Developing tourism statistics at the sub-national level: A response – Issues and Insights

Requested comment

Steve MacFeely

Summary: Across many countries, marked disparities in the economic performance and development of regions exist and persist. Trying to measure and understand these regional or sub-national inequalities is a complex task. For tourism statistics, this is not a trivial or a purely academic question as tourism is a local or “place” centred phenomena. Consequently it is appropriate that the sub-national perspective should be considered a priority.

Antonio Massieu in his paper Developing Tourism Statistics at the sub-national level: the measurement of flows of trips and visitors covers a wide range of topics and highlights a number of very important issues. It would be impossible to address all the issues raised and do them justice. Consequently I have selected a few issues where I hope the experiences and developments in Ireland may offer some universal insights.

The potential overlap between tourism and transport statistics is an important theme addressed. One could go even further however. Tourism is not simply an economic phenomena but an important social one too. If we examine the overlaps between tourism, transport and social statistics, the similarities between, for example, a Same Day Trips Survey, a Passenger Mobility Survey and a Time Use Survey are striking. This paper will discuss ongoing efforts in Ireland to develop an integrated household survey that among a suite of issues will attempt to address a number of overlapping tourism, transport and time use data requirements.

The use of administrative or existing data is a central issue highlighted. If an appropriate administrative data source can be found, it can potentially reduce respondent burden and compilation costs. Furthermore, administrative data may be the only realistic way of compiling sub-national data, as the cost of running sufficiently large surveys to generate robust sub-national level data may well be prohibitive. In Ireland administrative data has been used, both successfully and un-successfully to generate tourism statistics. These experiences are detailed in the paper.

Technology is advancing at an incredible speed. With this advance, multiple potential data sources are emerging. Within transport statistics GPS has long been recognised as an immensely powerful tool for tracking persons and vehicles. With the high penetration rates now achieved with mobile phones arguably a better source of data is emerging, particularly as the next generation of phones will be equipped with AGPS (or assisted GPS) giving a more accurate location than that currently available. For tourism statistics, passive mobile “roaming” data offers enormous potential for developing international tourism measures, including origin, destination and route. But can mobile phones also be harnessed as a survey tool suitable for completing expenditure diaries? This paper will outline some initial feasibility work being done in Ireland.

---

1 Director of Business Statistics – Central Statistics Office, Ireland
## Index

Introduction

Part 1

The importance of sub-national statistics

Determining appropriate Tourism regions

The importance of tourism expenditure for consumer prices

Overlaps with other statistical domains

Part 2

Administrative Data
- Airport – pairings database
- Visa data
- Credit/Debit card data

Harnessing technology
- “Digital” tourism diaries
- Passive data

Ethical issues

Conclusion

References
INTRODUCTION

The Central Paper covers a wide range of issues. In the first part of the paper I have selected and addressed only a few of the issues discussed, namely: the importance of sub-national statistics and determining the appropriate tourism regions; the importance of framework, structure and metadata; the potential links with other statistical domains, in particular transport statistics but also consumer price indices.

In the second part of the paper I will discuss ongoing attempts by the Central Statistics Office in Ireland to harness administrative, other commercial data holdings and technology in order to develop new tourism statistics. Much of this work is ongoing and success or failure is as yet uncertain.

PART 1

THE IMPORTANCE OF SUB-NATIONAL STATISTICS

The Central Paper highlights the emphasis and importance that UNWTO, the IRTS and the TSA:RMF place on the sub-national issue. Given the local nature of the tourism product this emphasis is justified. The importance of robust regional statistics is particularly important if "bottom up" data are considered a preferable to "top down" – and I would argue they are.

It is also important to recognise that the importance of sub-national statistics is not limited to tourism. Regional statistics are emerging as one of the most important issues across a number of statistical domains. In 2006 the National Statistics Board in Ireland conducted a survey of statistics users and asked them to review the output from the Central Statistics Office (NSB, 2007). Insufficient regional data emerged as one of the most frequent shortcomings raised across all statistical domains and user categories. This widening interest in regional issues will ultimately be of benefit to tourism statistics, as improved regional data across different statistical domains will allow the compilation of more robust regional Input-Output (I-O) tables and in turn regional TSA.

