involved in adding new data sources to the database will depend on the amount of new data added to the database each year as well as the extent to which the organizations sponsoring the surveys can submit the information and survey data in a standard format. Even so, some amount of review, checking and editing will be required to maintain data quality and avoid excessive user problems or questions. It is assumed for the initial estimate of required resources that 25 new surveys will be added to the database each year, and that each survey will require an average of one day of staff time to review, edit and add to the database. The estimate of the number of new surveys is based on each of the top 100 airports in the U.S. performing a new air passenger survey once every five years on average (including surveys performed at those airports by metropolitan planning organizations or other agencies), together with a further five surveys per year at smaller airports.

It is assumed that technical support for the website and supporting software will require about one half day per month, while user support will require an additional half day per month, based on one request per week that requires an hour to resolve and respond. Estimating resources required for further development of the database and website is more open-ended,

since it is not clear what enhancements might be required. However, for the purposes of this initial estimate of resources, it is assumed that this activity could require about 10 person-days per year.

In summary, database maintenance and support could require about 4 person-days per month of staff time, or about 20 percent of a full-time staff position.

4. Sources of Air Passenger Survey Data

There are three principal sources of air passenger survey data that could be included in a national database:

1. Surveys performed at specific airports by the airport authority
2. Surveys performed at specific airports by state or regional transportation agencies or other organizations
3. Household travel surveys that include questions about air travel.

Airport Authority Surveys

These are the most common type of air passenger surveys. Most airport authorities perform some such surveys periodically to gather information on their customers travel patterns and sometimes their perceptions of the quality of the services provided by the airport. The information from these surveys is used in airport planning studies and for airport operational decisions. The size of the sample and the questions included in the surveys varies widely, depending on the interests of the airport authority at the time. One fairly common purpose of such surveys is to gather information on air passenger spending patterns, both in the airport and on their trip generally, for use in planning and managing airport concessions and in preparing airport economic impact studies. Another very common purpose is to gather information on air passenger ground access and egress mode use for airport landside and ground transportation planning.

These surveys are typically conducted in the airport boarding lounge or gate area, since passengers usually spend some time waiting in this area and have the time to answer what may be a fairly extensive number of questions. Such surveys are inherently limited to enplaning (departing) air passengers. Surveying passengers arriving at the airport by air is much harder, since there is no common location where they spend any time waiting and they may be in a hurry to get to their ground transportation and unwilling to take the time to answer questions. Therefore questions about air passenger ground transportation use on arriving at the airport (if asked at all) are typically asked in surveys of departing passengers. While air passengers visiting a region will of course know how they left the airport when they arrived on their trip,

residents who are departing on their trip may not yet have made a decision how to leave the airport when they return.

Less commonly, some air passenger surveys have been performed as in-flight surveys in cooperation with the airlines. While the logistics of doing so are harder, and the airlines may not be willing to cooperate and distribute or collect the surveys on the flights, there are some advantages to this approach. Air passengers generally can take as much time as necessary to complete the survey questionnaire and the sample is not subject to bias resulting from the time before flight departure at which the air passengers arrived at the boarding lounge. However, this approach is limited to self-completed surveys.

Airport Surveys by Other Organizations

Airport surveys may also be performed by state and regional transportation planning agencies or other organizations, generally in association with the airport authority at the airports involved.

A number of metropolitan planning organizations (MPOs) in multi-airport regions have sponsored air passenger surveys at the major commercial airports in their region as part of the regional airport system planning process. These surveys are generally similar to those performed by airport authorities, although they typically focus on the airport ground transportation patterns of air passengers rather than on their perceptions of customer service at the airport or their spending patterns. They often have an advantage over surveys performed by individual airport authorities that they use the same (or essentially similar) questions at each airport and are usually performed at each airport at about the same time.

A number of private firms and industry organizations conduct surveys at airports to generate information on passenger satisfaction and other factors, which they then market to the airport authorities and airlines. It is not expected that these organizations would be willing to place this information in a national database.

