Global Distribution Systems and the U.S. Commercial Air Industry

Gathering Real-Time Airline Flight and Fare Information for Spatial and Economic Analysis

*Tony H. Grubesic, Department of Geography, 400 B Braunstein Hall, University of Cincinnati, Cincinnati, OH 45220-0131. (email: tony.grubesic@uc.edu)

Mark W. Horner, Department of Geography, Florida State University, 323 Bellamy Building, Tallahassee, FL 32306-2190. (email: mhorner@fsu.edu)

Matthew A. Zook, Department of Geography, University of Kentucky, 1457 Patterson Office Tower, Lexington, KY 40506-0027. (email: zook@uky.edu)

Thomas R. Leinbach, Department of Geography, University of Kentucky, 1457 Patterson Office Tower, Lexington, KY 40506-0027. (email: leinbach@uky.edu)

*Corresponding Author

WORD COUNT

Body text (including abstract and references):	5, 429
Figures and Tables: @ 250 words each	1,250
Total:	6,679

The rapid growth of publicly available global distribution systems (GDS) and central reservation systems (CRS) for air passenger bookings has generated a dramatic shift in the U.S. commercial airline industry, with more than 35 million Americans making travel arrangements on the Internet in 2003. GDS/CRS data for domestic airfares are widely available through consumer portals such as Expedia Travel, Travelocity and Orbitz. Basic travel queries to these portals reveal the real-time airline and fare information for a variety of competing carriers. Information regarding flight routes, departure times, duration and connections are also presented by the GDS. The purpose of this paper is to present a method for acquiring and utilizing GDS data for transportation analysis. In addition, we also provide a basic comparison between these data and the popular origin-destination survey conducted on a quarterly basis by the United States Department of Transportation. Finally, to provide an example of how this GDS data can be used a gravity-based accessibility model, which utilizes network information and itinerary pricing mined from the GDS engines is presented.

INTRODUCTION

The evolution of telecommunication technologies over the past decade has left virtually no sector of the U.S. economy untouched, including the commercial air industry. One of the most significant transformations is the rapid improvement of consumer-oriented applications such as the World Wide Web for accessing itinerary information and booking airline tickets. Today, consumers seeking to make air travel arrangements are able to consult a myriad of sources for travel information. For instance, the public interfaces to central reservation systems (CRS) and global distribution systems (GDS) such as Expedia Travel, Travelocity and Orbitz, along with direct reservation sites operated by individual airlines, allow consumers unprecedented access to information on price, travel times and schedules.

Access to this information also provides a unique opportunity to measure and monitor three of the most important aspects of the commercial air industry; travel time, airfare pricing and accessibility. Unlike the 10% ticket sample compiled by the U.S. Department of Transportation on a quarterly basis, data acquired from global distribution systems is nearly real-time, extremely detailed and provides the ability to track significant changes in routing, pricing and travel times for any origin-destination pair in the United States on a weekly or even daily basis. Moreover, GDS data represent the actual choice-based information provided to consumers during the booking process, mirroring seat inventories and connection requirements for a given itinerary at a single point in time.

Given the aforementioned changes in the U.S. commercial air industry, a real-time, data collection methodology for monitoring fluctuations in airfare pricing, city accessibility and travel times offers immense potential for exploring and documenting the spatial and economic dynamics of the US air passenger market. The purpose of this paper is to present a method for acquiring and utilizing GDS data for transportation analysis. First, we explore the evolution of global distribution systems and their importance in the travel industry. Then, a methodology for mining these data is presented. This is followed by a discussion of data integrity and limitations. Finally, we provide a simple empirical example where GDS data are used to explore the spatial and economic landscape of commercial air passenger accessibility in the United States.

THE EVOLUTION OF GLOBAL DISTRIBUTION SYSTEMS

Global distribution systems (GDS) have evolved from the first computer-based reservation systems implemented by several U.S. airlines in the late 1960s and early 1970s. A complete account of the historical evolution of GDSs is too detailed to present in this paper. There are, however, several key milestones and technological innovations worth highlighting. American Airlines (AA) was the first company to develop a real-time computerized reservation system. In a joint venture with International Business Machines (IBM), the Semi-Automatic Business Research Environment (SABRE) was launched in 1964 and helped American process nearly 26,000 reservation requests per day (1,2). By the late 1960s and early 1970s, virtually all of the major carriers, including United, TWA and Delta were operating their own central reservation systems (CRS). It was not, however, until 1976 that these systems were installed in travel agencies, allowing agents to both book and change reservations directly in the system using remote access terminals. Clearly, the ability for airlines to operationalize their booking systems in geographically dispersed markets was a major competitive advantage for those who participated.

This distributed system was paralleled by increases in computing and storage power; by 1978, SABRE was available in over 130 locations and could store 1 million airfares (2). After deregulation in 1978, the importance of computerized reservation systems became even more apparent. The rapid entrenchment of CRSs with travel agents enabled bias in the reservation process. For example, SABRE might decide to withhold inventory from other CRSs in an effort to gain a competitive advantage. Further, a CRSs owned by a particular airline could be programmed to display their airline's information ahead of their competitors in an effort to distort consumer choice (3). This was compounded by the natural tendency of travel agents to book flights with the carriers providing the CRS terminals – known as the "halo-effect" (3-5). As a result, in 1984 the Civil Aeronautics Board began to regulate these systems to insure a free and competitive market for both travelers and airlines.