Coherent and sustainable regional statistics are expensive, both financially and in terms of time. At the national level in Ireland there exists a high level of cooperation and understanding between the Dept. of Arts, Sports and Tourism (DAST), the tourism agencies and the CSO. This has evolved over time thanks to circumstance and the efforts invested by all the parties concerned. At the regional level however, no comparable relationship or liaison exists between all the relevant players (i.e. between CSO, DAST, the tourism agencies, the local authorities, city and county councils and the regional assemblies). All of these bodies need to come to the table if regional data are to be produced on an ongoing basis. An agreed understanding and common goals must be agreed upon. Harmonised metadata, such as classifications, definitions etc. must be adopted as far as possible, so that local, regional and national data are consistent and comparable.

DETERMINING APPROPRIATE TOURISM REGIONS

 Appropriately then, one of the first issues addressed in the Central Paper\(^2\) is what the appropriate level of sub-national territorial disaggregation for tourism statistics should be. In considering this question we must first examine what a region or sub-regional territory is? Is

---

\(^2\) See Point 10 of Central Paper
it is simply a geographic entity, or something else perhaps, with a unique cultural identity or set of characteristics? Can or should a tourism “region” be different from other institutional or statistical territories? Whatever our choice, it must be measurable and pragmatic.

Over recent years a number of Government plans in Ireland have targeted balanced regional development as a priority\(^3\). This is in turn has generated much debate as to what exactly the most appropriate regions are and how balanced development for those regions can be achieved.

What the most appropriate regions are is a difficult question to answer. Membership of the EU has led to the creation and adoption of the NUTS regions in Ireland. While the NUTS 4 level regions correspond to the long standing county structures in Ireland, the NUTS classification doesn’t otherwise correspond to any existing sub-national frameworks or institutions. This has posed a challenge for sub-national or regional statistics in Ireland. Official statistics in Ireland are compiled and disseminated using the NUTS classifications. Consequently any sub-national I-O tables that might be constructed will necessarily correspond with the NUTS classifications. However many other state bodies and institutions do not organise their work or compile their data on the basis of NUTS classifications. Consequently, in Ireland there are a myriad of regional classifications, ranging from health, environmental or police regions to tourism regions, none of which correspond with NUTS. There are of course very good reasons for different domains and interests have differing regions, but equally there are good reasons for a small country like Ireland to have a single regional structure and classification system.

The TSA at a national level provides a powerful tool to help the analyst develop an understanding of the tourism market. However, given the regional disparities that exist in many countries, and certainly in Ireland, a national or top down approach loses many of the subtleties of the tourism phenomena and limits the modelling potential of the data. Ideally, we should strive for regional or sub-national TSAs, such as exist for some regions of Spain\(^4\) where different tourism markets can be identified and differentiated. For example, many countries have a number of quite different tourism products such as maritime, coastal or mountain tourism. Certainly in Ireland, the tourism product available in the east (predominantly urban) is quite different to that available in the west (predominantly rural).

What the UNWTO have proposed seems quite sensible and pragmatic then. The regions should be the administrative or political units (corresponding with NUTS in the case of the EU) but that sub-regional territories or “local tourism destinations” will most likely also exist. These LTDs should aggregate up to the NUTS or other administrative units. This is a vitally important principle, for while tourism regions may differ, at some point the tourism regions must be able to map to the administrative regions.

As a general rule, the LTDs should be as small as is practicable to allow for the greatest flexibility in aggregation/disaggregation e.g. density - rural/urban, geographic – maritime/coastal etc. Ideally they should also have some sort of independent meaningful identity or boundary that will make their existence useful. This is vitally important if people are to identify with LTDs. One of the main criticisms of the NUTS structure in Ireland is that it lacks any meaningful relationship with existing structures, regions and institutions and that NUTS 2 and NUTS 3 were too large to be meaningful or of use. So the LTDs will have an important role in the provision of meaningful regions but it is crucial that these LTDs can be mapped to NUTS or equivalent so that tourism data can integrate with the wider dataset. This will be especially important if sub-national TSAs are the ultimate aim.

