

Introducing a New and Exciting Global Airline Database: 'MIDT'

If a research project aims to present a network analysis of global air passenger transport, then having suitable data is an unquestionable requirement. However, standard data sources do not always live up to this requirement. In this short paper, we therefore introduce an alternative database, the so-called MIDT ('Marketing Information Data Transfer'), which is able to overcome some of the limitations associated with other data sources. The MIDT is an almost untapped data source which contains information on the airline connections of more than half a billion passengers for the year of 2001. A preliminary analysis of the data reveals and confirms interesting connection patterns.

By Lomme Devriendt, Ben Derudder and Frank Witlox

This short paper introduces an exceptional airline database, the so-called 'MIDT' (Marketing Information Data Transfer). This database gives detailed information about the worldwide booking activities of travel agencies and airline carriers. The MIDT database is specifically designed to provide airline companies with valuable competitive information so that they are able to make well-informed decisions in their business strategies. While this data source can also be interesting for analyses on the importance and precise function of different airports, such an analysis has not yet been carried out (notable exceptions are Derudder and Witlox, 2005a, 2005b). This paper aims to rectify this situation. First, we introduce the MIDT database, through which we point out several advantages compared to other airline databases. Second, we demonstrate some of the basic patterns that emerge from the data by deriving a detailed ranking of the major passenger airports in the global airline network.

Standard Airline Data Sets

If a research project aims to present a network analysis of global air passenger transport, then having suitable data is a straightforward requirement. At first sight, it appears that air transport economists and geographers can rely on a variety of suitable data sources, such as the ones provided by IATA (International Air Transport

Association), AEA (Association of European Airlines) and ICAO (International Civil Aviation Organization), in addition to information gathered by Boeing/Airbus, OAG (Official Airline Guide) data, GMID (Global Market Information Database), Ministries and/or Departments of Transportation, Departments of Civil Aviation, International Air Travel Statistics, etc. Despite the interesting insights that can be derived from these sources, it should also be noted that they are not very appropriately geared towards a systematic analysis of global (i.e. on world-wide) airline connections. When considering all of them, one can discern five important limitations:

- implicit state-centrism in the data;
- lack of comparability between different data sources;
- information biased through the use of selected carriers;
- lack of origin-destination data;
- use of proxy variables such as scheduled flights or services.

The first problem is that most data sources detail international rather than global flows. The resulting bias towards inter-state rather than trans-state flows tends to undervalue relations between airports in large and/or economically significant nation-states (a notable example is the United States). For example, data provided by the ICAO report on international con-

nections rather than on global connections. This results in a downgrading of U.S. airports because an important connection such as Chicago O'Hare - Newark is not incorporated in the data set. This means that if no other information sources are used to complement the ICAO data, we would have no information on the relations between two airports situated in the same state.

Second, data provided by airline associations (e.g., AEA) often focuses on a specific world region. Thus, the provided information does not cover the entire global economy. This problem implies that several data sources have to be consulted if one wishes to thoroughly analyze flows between several geographical regions. Two additional problems arise here: (i) data is not necessarily measured and gathered in the same way and (ii) is not necessarily published in the same year. The result is that an analysis based on aggregated data will very likely be devalued, because of a lack of comparability between different data sources.

A third limitation of some of the standard data sources is that, sometimes, information comes from member airlines of airline associations. This implies that airlines servicing a certain region without being members of the association will not be included in the dataset. This may cause biases in the database. For example, Air Scotland,

Corsair and Air Polonia are not members of the Association of European Airlines (AEA), which implies that their connections are not included in the Association's dataset, even if they serve the same region as members of the AEA.

A fourth major obstruction in the 'translation' of airline data into global airport analyses is the lack of origin/destination information. Most airline data records the individual legs of a trip rather than the trip as a whole, so that, in case of a stopover, the origin/destination links are replaced by two or more 'origin/destination' links. This implies that we do not have a clear and detailed view of the real 'function' of the airport (hub airport, regional airport, international airport).