By the early 1990s, CRSs had evolved into more complex systems. In part, this was motivated by strategic alliances and mergers between U.S. and European carriers, spawning true global distribution systems (6). In addition, the ability to leverage the Internet as a tool for expanding the presence of these booking systems was promising. In response to these developments, the United States Department of Transportation (USDOT) increased regulation of the GDS industry, mandating that GDS owners (7):

- cannot favor or disfavor any airline in setting the display order of flights or booking fees;
- must provide data to all carriers on a non-discriminatory basis;
- must participate in all other CRS systems (for airline CRS owners); and
- must offer travel agents contracts of three years or less.

As the GDS industry continued to evolve in the 1990s, the availability of public interfaces to these systems greatly expanded, particularly with the rollout of the Internet and World Wide Web. As a result, consumers gained unprecedented access to information concerning price, travel times and schedules. Today, airlines no longer have an ownership interest in GDSs and the regulation provided by the USDOT is expired. There are currently four major GDSs in operation, Amadeus, Galileo, Sabre and Worldspan, all of which provide real-time flight information to travel agents and consumers. The public interfaces to the GDSs can be categorized in the following way: 1) airline websites; 2) GDS-based online travel agencies such as Travelocity, Expedia Travel and Orbitz; 3) opaque sites that require some type of bid/payment before knowing the actual travel schedule such as Priceline; 4) specialty low-fare sites which are analogous to a tip-sheet for selected bargains; and 5) "screen-scraper" sites which actually reads fare information from the screens of other sites and reports them to the consumer (8).

This paper is primarily concerned with mining information from the GDS employed by sites such as Expedia Travel and Orbitz, although the proposed methodology is equally effective for mining data from individual airline websites. There are two distinct advantages in acquiring data from a source like Expedia Travel. First, the booking engine used by most of these travel sites is Worldspan GDS, which processed more than 65% of all online agency travel bookings in 2003. Currently, Worldspan supplies availability, fare and rate information for more than 455 airlines around the world, thus providing an excellent snapshot of the market. Second, the information presented by sites such as Expedia Travel and Orbitz represents real-time airline flight and fare information, which is exactly what the consumer uses to make decisions regarding air travel.

DATA ACQUISITION

The data acquired for this paper was derived via a software program designed to automate the search process of individuals looking for airline flights on a public interface to a GDS such as Expedia Travel. This approach is equally applicable to any website (including individual airlines), supplying itinerary information. The data acquisition program was written in the programming language Perl and relies upon a set of object-oriented modules known as LWP (Library for World Wide Web in Perl). LWP provides a set of tools through which web queries can be automated, allowing a researcher to "scrape" data from web pages and web based applications. Both Perl and the LWP modules are freely available for download from the Internet.

Using LWP and Perl

Central to LWP's operation are a set of data and procedural definitions known in programming terms as "objects". The relevant objects for LWP and this project are **UserAgent** (an object which emulates a web browser) and **Response** (an object containing the http code returned to a web browser). LWP allows one to construct a UserAgent based query, submit the query to a web server using LWP's “get” function, and collect the result into a Response object. For example, the following code collects the homepage of the New York Times.

```
use LWP;
$web_browser = LWP::UserAgent->new();
$web_response = $web_browser->get("http://www.nytimes.com/");
```

Upon receiving a Response object, the program then searches the returned html code to find the desired data. Because Perl was designed as a text processing computer language, it is particularly suited for extracting data via regular expression analysis, *i.e.*, a technique embedded with Perl that matches particular patterns within a search string. A review of html code from the New York Times webpage reveals that headlines are preceded by "SIZE="+1">" and followed by "". This coding is unique to headlines. The following example outlines one headline and its associated html code.

```
SIZE="+1"><strong>As All Washington Guessed, Bush Zeroed In on His
Choice</strong>
```

The text of the headlines for the webpage can be extracted by the following regular expression and placed in an array named @headlines.

```
@headlines = $web_response->content =~ m{\+1\"><strong>([^<]*)</strong>}g;
```

The data in this array is then cleaned and the results printed with the following code.

```
$i=0;
while (@headlines[$i]) {
   $order = $i+1;
   @headlines[$i] =~ s/\r//g;  # removes line breaks in headlines
   @headlines[$i] =~ s/\n//g;  # removes line breaks in headlines
   print "\n$order. @headlines[$i]";
   $i++;
}
```

Combined, these lines of code produced the following output when run on July 20, 2005:

```
1. Bush Urges 'Timely' Hearing; Democrats Promise Scrutiny
2. As All Washington Guessed, Bush Zeroed In on His Choice
3. Sunnis Suspend Participation in Iraqi Constitution Panel
4. Greenspan Signals Further Rate Increases in Testimony on Hill
```

This small example illustrates the power of LWP and Perl but only scratches the surface of what is possible. For more detailed information on using LWP and further reading, see (9).

Structure of the Expedia Travel Search

Building upon the basic steps illustrated above, the data for this paper was acquired via a custom Perl program using the LWP modules. Because the program code is rather long, Figure 1 offers a logical overview of the approach used for mining the Expedia Travel data.