---


\(^4\) For example, the Tourism Satellite Account of Andalusia for 2000
THE IMPORTANCE OF TOURISM EXPENDITURE FOR CONSUMER PRICES

The Central Paper raises a very important issue with regard to national and regional price indices and how tourism expenditure and consumption plays a central role in their accurate measurement. For example, the Harmonised Index of Consumer Prices (HICP) is the agreed approach to measuring consumer inflation across the EU where coverage is household final monetary consumption as defined by national accounts concepts for the European System of Accounts (ESA 1995). For many countries, including Ireland, the HICP is a sub-index of the national Consumer Price Indices. Consequently the rules and definitions applied to the HICP also apply to many national CPIs.

Article 1 of Council Regulation No 1688/98 stipulates that “household final monetary consumption” is defined as that part of final consumption expenditure which is incurred by households irrespective of nationality or residence status [...] and on the economic territory of the member state…” The Eurostat manual (Eurostat, 2004) for users notes one of the practical consequences of using “household final monetary consumption expenditure” is that the geographic and population coverage includes all purchases by households within the territory of a country i.e. those by both resident and non-resident households. This is usually referred to as the “domestic concept”.

That HICPs (and many CPIs) are compiled on a domestic rather than a resident concept is very important, for as the Central Paper notes it may result in a serious problem. Expenditure weights (and the basket of goods) for the HICP must properly reflect private monetary consumption on the respective domestic economic territory of resident and non-resident households. Consequently, in terms of domestic concept, there are three elements that must be taken into consideration when constructing the base weights for the HICP:

1. Total resident private household monetary consumption
2. Less resident private household monetary consumption beyond the domestic territory
3. Plus non-resident private household monetary consumption within the domestic territory

This also implies that the basket adequately reflects non-resident household (tourism) consumption. This has both product and geographic implications and given the importance of HICPs to the EMU project, it is not unreasonable to ask on what basis are EU member states making these adjustments. The legal instruments pertaining to the HICP obviously don’t extend beyond the EU. However for non-EU countries, their CPIs are either compiled on a domestic basis (in which case the same challenges as outlined above apply) or on a resident basis. A CPI based on the resident concept, will still require robust and detailed tourism expenditure data for outbound tourism.

In practical terms, this requires the CPI compiler to have good data available, not only on the total volume of inbound and outbound tourism expenditure, but also the nature of those transactions and where they occurred. How much did foreign tourists spend on the domestic territory? What goods or services were purchased? Where were those goods and services purchased? The later question is very important, not only making sure pre and post payments are correctly captured but also if regional price indices are being constructed.

5 See point 31 of Central Paper
7 Manual for users - Eurostat
The source of the tourism expenditure data is also critically important. In Ireland, the principle source of data for CPI and HICP weights is the Household Budget Survey (HBS). However, comparative studies between the HBS in Ireland and Household Travel Survey (HTS)8 revealed that HBS expenditure for comparable trips (i.e. domestic and international trips with 4+ nights) was consistently lower than expenditure reported by the HTS (MacFeely, 2005).

These issues all highlight the importance of having good quality tourism expenditure data. For the HICP and national CPIs, the resulting adjustments may not be trivial but rather may be very significant, particularly for regional price indices in a tourism dependent region. In Ireland where tourism accounts for in or around 7% of GNP (Deegan et al, 2004) the importance should be obvious. The importance of tourism expenditure to the accurate compilation of the HICP only reinforces the importance of tourism expenditure to both the National Accounts and the Balance of Payments.

OVERLAPS WITH OTHER STATISTICAL DOMAINS

The potential overlap between tourism and transport statistics is an important theme addressed in the Central Paper9. Arguably one could go even further. The potential links and overlaps go wider than transport. Tourism is a complex phenomenon involving physical movement, economic impact and social consequence. The WTO-IRTS 2008 draws attention to this fact in its very first paragraph “Tourism is a social, cultural and economic phenomenon related to the movement of people…” (UN-WTO, 2008).

Consequently the measurement of that phenomenon has important implications for National Accounts, Balance of Payments, Social and Business statistics and as we have seen above, for Price statistics. Tourism statistics also have an important role to play in the development of rural, regional, spatial, sustainable development, energy, transport and environmental domains. Only by recognising and embracing the cross-cutting nature of tourism can we begin to exploit tourism statistics to their fullest potential.