Household Travel Surveys

In addition to surveys performed at specific airports, various agencies undertake household travel surveys that ask about the respondents recent travel behavior. In some cases, these trips by household members include air trips. The data for these air trips can be extracted

from the survey data for all types of trip and presented in a comparable format to air passenger surveys performed at airports. These survey data provide two very valuable potential additions to the information available from airport surveys. The first is that by being derived from surveys of household travel behavior, they allow an analysis of the way in which air travel patterns are related to household characteristics and other types of travel behavior. While some of this information could in principle be obtained from airport surveys, by their very nature airport surveys will not capture travelers who elect not to use air for a specific trip or information on households that do not make air trips. The second valuable addition is the ability to study air travel behavior without the inherent bias that results from performing surveys at specific airports. This is particularly useful in studying airport choice behavior, in integrating information from multiple air passenger surveys, and in performing leakage studies.

Such household travel surveys are performed at both the national and regional levels. The two most extensive national household surveys are the National Household Travel Survey (NHTS), that was last performed in 2001 by the U.S. Department of Transportation,¹ and the Bureau of Transportation Statistics (BTS) Omnibus Survey that has been performed on an on-going basis from August 2000 to October 2003 for the U.S. Department of Transportation.² The NHTS has a much larger sample size than the BTS Omnibus Survey but only provides a profile of travel patterns for the year of the survey. The only NHTS performed to date (in 2001) updated information from two earlier surveys that were last performed in 1995, the National Personal Transportation Survey and the American Travel Survey. In contrast, while the BTS Omnibus Survey provides a smaller sample size each month (about 1,000 households), its monthly frequency allows trends to be more quickly identified and permits analysis of seasonal travel patterns.

Regional household travel surveys are typically performed by regional transportation planning agencies as part of their travel demand modeling process. The extent to which they capture information on air trips is likely to vary from agency to agency, and is a subject that is deserving of further research.

¹ Descriptive information is available at http://www.bts.gov/programs/national_household_travel_survey/

² Descriptive information is available at http://www.bts.gov/programs/omnibus_surveys/

Household travel surveys are typically performed by telephone or mail (often both) and may include requesting the responding household to maintain a travel diary for the period of time covered by the survey. While information on local travel is usually restricted to specific days, information on intercity travel is often collected for the most recent trip, due to the lower frequency of such trips. In the case of air travel, this trip may be some time before the date of the survey.

Because of the wide differences in the questions and approach used in regional household travel surveys, it is not envisaged that the data from such surveys would be included in the national air passenger survey database, although information about the surveys could be. The data from the national household travel surveys is already available from the BTS, so there is no need to include this in the national air passenger survey database. However, it may be useful to extract the subset of air travel information from the larger dataset of national travel patterns by all modes and include this in the national air passenger survey database to facilitate comparison with information from air passenger surveys conducted at airports.

5. Integration of Air Passenger Survey Data from Multiple Sources

The full value of having access to data from multiple air passenger surveys lies in the ability to integrate the information from more than one survey. This is particularly important in performing analysis of air passenger travel patterns in multiple airport regions, or performing studies of air passenger use of more distant airports that may offer better air service (sometimes referred to as “leakage studies”). It may also be useful in corridor studies, where air passenger ground travel patterns are needed at both ends of the market, or in expanding the sample size of information on air party characteristics at a given airport by including data for passengers traveling to (or from) that airport who were surveyed at other airports on their trip.

However, the integration of information from multiple air passenger surveys involves a number of difficult issues that need to be carefully addressed if the resulting data is not to become misleading or invalid. Some issues are fairly obvious but others are more subtle. The most obvious issue is the consistency (or lack thereof) in the way that the questions are worded or responses are coded. Inconsistency in coding can often be addressed through recoding responses, although difficulties arise when the response categories in two surveys overlap. For example, one survey may have given respondents the option of stating that their trip purpose combined business and leisure, while another survey may not have provided this option and asked respondents to state the *main* purpose of their trip. Inconsistent response categories are particularly common in the way airport ground access or egress modes are defined. Less obvious, but potentially important, are more subtle differences in question wording that may have resulted in survey respondents giving different answers in the two surveys for essentially the same situation. For example, one survey may have asked how many bags the respondent checked, while another survey may have asked how many bags were checked by the respondent’s travel party.