The first step was to input the desired search dates for a trip. Although the data searches for this project were conducted for identical departure dates of (3/23/05) and return dates of (3/30/05), this window is easily changed. For example, one could conduct the search for a business itinerary, with departure and return on the same day. In terms of geographic scope, all U.S. airports listed by the DOT with commercial departures or arrivals were selected for data retrieval (n = 206) (10). This group of airports was used to generate a matrix of origin and destination locations representing all possible itineraries between the selected airports. Because data is only collected for the outbound leg of a trip, both configurations of each airport pair (LEX-CVG and CVG-LEX) were included. This resulted in a total of 42,230 unique searches.

A query to Expedia Travel was then conducted for each city-pair. The first step in this query is the construction of a UserAgent object ($browser). While this is sufficient for many webpages, the Expedia Travel website requires that browsers (e.g. Netscape) record identifying information in what is known as a "cookie". Any queries to Expedia Travel without this stored cookie information will yield an invalid response. Therefore, we used the LWP modules to create a cookie variable for the UserAgent object and placed within it the data required by Expedia.

After defining the UserAgent object (including the relevant search dates and airport pairs) and its cookie variables, a query was submitted to Expedia Travel and a response received. If the response was invalid (i.e. data packets lost during the transfer across the Internet), it was resubmitted until a valid response code was received. While the vast majority of Internet requests proceed without problem, this safety measure guarded against invalid responses due to Internet router congestion or other problems on the network. When a valid response was received, data were extracted for all itineraries between the selected origin and destination. In order to minimize the load on the Expedia web server, the program paused for approximately 20 seconds after each search.

Searches were ordered by multiplying the number of arrivals at the departure airport by the number of arrivals at the destination airport. These totals were then sorted in descending order and the Perl script queried the top-ranked city pairs first. As a result, the busiest airports were searched first and the least busy airports were searched last. As noted previously, queries were constructed around a departure on 3/23/05 (Wednesday) and a return on 3/30/05 (Wednesday) with no time specified. Obviously, this includes a Saturday stay, which frequently facilitates a lower overall fare. The search commenced on February 15, 2005 and ended on February 28, 2005. There were approximately 150 searches conducted per hour during this time.

A typical search result is displayed in Figure 2. Flights were ordered by Expedia Travel based on price, with the Expedia engine returning anywhere from 1 to 54 flight possibilities. The summary statistics for flights returned by the Expedia engine were as follows: mean = 18.16; median = 18; standard deviation = 7.56. Price, departure time, flight duration, number of connections and routing data were collected for all flights. However, due to the structure of Expedia's search system, only outbound data were collected for each query

DATA INTEGRITY AND LIMITATIONS

Data acquired through this methodology are extremely detailed. However, in order to provide a more complete picture regarding the utility of these data for spatial and economic analyses, it is important to compare and contrast the mined GDS data to alternative sources of flight information provided by other outlets. For example, the Bureau of Transportation Statistics (BTS) collects data on a host of transport modes operating in the U.S. and throughout the world. This information includes several aviation-related databases. For instance, the DB1B database, analyzed in a variety of studies (e.g. 11, 12, 13), shares certain commonalities with the GDS data. However, there are some major differences between the two databases and we will illustrate the most important of these to demonstrate the utility of the GDS approach.

Overview of DB1B

Formally referred to as the Airline Origin and Destination Survey, DB1B is a 10% sample of airline tickets as reported by commercial carriers and compiled by the Office of Airline Information at BTS. Data items included are the origin, destination and other itinerary characteristics of passengers transported by commercial air carriers. As the BTS suggests, the DB1B database has found use in modeling air traffic patterns, delimiting air carrier market shares and estimating passenger flows. Because the individual data are tagged with airport codes describing the individual's origin and/or destination, these data are easily imported and joined with spatial data describing airport locations in a geographic information system (GIS). The most recent DB1B data were downloaded from BTS's web site (www.bts.org) and imported into TransCAD GIS to illustrate their characteristics. Figure 3 displays the locations of commercial U.S. airports and the average market fare paid by departing travelers during the fourth quarter of 2004.

The DB1B is organized into three parts. Coupon records report information on sampled passengers for each flight segment of their travel itinerary. The locations of intermediate stops made by these sampled passengers are also reported. The second set of records is summarized by market, which consists of information on the trip origin and the ultimate destination. Intermediate stops are not reported. Lastly, ticket records include basic data on the trip, such as pricing and trip length, by origin.

DB1B or GDS Data?

There are several important distinctions that need to be made between DB1B and GDS data. First, it should be clear that the two sources for airfare and ticketing information are dramatically different. DB1B data contain information on *actual* market transactions and do not make any distinction as to *when* the transactions took place, other than the quarter in which the data were compiled. In contrast, GDS data contain information on fare and service offers at a single point in time, and unlike the DB1B data, only represent the choices available to the consumer – not actual tickets booked.

That said, a major strength of the GDS data is the ability to monitor consumer ticket options for virtually any time period. If itinerary information is needed for a specific day, week or month, then those data may be obtained. Recall that the temporal resolution with DB1B is limited to the quarter, and fare information cannot be tagged to a specific time within that three-month period. Therefore, the ability to monitor and track near-real time fare information is helpful in cases where researchers are seeking to analyze industry responses to surges in fuel prices, labor strikes, bankruptcies or unexpected events, such as terrorist attacks, natural disasters or other anomalies.