For example, studies in Ireland have indicated that nationality could be an important determinant in outbound travel patterns of Irish residents (MacFeely, 2006), as migrants continue to return regularly to their native countries for several years after emigrating. The addition of other socio-economic variables to the tourism dataset could also open up a number of other interesting and insightful research avenues. For example, the principle economic status, employment status or income of householders could provide new insights into travel patterns and could widen our understanding of topics not typically associated with tourism, such as relative deprivation.

The accurate measurement of migration is a perennial challenge for official statisticians, with serious implications for demographic, labour market and tourism statistics (to name a few). Understanding migration is important not only for the accurate measurement of tourism numbers but also for understanding route and seasonal travel patterns. Recent controversies in the UK over migration measurement have served as a timely reminder on just how important the links between tourism and demography are (BBC, 2008).

The overlap between tourism, travel and transport is obviously substantial. All of these domains involve the movement of people, their origin and destination and their purpose of journey etc. If we examine the data requirements for a household transport or passenger mobility survey, a same day trips survey and an overnight trips survey, the potential overlap

---

8 The Household Travel Survey is the main source of domestic and outbound tourism data in Ireland.
9 See Section D of Central Paper
is significant. To begin with there are immediate overlaps for socio-economic (age, gender, marital status etc.) and geographic (county, NUTS region etc.) type variables, which are common to most household surveys.

**Figure 4.1 – Potential overlaps between transport and tourism data requirements**

<table>
<thead>
<tr>
<th>Basic Journey Details</th>
<th>Passenger Mobility</th>
<th>Same-Day Trips</th>
<th>Overnight Trips</th>
</tr>
</thead>
<tbody>
<tr>
<td>Origin</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Destination</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Distance travelled</td>
<td>x</td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Participants (number/details)</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Date of departure</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Length of Journey (hours)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Length of Journey (days)</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Purpose of Journey</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Mode of Transport</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Type of Accommodation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Method of booking (Transport)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Method of booking (Accommodation)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regular/Routine Journey?</td>
<td></td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Expenditure - Total</td>
<td>x</td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Expenditure - Prepayments</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Expenditure - Details</td>
<td>x</td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Airport/Port of departure (if outbound)</td>
<td></td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Airport/Port of return (if outbound)</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Number of nights in Ireland en route (if outbound)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expenditure in Ireland (if outbound)</td>
<td></td>
<td></td>
<td>x</td>
</tr>
</tbody>
</table>

For the basic journey details there is also quite a significant overlap in the data required. Figure 4.1 presents a simplistic example of the types of overlap that exist. For a passenger mobility survey, a lot of additional data on trip details and journey legs would be required, that would not be required from a tourism perspective but might nevertheless be useful from a sub-national perspective or in determining whether a same day trip qualifies as a tourism trip or not. Qualy, the overlap between such surveys and a Time Use Survey is noteworthy, albeit from a different perspective.

Exploiting overlaps between statistical domains has a number of important advantages. It immediately widens the potential audience and thus the potential use of the data. Either through the linking of datasets, or by expanding the dataset, tourism can widen its appeal beyond travel and balance of payments to include labour market, migration, social inclusion,relative poverty and deprivation and environmental analysis. With this in mind, the Central Statistics Office in Ireland is currently proposing to introduce a new general household survey vehicle which would have labour market statistics at its core but with several satellites orbiting the core at different frequencies. The motivation for such a development is both financial efficiency but also to realise potential data linkages that cannot currently be easily explored.
PART 2

ADMINISTRATIVE DATA

Response burden, cost and cost effectiveness are familiar issues for any institution concerned with the compilation of statistics. Such constraints must be balanced against a steady growth in demand for data, and in particular as noted in part 1, a demand for improved regional data. In Ireland, the Central Statistics Office has been actively attempting to harness administrative and other existing data sources to generate more comprehensive official statistics, including tourism statistics.10

It comes as no surprise, and it is very appropriate that the use of administrative or existing data is a major issue highlighted in the Central Paper. If an appropriate administrative data source can be found, it can potentially reduce respondent burden and compilation costs. In other cases administrative sources may simply supplement rather than replace existing survey data. Furthermore, administrative data may be the only realistic way of compiling robust sub-national level aggregates, as the cost of running sufficiently large surveys to generate good quality sub-national level data may well be prohibitive.