Another issue that requires careful attention is that of sample size. The expansion of survey responses to total air passenger characteristics for an entire year (or any other period) is an issue that arises in all surveys. However, the errors that are introduced by simply factoring up the survey responses to the total passenger count become much more important when the results are to be combined across multiple airports, particularly if the surveys were performed at

different times. Many organizations performing air passenger surveys generate response weights that adjust each survey response so that analysis of the weighted data provides the air party characteristics on an annual basis (or other specified period). If this is done correctly, then this facilitates the combination of data from multiple surveys. However, in many cases these weights are normalized to give the total number of survey responses rather than the total number of annual passengers (or total passengers in some other period of interest). Therefore the weighted data need to be expanded to reflect the total traffic in the period.

Among the issues that need to be considered in performing this expansion are how missing data are handled in the surveys and how connecting or through passengers were handled compared to originating or terminating passengers. Since many surveys are primarily interested in the characteristics of originating or terminating passengers (who utilize the airport ground access system) it is not uncommon to exclude connecting passengers from the surveys, or at least limit the data that are recorded for them. However, most airports only collect data from the airlines using the airport for enplaned and deplaned passengers and do not have good data on the number of connecting passengers. This makes it difficult to calculate appropriate weighting factors for local (originating or terminating) passengers and connecting or through passengers.³

The U.S. Department of Transportation (DOT) collects data from the airlines on a ten-percent sample of passenger itineraries every quarter that can be used to estimate the proportion of enplaned passengers connecting between flights at each airport on a quarterly basis. However, most airports do not perform this calculation on a routine basis. Calculating these statistics for each airport and including them on the air passenger survey database website would provide a significant value-added function that would facilitate interpretation of the air passenger survey data available on the site. These statistics can even be destination- or airline-specific. The U.S. DOT also collects monthly data on passengers carried by flight segment. These are total counts, as distinct from a sample of trips, but do not distinguish between connecting passengers and those beginning or ending their trip at either end of the segment (termed local traffic). These data can be used to develop survey expansion factors on a flight segment basis.

³ A connecting passenger is one transferring between two different flights at the airport. A through passenger arrives and departs on the same flight, but may leave the aircraft at the stopover airport to use the terminal building facilities and then reboard. Through passengers are not counted in airline enplaned and deplaned passenger statistics, while connecting passengers are.

Previous Efforts to Compare Data from Multiple Air Passenger Surveys

In spite of the importance of being able to combine information from multiple surveys, there appears to be very limited discussion in the literature of the issues involved and how this can best be done. There are a large number of studies of air party choice behavior involving multiple airports that required the use of data from multiple surveys, but the authors generally do not discuss how they addressed these issues. In fact, in some cases, one suspects that they did not address these issues and the data on which they estimated their models may well be invalid. Many of these studies are reviewed in two recent reports, a literature review that was performed for the Southern California Association of Governments (Gosling, *et al.*, 2003) and a study performed for the FAA (GRA, 2003).

A comparative analysis of air passenger survey data from multiple surveys was undertaken as part of a study for the Transit Cooperative Research Program (TCRP) on *Improving Public Transportation Access to Large Airports* (Leigh Fisher Associates, *et al.*, 2000). This study documented air party travel characteristics, particularly their use of ground transportation, at some 35 U.S. airports, but did not attempt to combine data from multiple airports.

One recent report that specifically addresses the issues involved in combining information from multiple surveys has been prepared as part of a larger model development study for the Southern California Association of Governments (SH&E, *et al.*, 2003). Similar issues have been addressed in an on-going regional airport planning study for New England (www.nerasp.org). In the case of the New England study, the difficulties involved in combining information from multiple existing air passenger surveys at airports in the region proved so intractable, that it was decided to undertake a new survey using a consistent methodology and questionnaire.⁴

However, in the case of the Southern California study, it proved feasible to combine the survey information from multiple airports to provide data on air party trips originating in an area that accounted for about 46 percent of the air passenger trips in the Southern California Association of Governments (SCAG) region, in order to support the development of a model of air passenger trip generation at a regional level. The development of an integrated dataset for this area was feasible because survey data was available using consistent survey questionnaires

and performed at the same time for the two airports (Los Angeles International Airport and Ontario International Airport) that accounted for over 80 percent of the trips in the selected area (Cambridge Systematics, Inc., 2003). Outside this area, large proportions of the air trips used airports for which less comprehensive survey data was available, creating the same problem that was encountered in the New England study.