A second caveat with DB1B data is the sampling methodology. While 10% is generally considered a rich sampling rate, there could be instances in the database where some passengers are not sufficiently sampled (e.g. low-density routes). Clearly, this would render the estimates of certain segment and market fares questionable. Similar limitations exist with GDS data. Although the data acquisition approach outlined in this paper has the ability to derive a more complete picture of flights and fares available to consumers at a given point in time, the saliency of large-scale collection efforts is impacted by subsequent pricing adjustments made by the airlines. With constantly changing inventories, fare schedules and adjustments to the revenue management systems made by commercial carriers, GDS data are the very definition of a snapshot in time.

One perceived disadvantage in using the GDS data from Expedia Travel or Travelocity is that itineraries available from carriers such as Southwest and Jet Blue, which do not participate in the GDS system, will not be included in the analysis. While it is true that such fare and route information is not directly accounted for by the GDS, it can be argued that the impact of low-cost carriers like Southwest is already priced into the fare structures of competing airlines (13). However, while legacy carriers in certain market do make attempts to match the pricing of a low-cost carrier, it is unclear how many seats are available at the discounted price.

A second perceived disadvantage is that finding low airfares is a matter of opportunity and not a reflection of the market. For example, travel sites like Expedia, Travelocity or the web site of a particular carrier occasionally offer special fares. However, these temporary fluctuations in market fares are exactly why mining data from the appropriate web sites can provide a unique window to the spatial and economic landscapes of commercial air travel. These special fares are masked by the DB1B reporting system, however, given a suitable search window, these fares can be accounted for by the methodology proposed by this paper. Finally, it is important to note that the data collection methodology outlined in this paper is *flexible*. While Expedia Travel was selected as a case study in this instance, minor changes in the Perl code will allow one to collect information directly from airline websites, *e.g.,* JetBlue or Southwest. In sum, the use of both the BTS DB1B data and information mined from GDSs or individual airline websites can provide an extremely rich snapshot of the commercial air industry for a given time period.

AN EMPIRICAL EXAMPLE USING GDS DATA

To demonstrate the utility of the GDS data for empirical analysis, the following example estimates a basic airport-based accessibility index. This index can be used to compare the overall accessibility of airports to the U.S. population while controlling for the price (i.e. airfare), associated with reaching those locations.

Model Specification

There are two basic premises in formulating a good accessibility index. First, the accessibility for a given location should increase as the size of the opportunities (e.g. population), available to that location increase. Second, accessibility should decrease as the costs to access the identified opportunities increases (14,15). These concepts are well-established by previous studies outlining issues of accessibility (16,17). A basic formulation for an accessibility index is as follows:

$$A_i = \sum_{j=1, i \neq j}^{m} O_j f(C_{ij}) \tag{1}$$

where

A_i is the accessibility at point i,

O_j is the opportunities at point j,

$f(C_{ij})$ is a function of the travel cost from i to j,

This index (1) is a general gravity-based model that helps account for the relationship between cost and accessibility (17,18). In this particular application, indices i and j represent origin airports and destination airports, respectively. Many possible functional forms can be employed to express the relationship between cost and accessibility $f(\ )$ in equation (1) (18). Because of its well-understood properties and frequent use in accessibility modeling (17), the exponential function is utilized. A parameter in the function denoted beta (β)accounts for the relationship between fare cost and accessibility. When the model in (1) is updated to reflect the modified functional form, we have:

$$A_i = \sum_{j=1, i \neq j}^{m} O_j \exp(-\beta C_{ij}) \tag{2}$$

Application

The application of (2) yields an origin-based index of airport accessibility. Specifically, this gravity model considers all opportunities reachable from a given origin, weighting these opportunities based on the importance assigned to travel costs, which is governed by the analyst's choice of β.

As noted previously, GDS data were obtained for the third week of March 2005. Although the richness of these data would allow for numerous avenues for analysis, several limitations in the scope of the accessibility model were made to expedite the generation of results and to maintain the illustrative and instructional theme of this paper. For instance, this particular empirical example is limited to non-stop flights between U.S. commercial airports (n = 206) including Hawaii, Alaska and Puerto Rico. Further, average fare prices were used as the measure of costs (C_{ij}) separating airports in the United States. However, it is important to note that average fares between origins and destinations are not the only choice for analysis. Instead, one could examine morning flights/fares versus evening flights/fares; lowest versus highest, etc. In this particular case, (C_{ij}) was calculated as the average of all listed flights for an origin-destination (O-D) pair. As a result, the accessibility model is based on a simple view of connectivity – if a single flight was made between a city pair on the queried day, they are considered connected. In this simplified scenario, there are no bonuses for an increased daily service frequency between city pairs. If that were the case, cities providing shuttle services in the Northeast would certainly benefit. Future work can consider this dimension of connectivity.

The measure of opportunity used in this analysis, O_j is broadly defined as the population in the j^{th} airport's service area. Again, because the model is for illustrative purposes, a detailed assessment of airport service areas (e.g. their spatial extents and airport competition) was not attempted. A more sophisticated opportunity measure would enhance the model's utility in future work (24). Here we utilize a fixed distance of 100 miles around each airport to proxy the market extents. This was calculated in a GIS using standard overlay techniques to estimate the population within each of the buffers. The original population surface was based on a GIS layer of census tracts for the US. These are fairly disaggregate spatial units and should ensure that the airport market area population estimates are reasonable.