In Ireland administrative data has been used, both successfully and un-successfully to generate tourism statistics. Two case studies are briefly detailed: The first is the successful development of an Airport-Pairings database, the second is an un-successful attempt to use visa data to estimate visitor numbers for countries with small traffic flows to Ireland (sometimes we can learn as much from failure as we can from success). Finally a new initiative to use credit/debit card data is also outlined briefly.

Airport – pairings database11

During 2005 discussions with Irish airport authorities indicated that, rather than compiling special aggregates for the CSO, it was easier to provide highly disaggregated administrative route data (i.e. the total number of passengers embarking and disembarking on each airport to airport route). Consequently, while the CSO now receives much more detailed data, the response burden to the airport authorities has been reduced and the transmission time lag has been reduced (by as much as a month in some cases).

CSO realised that a central dissemination resource for these data would be of significant benefit to the travel industry and beyond. Furthermore, CSO saw this as an opportunity to provide a new statistical output at a relatively low cost.

The Airport-Pairings Database12 was launched in February 2008 at a business statistics seminar13. This database provides monthly information on every direct aviation route in and out of Ireland. Although it varies by month, the database typically comprises a matrix of the 9 Irish airports and approximately 500 foreign airports, resulting in around 4,000 populated cells per month. The database allows users to navigate or search, subset and structure data using PC AXIS before viewing on screen or saving to a file type of their choice. Data can be compiled at four different levels of aggregation airport (IATA14 coded), city, country or continent. The data are lagged by 6 months in agreement with the data providers in order to safeguard market sensitivities.

10 See CSO - SPAR reports CSO (2003, 2006) and NSB Report (2005) for details
11 For more information on the background to this database see MacFeely & O’Hanlon (2007)
12 To view the Airport – Pairings Database go to http://www.cso.ie/statistics/Tourism_and_Travel.htm
13 http://www.cso.ie/pressreleases/business_statistics_seminar_programme.htm
14 IATA – International Air Transport Association
The database also has a number of valuable cross-cutting qualities, as the data are also very useful from a transport perspective. The CSO are currently developing a second generation of the database that in time should offer some visualisation facilities such as maps, route-KMs and a country of residence (Irish/Non-Irish) passenger split by route. These latter additions will facilitate the calculation of an Irish/Non-Irish passenger-KMs by route.

For example, tentative estimates for overseas air passenger-KMs are presented in Table 5.1. By combining the 2007 Airport-Pairings Database and Country of Residence Survey results, and using great circle routes to estimate the air-route distance between airports, total and Irish air passenger-KMs are derived. Direct flights in and out of Ireland generated 40.8 billion air passenger-KMs in 2007. Irish passengers accounted for 23.2 billion passenger-KMs or 57% of the total. Of course, Irish tourists travel to many destinations beyond the range of direct flight (such as Australia etc.).

Using the Household Travel Survey results to estimate transfers (by comparing ultimate destination with direct flight destinations) then Irish passenger-KMs in 2007 are likely to have approximated 33.2 billion. This latter estimate is a rather approximate number, as detailed route and destination information were not available. Nevertheless, it serves to highlight the scale or magnitude of air passenger-KMs generated for a single year.

Table 5.1.- Total Air Passenger-KMs generated by Air Travel to and From Ireland, 2007

<table>
<thead>
<tr>
<th></th>
<th>Direct Flights - Number of journeys</th>
<th>Direct Flights - Air Passenger KM</th>
<th>Direct + Indirect - Air Passenger KM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Million</td>
<td>Billion</td>
<td>Billion</td>
</tr>
<tr>
<td>Irish Residents</td>
<td>14.6</td>
<td>23.2</td>
<td>33.2</td>
</tr>
<tr>
<td>Overseas Residents</td>
<td>13.9</td>
<td>17.6</td>
<td>26.3</td>
</tr>
<tr>
<td>Total</td>
<td>28.5</td>
<td>40.8</td>
<td>59.5</td>
</tr>
</tbody>
</table>

The development of this database is a good example of how more data can be made available to the user and yet overall cost and burden imposed on the data providers has been reduced. As it evolves, the database should make a valuable contribution, not only to tourism and transport statistics, but also for regional, energy and environmental statistics.