Fifth, global air traffic data should preferably reflect real passenger movements rather than estimated connections derived through proxy variables such as scheduled flights or services (e.g. OAG).

Since we intend to pursue a truly global analysis of major passenger airports, the data should preferably not suffer from the above-mentioned drawbacks. In other words: our data must be global rather than international. Furthermore, we need information on origin/stopover/destination travel (i.e. the full trip), thereby paying attention to the particular function of the airport. In addition, the data will need to comprise all passengers traveling on a specific origin-destination pair. This means that as many airlines as possible will have to be included. And finally, the information we use, should pertain to real passenger movements. The data set presented in the next paragraph, contains such information.

The MIDT Data Set

Our analysis of major passenger airports around the globe is based on the so-called 'Marketing Information Data Transfer' (MIDT). This data set contains information on global airline bookings, and provides information such as airline codes, cities of embarkation and disembarkation, booking status code, passenger count, flight number, class of service, booking date, departure date, agency name,

ARC/IATA number, pseudo city code, PNR locator and a cancellation indicator. All the bookings recorded in the MIDT system originate entirely from so-called Global Distribution Systems (GDS), which are electronic platforms used by travel agencies to manage airline bookings (i.e., selling seats on flights offered by different airlines), hotel reservations, and car rentals purposes. Examples of GDS are Galileo, Apollo, Sabre, Worldspan, Amadeus, Topas, Infini and Abaccus (Shepherd Business Intelligence, 2005).

A potential disadvantage of this MIDT database is related to its pure GDS-origin (The BoydGroup, 2005): bookings made directly with an airline are often excluded from the MIDT dataset. Airlines choose this direct booking option to avoid fees charged by travel agencies. Direct bookings via the Internet are estimated to cost an airline \$1, while bookings through travel agents cost an estimated \$10 (Goetzl, 2000). Low-cost carriers such as Southwest, EasyJet, Virgin, and Ryanair specifically restrict themselves to direct sales to avoid the extra costs associated with the use of GDS. Given that our research interest lies on the global scale and that no-frill airlines usually operate more locally, this is not a very big problem here (see De Groote, 2005; Alderighi et al., 2005).

With the co-operation of one airline, we were able to obtain a partial MIDT

database, which comprises information on origin and destination, stopovers and carriers of 547.410.397 passenger movements. These movements cover the period January - August 2001. Working with data before September 2001 has the advantage that they are not by distorted by the terrorist attacks of 11 September 2001. During the months following that date, many destinations were not serviced since numerous connections were (temporarily) shut down, and some carriers disappeared from the market (see e.g. Lai and Whei-Li, 2005). The Association of European Airlines (AEA) announced an intra-European traffic decrease of 11.6% during the last 4 months of 2001 when compared to the same period in 2000. On the North-Atlantic routes, air traffic dropped severely in November 2001: 30-35% relative to the previous year. Thereafter, the losses diminished and stabilized at 20% at year-end. On the routes to the Far East, the number of passengers decreased to more than 20% in October and November 2001. However, by March 2002, traffic had almost returned to levels seen in the previous year (AEA, 2002). None of these traffic changes will therefore bias our database. Conversely, we have the disadvantage that our data set covers only one period, so that no comparisons are possible; we are not able to track changes over time and study the evolution of specific connections.

Airport	City	No. of pax
London Heathrow Airport	London	26680723
Los Angeles International Airport	Los Angeles	22275989
Charles De Gaulle Airport	Paris	17863633
Kennedy International Airport	New York	15650776
Chicago-O'Hare International Airport	Chicago	14250119
Newark International Airport	New York	13758248
Chek Lap Kok Airport	Hong Kong	13441401
Miami International Airport	Miami	13035628
International Airport	Frankfurt	12888709
San Francisco International Airport	San Francisco	12297086
William B Hartsfield Airport	Atlanta	11911059
Pearson International Airport	Toronto	11611632
Barajas Airport	Madrid	11388947
Bangkok Don Muang International Airport	Bangkok	11356105
Fiumicino Airport	Rome	11037336
Orlando International Airport	Orlando	10953141
Amsterdam Schiphol Airport	Amsterdam	10949204
Logan International Airport	Boston	10527606
Changi Airport	Singapore	10514931
Guardia Airport	New York	10214707

Table 1: Top 20 Airports Worldwide (by passenger numbers).