Spatial data were imported into TransCAD GIS and the accessibility model was estimated based on matrix multiplication routines in the software. One of the major advantages to Model (2) is the flexibility afforded by the beta parameter (β) when considering costs. The β chosen for this analysis was equal to 0.0026866, which ensures that exactly 50% of city j's population will count into the city i's accessibility *if* the fare between the two cities is equal to the average of all nonstop fares (the average was estimated at $258). Cities that are more expensive to reach would see their populations count less into the accessibility index, and vise versa.

Application of the model in equation (2) yields a basic accessibility index. The produced numerical results are mixed units of population and weighted cost which have no direct interpretation. Therefore, it is customary to compare indices among origins (17). Because the calculations often result in large individual accessibility values, it is also customary to scale the accessibility indices by a constant to make them more manageable. This does not change their distribution or ordering. Each calculated airport accessibility score is divided by the maximum score in the database (468,977,216) such that the most accessible airport by this measure has a score of 1.0. Results are displayed in Figure 4.

The patterns displayed in Figure 4 are not particularly surprising, and confirm what is known about general airline conditions in the U.S. That is, most of the airports in large or well-known hub cities are highly accessible as defined by cost. As a result, there is a slight geographic bias to the most centrally located hub airports in the U.S. For example, Cincinnati (CVG), Cleveland (CLE), Chicago (ORD) and Detroit (DET) are very accessible (Table 1). The emergence of secondary airports in the United States is also evident in the results. For example, both Memphis (MEM) and Fort Lauderdale (FLL) are included in the top 25 (Table 1). In part, this is due to the fact that many of the major U.S. airports are nearing their maximum capacity, becoming congested and less passenger friendly (19). These pressures are partially responsible for the increasing popularity of secondary airports, located on the periphery of many urban areas. In fact, the available capacity at secondary airports, in terms of runways, is nearly double the existing capacity at core airports for many urban areas (e.g. Los Angeles, San Francisco, Miami, Atlanta and others) (19). Moreover, the entry of low-cost carriers to secondary airports is also spurring an increased level of enplanements.

Future Research on Accessibility Issues With GDS Data

As we have suggested, accessibility issues are key considerations for both travelers and air service providers. The approaches illustrated in this paper for both mining and utilizing GDS data allows researchers a unique window into the spatial, temporal and economic landscapes of air travel in the United States. Although the analysis presented in this paper focused on one time period and a very broad geographic scale (i.e. the national level), some interesting patterns were realized, but again, the true power of GDS data are exploited in applications where panel data are extracted for multiple locations. This type of fine spatial and temporal resolution is not possible with other databases, notably the DB1B. In an effort to establish a useable data series, data similar to that presented here for Q1 2005 has been collected for five quarters, ranging from September 2004 to September 2005. We plan to expand our analysis of these data and hope that funding will allow us to standardized it so that it can be available to other researchers as well.

It is also important to note that he possible permutations on the basic accessibility analyses presented here are virtually limitless, but again, it must be recognized that they are only possible because of the availability of GDS data. Future empirical work of interest to both public and private research agencies might include conducting state or regional level accessibility analysis, tracking accessibility at a given airport(s) over time in a performance assessment, or assessing inter-airport competitiveness – all of which are contingent on having representative airfare and service data. Because GDS data are a true snapshot airfares for a given point in time, and can be collected for any desired time period, we feel their use in research should be strongly encouraged.

CONCLUSION

In addition to exploring the evolution of global distribution systems and their relative importance in the commercial air industry, this paper has introduced a flexible methodology designed to acquire GDS data for analysis, with a specific focus on the United States. In particular, this nearly real-time data collection methodology has the potential to allow analysts to monitor fluctuations in airfare pricing, city accessibility and travel times for virtually any origin/destination pair in the U.S. We also demonstrate the potential utility of these data for spatial and economic analysis of the U.S. air passenger market with a small, instructional example focusing on airport accessibility.

This paper also compares and contrasts GDS data with alternative data sources currently available from the USDOT, namely, the popular origin-destination survey, DB1B. As noted previously, the simultaneous use of both databases has the potential to greatly expand our understanding of the commercial airline industry in the United States. Clearly, the ability to leverage GDS data, or similar information from airline websites, is a critical measure of the actual market dynamics for travel in the U.S. and beyond and provides the opportunity to measure the impacts of unplanned events, labor disputes or natural disasters on pricing and scheduling. Eventually, these data can be compared with the actual information regarding booked itineraries provided by the 10% USDOT ticket sample in an effort to measure both short and long-term economic cycles in the commercial air passenger industry. In conclusion, we hope this brief introduction to GDS data will motivate further work in this area and that it will spur additional developments in the field of transportation analysis.

REFERENCES

1. Anthes, G. Sidebar: Sabre Timeline. *Computerworld*, 2004. URL: http://www.computerworld.com/managementtopics/management/project/story/0,10801,93453,00.html?from=story_package

2. Sabre Travel Network History, 2004. URL: http://www.sabretravelnetwork.com/about/history.htm

3. Permberton, J.D., Stonehouse, G.H. and C.E. Barber. Competing with CRS-generated information in the airline industry. *Journal of Strategic Information Systems*. 10, 2001, 59-76.