Visa data

During 2005, following a number of requests from Tourism Ireland¹⁵ for detailed information on visitor numbers from specific countries such as China and India, the CSO began to look at alternate sources of data for non-EU visitors. While the CSO has a good Country of Residence Survey, which produces robust estimates for the main visitor countries (data for 26 countries are published on an individual basis), the number of visitors to Ireland from countries such as China or India is relatively small and consequently our estimates would not be sufficiently robust to publish. For example, China is included in the Eastern Asia aggregate, which for 2007 only accounted for less than 0.3% of all visitors to Ireland. Similarly India is included in the Other Asia residual, which in 2007 accounted for less than 0.5% of all visitors to Ireland.

¹⁵ Tourism Ireland is the All Ireland tourism agency responsible for marketing the island of Ireland abroad.
One source that appeared to have great potential was data on visas issued by the State to those countries outside the EU or that do not have a bilateral agreement with the Republic of Ireland. In Ireland the Department of Justice, Equity and Law Reform has formal responsibility for issuing visas. However under a delegated sanction, the Department of Foreign Affairs actually issue and process approximately 70% of all visa applications, including applications made directly to embassies in Moscow, Beijing and New Delhi. The Department of Justice, Equity and Law Reform deal with all other applications. The principle visa types issued are for the purposes of (1) Visit, (2) Study, (3) Business, (4) Conference and (5) Other (which include join spouse, family re-unification).

That two separate departments are both issuing visas made the task of constructing a workable dataset quite complex. The Department of Foreign Affairs could not provide any details regarding the visas they had issued. The Department of Justice, Equity and Law Reform could provide quite detailed information, but as they only account for approximately 30% of visas issued, and as the majority of visas issued to China and India would have been issued in Beijing and New Delhi via delegated sanction and that no data were available on the number of issued visas actually exercised (i.e. once a visa is issued there is no link with the border control to determine whether issued visa was actually used), or what proportion were multi-entry visas, this project was not pursued any further.

Ultimately, this project did not offer any useable data in the short term. However, as interest in immigration grows and concerns over security heighten it is likely that these data sources may improve in time. For example, recently unveiled plans by the Department of Justice, Equity and Law Reform to collate all international travel itineraries (sea and air) in a centralised database, as part of a plan to establish an electronic border control system may result in new data or in a link between visas or permits issued and border crossings being established.

Credit/Debit card data

This project began largely as a Retail Sales project, prompted by some innovative developments in New Zealand on constructing an experimental RSI (Barrow, 2004 & SNZ, 2008). While the prevalence of credit/debit card usage is nowhere as high in Ireland as it is in New Zealand, using electronic card transaction data could potentially supplement our existing retail statistics. They could also provide some interesting data on tourism expenditure both outbound (i.e. Irish cards used outside Ireland) and inbound (i.e. non-Irish cards used in Ireland). The sheer volume of data is breathtaking, for example in New Zealand in July 2008 alone, 84 million electronic transactions were made, valuing $NZ 4.6 billion.

The CSO is now working with the Irish Banking Federation and the Irish Payments Services Organisation (IPSO) to try and secure access to credit/debit card transaction data. The detailed credit card data can only be sourced from international payment associations, such as MasterCard International or Visa International. The domestic debit card (branded Laser) data can be sourced from the IPSO, however as this single domestic brand will be replaced by international card brands in the coming years, in line with the introduction of the Single Euro Payments Area in 2010, steps are being now to ensure data streams are safeguarded for the future.

In broad terms, CSO is interested in accessing electronic transaction data in order to examine whether we can improve our knowledge and understanding of the following activities:

- Irish expenditure abroad by tourists/travellers
- Cross border expenditure by Irish residents - furniture, alcoholic drink, etc.
• Expenditure in Ireland by non-residents (tourists)
• General appreciation of ICT growth for purposes of retailing

In order to do this, CSO in cooperation with IBF and IPSO is trying to establish three data streams:

• Internet transactions
• Cross-border transactions
• Transaction classified by merchant or product

In general there appears to be more data available for credit cards than for debit cards. For example, while both cards can provide the number and value of transactions, credit cards can also a CP/CNP\(^{16}\) breakdown. The CNP category can be further broken down into internet and mail order/telephone order (MO/TO) payments. Similar breakdowns are not currently available for debit cards in Ireland.