Some Preliminary Results

To conclude this short paper, we present some preliminary results of our MIDT-analysis. The results are shown in Table 1 and Figure 1. These results were derived from a transformation of the raw data in the MIDT. Note that these numbers are based on actual passenger arrivals and departures (hub passengers are therefore excluded). Table 1 ranks the 20 major airports in the global airline network (see also Figure 1). The size of the nodes in Figure 1 is based on the total number of travelers. For further details, we point to Derudder and Witlox (2005a, 2005b) and Derudder et al. (2005a, 2005b). These papers can be downloaded at <http://www.lboro.ac.uk/gawc>.

Conclusions

In this article, we focused on the advantages associated with an exceptional, albeit thus far almost untapped database. The information contained in the MIDT is unique in the sense that it does not suffer from shortcomings associated with some of the standard statistical sources. As a source for global airline connectivity research, this data source is as yet far from fully explored. In future research, we would like to focus on a number of topics, such as extensive hub analyses, com-

parisons with existing research on the emergence of a so-called 'world city', and in-depth regional analyses.

References

- AEA (Association of European Airlines) (2002) Yearbook 2002. (<http://www.aea.be>).
- Alderighi, M., Cento, A., Nijkamp, P. & Rietveld, P. (2005) Network competition - the coexistence of hub-and-spoke and point-to-point systems. *Journal of Air Transport Management* 11, 328-334.
- De Groote, P. (2005) The success story of European low-cost carriers in a changing air world. *GaWC Research Bulletin* 174. (<http://www.lboro.ac.uk/gawc/>).
- Derudder, B. & Witlox, F. (2005a) An appraisal of the use of airline data in assessing the world city network: a research note on data. *Urban Studies* 42 (13), 2371-2388.
- Derudder, B. & Witlox, F. (2005b) On the use of inadequate airline data in mappings of a global urban system. *Journal of Air Transport Management* 11 (4), 231-237.
- Derudder, B., Devriendt, L. & Witlox, F. (2005a) Flying Where You Don't Want to Go: An Empirical Analysis of Hubs in the Global Airline Network. *GaWC Research Bulletin* 187. (<http://www.lboro.ac.uk/gawc/>).

- Derudder, B., Witlox, F. & Taylor, P.J. (2005b) United States Cities in the World City Network: Comparing their Positions Using Global Origins and Destinations of Airline Passengers. *GaWC Research Bulletin* 173. (<http://www.lboro.ac.uk/gawc/>).
- Doganis, R. (2001) *The Airline Business in the 21st Century*. Routledge: London and New York.
- Goetzl, D. (2000) Southwest airlines takes flight fare fight to the web. *Advertising Age* 71 (40), 20-22.
- Lai, S. & Whei-Li, L. (2005) Impact analysis of September 11 on air travel demand in the USA. *Journal of Air Transport Management* 11, 455-458.
- Shepherd Business Intelligence (2005) (<http://www.shepsys.com>).
- The BoydGroup, Aviation Consulting and Forecasting (2005) (<http://www.aviationplanning.com>).

Corresponding Author

If you would like to know more about MIDT, please contact:
Prof. Dr. Frank Witlox
Ghent University, Geography Department
Krijgslaan 281 (S8)
B-9000 Ghent
BELGIUM
Tel: (+32)92644553
Fax: (+32)92644985
email frank.witlox@ugent.be



Figure 1: Top 20 Airports Worldwide (by passenger numbers).