4. Humphreys, B. Les SIR (systems informatises de reservation). (The CRS (Computer Reservation Systems)). ITA Studies/Etudes and Documents. 18(0/1), 1990, 1-188.

5. Wardell, D. Airline Reservation Systems in the USA: CRS, Agency Dealerships and the 'Golden Handcuff." *Travel and Tourism Analyst*, January, 1987, 45-46.

6. Barber, C.E., Pemberton, J.D. and G.H. Stonehouse. Europe's Airline Developed Computer Reservation Systems: a turbulent course ahead? *European Business and Economic Development*. November, 1993, 32-49.

7. Hymel, E.M. and J.L. Gattuso. Regulation in Brief #5: Computer Reservation Systems. *The Heritage Foundation*, 2003. URL: http://www.heritage.org/Research/Regulation/regulation_brief111903.cfm

8. OAIA [Office of Aviation and International Affairs]. Efforts to monitor Orbitz. House Report 107-308. Washington, DC, July, 2002.

9. Burke, S. *Perl & LWP*. O'Reilly and Associates: Sebastopol, CA, 2002.

10. Air Travel Consumer Report by Aviation Consumer Protection Division of the Bureau of Transportation Statistics, 2004. URL: http://airconsumer.ost.dot.gov/reports/2004/0408atcr.doc

11. Goetz, A.R. Deregulation, competition, and antitrust implications in the US airline industry. *Journal of Transport Geography*. 10, 2002, 1-19

12. Goetz, A.R. and C.J. Sutton. The geography of deregulation in the US airline industry. *Annals of the Association of American Geographers*. 87, 1997, 238 – 263

13. Vowles, T.M. The Southwest Effect in multi-airport regions. *Journal of Air Transport Management*. 7, 2001, 251-258.

14. Pooler, J.A. Measuring Geographical Accessibility: a Review of Current Approaches and Problems in the Use of Population Potentials. *Geoforum* 18(3), 1987, 269–289

15. Thill, J.C. and L. Hyunwoo. Impact of Intermodalism on Regional Accessibility. Paper presented at the *Annual Meetings of the Transportation Research Board*, Washington, D.C.: January, 2004.

16. Harris C. The Market as a Factor in the Localization of Industry in the US. *Annals of the Association of the American Geographers* 44(4), 1954, 315–348.

17. Horner, M.W. and M.E. O'Kelly. Aggregate accessibility to population at the county level: U.S. 1940–2000. *Journal of Geographic Systems*. 5, 2003, 5–23.

18. Bhat, C., Handy, S., Kokelman, K., Mahmassani, H., Oinglin, C., Srour, I., and L. Weston. Assessment of accessibility measures. University of Texas at Austin- Center for Transportation Research. FHWA report number TX-01/4938-3, 2001.

19. Bonnefoy, P.A. and R.J. Hansman. Emergence and impact of secondary airports in the United States, 2004. URL: icat-server.mit.edu/Library/ Download/233_ATIO_P_Bonnefoy.pdf

LIST OF TABLES AND FIGURES

TABLE 1: Accessibility index rankings

Ranking	Airport Code	City/Airport	Scaled Accessibility Index
1	ATL	Atlanta	1.00
2	CVG	Cincinnati	0.94
3	ORD	Chicago-O'Hare	0.86
4	DTW	Detroit	0.85
5	DFW	Dallas-Fort Worth	0.75
6	CLE	Cleveland	0.73
7	IAH	Houston-Bush Intercontinental	0.73
8	MSP	Minneapolis-St.Paul	0.73
9	IAD	Washington D.C-Dulles	0.67
10	PHL	Philadelphia	0.66
11	CLT	Charlotte	0.63
12	DCA	Washington D.C.-Reagan	0.60
13	BOS	Boston	0.59
14	PIT	Pittsburgh	0.58
15	MCO	Orlando	0.56
16	EWR	Newark	0.55
17	PHX	Phoenix	0.55
18	LAS	Las Vegas	0.52
19	STL	St. Louis	0.51
20	DEN	Denver	0.50
21	TPA	Tampa	0.49
22	FLL	Fort Lauderdale	0.48
23	MEM	Memphis	0.48
24	LAX	Los Angeles	0.46
25	IND	Indianapolis	0.45