This project is only in its infancy. Access to the data is as yet uncertain. Nor is the quality of the metadata clear e.g. whether they have product classifications, geographic codes or personal/commercial card transactions. What is clear however is that it has been done in New Zealand, so why not Ireland?

HARNESSING TECHNOLOGY

Technology is advancing at an astonishing speed and with these developments powerful new data sources are being created. Transport statisticians have long recognised the potential of GPS to yield very detailed data on both vehicle and person movements (Stopher, 2008). More recently, given the high penetration rates that exist for mobile phones another source of data is emerging. As these mobile phone data already exist and don’t need to be created, the cost and completeness of coverage make these data very attractive. Even for those who need quite accurate location information beyond what is currently available from mobile phone data, the next generation of phones will be equipped with AGPS (or assisted GPS) providing a very accurate location.

There are already examples of how this can be achieved. The Massachusetts Institute of Technology (MIT) has been using anonymised data supplied by Mobilkom Austria and Telecom Italia from their mobile-phone networks (Economist, 2007). By tracking the movements of mobile phones from local cell to cell, researchers can by default track people and by monitoring direction, speed, frequency or regularity of movement derive a lot of additional information, such a mode of transport or even purpose of trip. Similar pioneering work has been progressing in South Africa (Krygsman et al, 2008). It should be noted that this is not a new idea. In the early 1960’s the Nobel Prize winning economist William Vickery installed a radio transmitter in his car to prove that people could be tracked (Vanderbilt, 2008).

The Central Statistics Office in Ireland is becoming increasingly interested in mobile-phone technology, both as a source of valuable data and also as a possible CAPI tool. Two ongoing development projects are outlined briefly: First off, the CSO is helping the University of Limerick (UL) in Ireland to test tourism questionnaires that use mobile-phone and Bluetooth technology. Secondly, with the proliferation of mobile-phones, both in Ireland and internationally, the CSO are trying to examine if the passive datasets be accessed and used to estimate outbound flows of Irish travel and inbound regional or sub-national travel patterns within Ireland.

\(^{16}\) Card holder present (i.e. in a retail outlet) /not present at place of transaction (e.g. phone transaction)
"Digital" tourism diaries

Compiling accurate tourism expenditure data is not straight forward. Many products are bundled often making tidy separation very difficult or impossible, services and products may be purchased and paid for well before the actual trip, others may not be paid for until long after the trip. Finally there are the added complications of memory decay and expenditure on consumer durables.

For some time, researchers at the National Centre for Tourism Policy Studies in UL have been working closely with Failte Ireland and the CSO in the development of a diary to measure tourism expenditure, as arguably this is the best way to compile accurate current expenditure data, overcoming typical recall difficulties and should also provide better product and service expenditure breakdowns.

To date, pilot studies have found:

- Diaries are comprehensible to the tourist population
- Diaries generate very rich data
- Diaries facilitate analysis of consumption patterns over the duration of a trip
- Diaries can allocate consumption to geographic location, which is vitally important for regional or sub-national analysis.

To date, pilot studies have been paper based resulting in some major feasibility issues:

- Diaries are expensive to provide and expensive and time consuming to key
- Paper based diaries tend to be bulky/cumbersome.

In order to try and overcome these issues, the Wireless Access Research Centre, the National Centre for Tourism in UL and the CSO have joined forces to examine if recent technological advancements in personal and hand held computing can overcome these issues. Using GSM/GPRS/EDGE, SMS and Bluetooth technologies, questionnaires or diaries can be transmitted to mobile-phones (or other devices such as Blackberrys) in an airport. Respondents could be asked to complete a questionnaire on the spot (suitable for travellers waiting in a departure lounge) or a diary (suitable for arriving tourists). By using respondents own mobile-phone/device as a CAPI tool, it should potentially reduce costs and reduce the need for training as respondents will be familiar with their own phone.

Before undertaking any field work CSO approached ComReg in Ireland to ensure there were no legal impediments. From a legal viewpoint, once any questionnaire transmitted by Bluetooth is clearly flagged as a CSO survey, immediately provides an Accept/Reject option and clearly flags any cost implications to the respondent, there are no legal barriers. However the use of SMS is subject to greater restrictions, as unsolicited SMS is considered SPAM from a legal perspective. The National Consumer Agency noted that as Bluetooth is free this would be preferable to SMS, as even if CSO provided a toll free number, foreign tourists using a non-Irish network would still most likely be charged a roaming fee.