⁴ E-mail correspondence with Ralph Nicosia-Rusin, FAA New England Region, 12/22/04.

6. Plans for the Remainder of the Study

The remainder of the study consists of three activities:

- Perform a proof of concept study that will assemble air passenger survey information from a sample of past surveys, covering about ten major airports, incorporate these into a prototype web-based database, and perform an analysis of the resulting survey data to demonstrate the utility of having access to the results of multiple surveys.
- Undertake a survey of potential users of the proposed national database of air passenger survey information to assess the support for and perceived usefulness of the proposed database.
- Perform the necessary research to define the necessary technical and organizational procedures that would be needed to implement an archive of air passenger survey data on a national level, and develop detailed estimates of the resources and costs involved in doing so.

Proof of Concept Study

The planned proof of concept study serves two purposes. The first is to demonstrate the utility of providing web-based access to air passenger survey data by allowing potential users to access data for a sample of recent surveys and by conducting a number of illustrative analyses of air passenger information obtained from multiple surveys. The second is to acquire practical experience in resolving the issues involved in implementing such a database and develop the understanding necessary to prepare a detailed implementation plan for a national database.

By establishing a web-based prototype database, potential users and FAA staff can obtain hands-on experience accessing selected data and can provide feedback on desired functionality and ease of use. It is envisaged that the air passenger survey data available on the prototype website will be sufficiently comprehensive that potential users will be able to perform their own analysis of the data.

Survey of Potential Database Users

In order to assess the support for and perceived usefulness of the proposed national database of air passenger survey information a survey of potential users of the proposed database will be undertaken. The survey will include a representative sample of airport authorities, state and regional planning agencies involved in airport or aviation system planning, aviation or airport planning consultants, relevant research organizations, and other potential users of the information. In addition to identifying the type of information most likely to be of use to different types of organization, and their anticipated willingness to contribute air passenger survey data to the proposed archive, the survey will also develop better statistical information on the frequency with which air passenger surveys are conducted by various types of organization.

Define Implementation Requirements for a National Database

This task will prepare an implementation plan and refine the estimates of the resources required to implement a national database of air passenger survey data from the initial estimates presented in Chapter 3. A detailed implementation plan will be prepared and discussed with FAA staff in the National Planning Division and other appropriate FAA offices. This implementation plan will be used as the basis from which to develop revised estimates of the resources required to implement the proposed national database.

In addition to the implementation plan, NEXTOR will define requirements for utility routines to access and combine information from multiple surveys and to facilitate uploading survey data to the archive in standard format, and will prepare conceptual functional designs and detailed functional specifications for these utility routines. These functional specifications can then be used to simplify the transition of the prototype database to a production platform and reduce the cost involved in developing the necessary support software.

References

- Cambridge Systemtics, Inc., Geoffrey D. Gosling, and SH&E, Inc., SCAG Regional Airport Demand Model – Trip Generation Working Paper (Draft), Prepared for the Southern California Association of Governments, Los Angeles, California, September 2003.
- Gosling, Geoffrey D., Cambridge Systemtics, Inc., and SH&E, Inc., SCAG Regional Airport Demand Model – Literature Review, Prepared for the Southern California Association of Governments, Los Angeles, California, June 2003.
- GRA, Incorporated, Alternate Airports Study, Prepared for the Office of the Assistant Secretary for Transportation Policy and the Office of the Assistant Secretary for Aviation and International Affairs, U.S. Department of Transportation, Final Report, Jenkintown, Pennsylvania, April 15, 2003.
- Leigh Fisher Associates, Matthew A. Coogan, and MarketSense, Improving Public Transportation Access to Large Airports, TCRP Report 62, Transit Cooperative Research Program, Transportation Research Board, Washington, DC, 2000.
- SH&E, Inc., Cambridge Systemtics, Inc., and Geoffrey D. Gosling, SCAG Regional Airport Demand Model – Survey Analysis Working Paper, Prepared for the Southern California Association of Governments, Los Angeles, California, June 2003.