FIGURE 1: A logical overview of mining GDS data from Expedia Travel

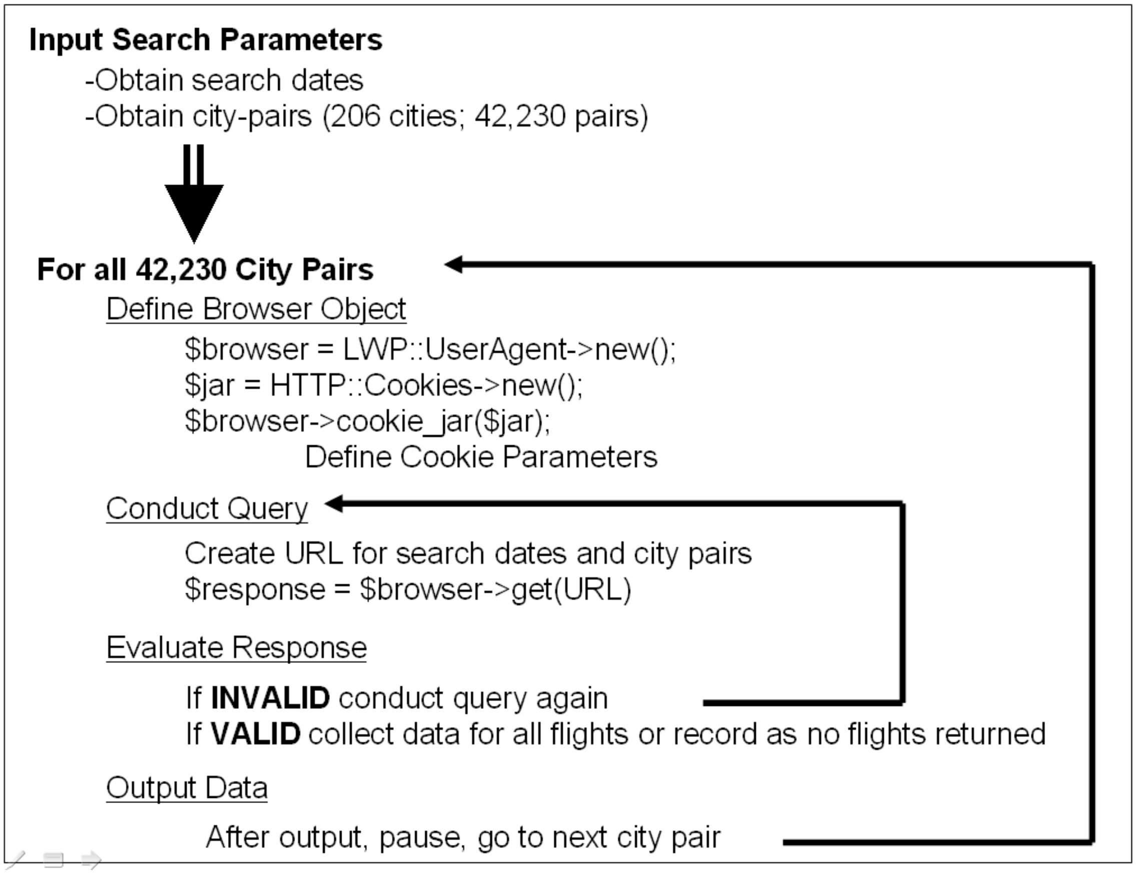

FIGURE 2: A typical Expedia Travel search result

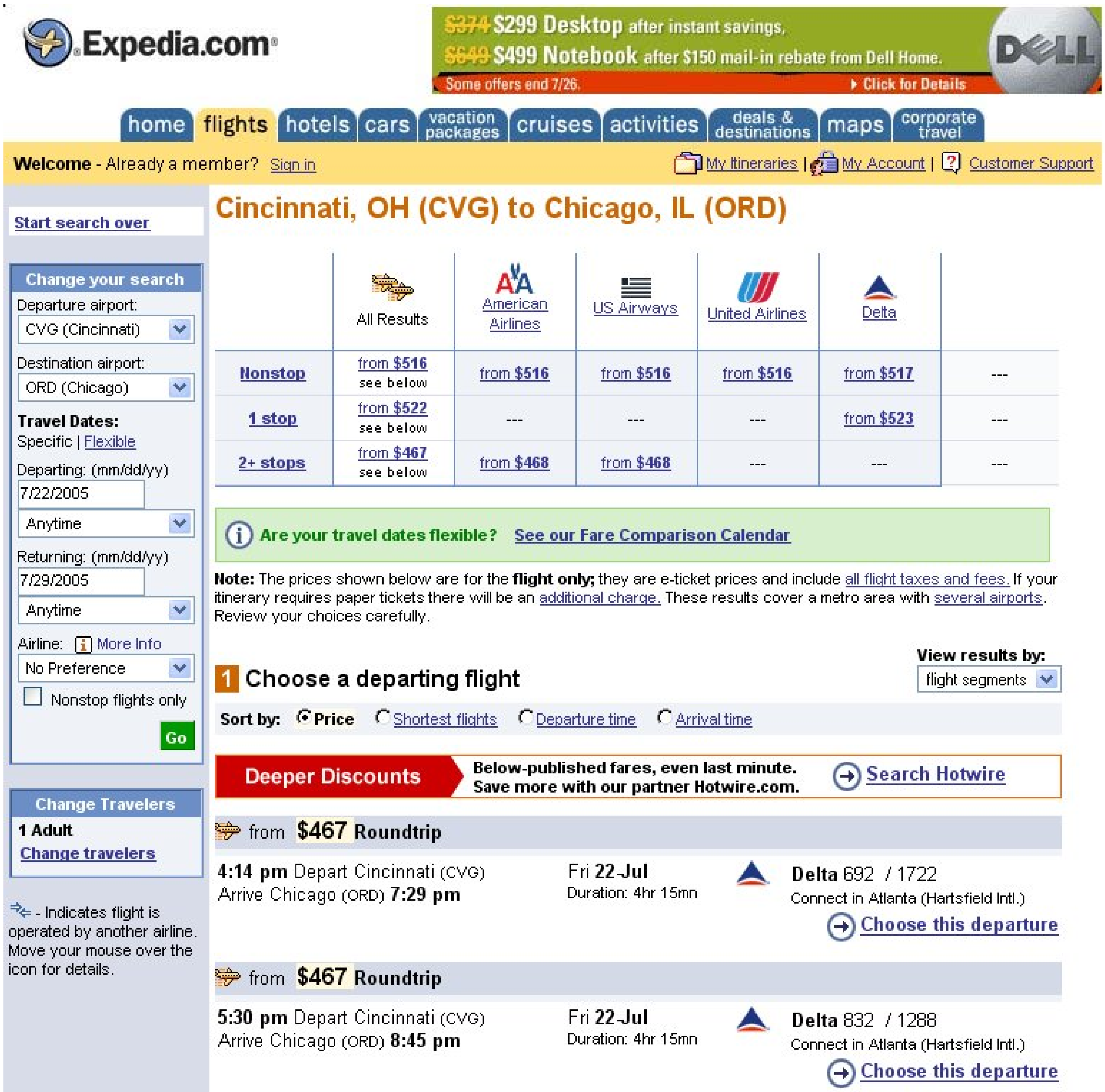

FIGURE 3: U.S. airports and average fares (Q4, 2004) (Source: BTS DB1B)

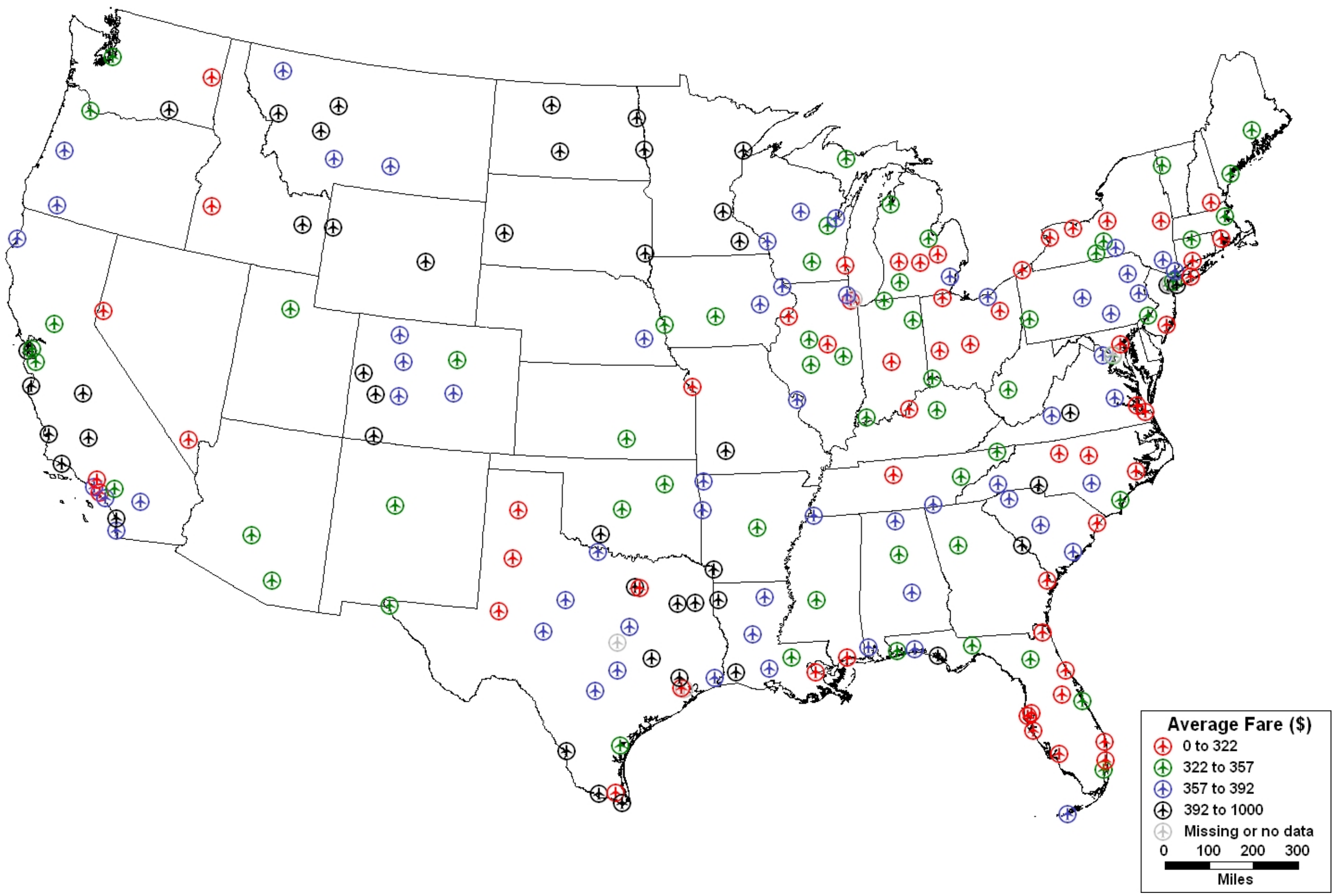

FIGURE 4: Accessibility index results