During the summer of 2008 the CSO carried out the first phase in a series of pilot studies. The first study, examining overall feasibility and probable biases, conducted in a number of selected Irish airports examined what proportion of travellers had mobile-phones and whether or not they had Bluetooth functionality. Some of the key results found were:

---

17 CAPI – Computer Assisted Personal Interview
18 ComReg – The Commission for Communications Regulation. The statutory body responsible for the regulation of the electronic communications sector in Ireland.
Developing tourism statistics at the sub-national level: A Response – Issues and Insights

- 96% of travellers had a mobile-phone with them;
- Of those, 65% had Bluetooth functionality, 7% didn’t know whether they had or not;
- Of those with Bluetooth, only 39% had it enabled, 10% didn’t know whether it was enabled or not;
- Of those with Bluetooth, 83% reported they knew how to switch on/off function.

Other interesting findings were that 77% of respondents said they would not accept an unsolicited questionnaire transmitted to their phone. However, 58% said they would accept a questionnaire transmitted to their phone if they had first been approached by a representative from CSO.

Figure 6.1.1 – Respondents able to operate Bluetooth by Age Category

![Figure 6.1.1](image)

The survey also detected a clear age bias regarding ability to use/knowledge of Bluetooth functionality, with 76.3% of those aged 60 or more responding they either couldn’t use Bluetooth or didn’t know whether they could or not (compared with 16.3% of those aged between 14 and 25). The trend in Figure 6.1.1. is clear and intuitive. The older respondents are, the less likely they will be to use Bluetooth. Consequently even if a respondent has Bluetooth, the older they are, the less likely they are to complete an electronic questionnaire or diary. However if the diary data are to be used as supplementary data, this is not an insurmountable problem.

The next phase of the project will be conducted in selected Irish airports during the autumn of 2008. This phase will involve testing a replica copy of the existing paper based Passenger Card Inquiry. The purpose of this phase will be to test actual response rates for the electronic version versus the paper based version. It will also allow a basic assessment of data quality and how successfully transmission and extraction of data will work in a busy hub. For instance, retrieval of data using Bluetooth is a technical issue, as it has a short transmission range.

From a CSO perspective, our real long term interest is in the possibility of getting diary data to supplement existing survey data, particularly at a regional level. Consequently the next steps will involve testing that the existing telecommunications infrastructure is adequate, particularly for determining location (either through AGPS or GSM positioning). It will also involve testing respondent willingness to cooperate with a more invasive diary or will they
require incentives? The other benefit to CSO arising from this project is the strengthening of our relationship with UL. The CSO has much to learn from academia, both in terms of keeping up to date with the latest technological developments but also with regard to their tourism research.

This project is at a very early stage. Initial tests suggest that the technology works. However it appears there may be some significant respondent biases, particularly relating to age. Nevertheless as phone technology develops and becomes increasingly multi-functional we are all likely to become more dependent on them for a greater number of day to day uses, it is possible that this problem may solve itself in the longer term.

Passive data

As noted above, international research has already demonstrated the usefulness of mobile-phone data for developing transport models. Does the same potential exist for tourism? In CSO we believe the answer is yes. In fact we believe the potential of “passive” mobile-phone “roaming” data alone offers enormous potential for developing international travel measures, including origin, destination and route. If these data could be linked to survey returns, the potential might be even greater (particularly for identifying the tourism component).

During 2008 CSO began negotiations with ComReg and the Data Protection Commissioner about obtaining access to mobile-phone aggregates and microdata from mobile-phone operators. There are three data streams of potential interest to CSO.

- Outbound “roaming” data i.e. where do Irish mobile-phones travel to (primary and secondary destinations), how long do they stay, what routes were travelled. This could be of huge importance for cross-border flows.

- Inbound “roaming” data i.e. where do foreign mobile-phones travel to etc. Again there may be huge potential here for developing cross-border and regional flows.

- Mobile-phone movements within Ireland i.e. Irish phones in Ireland. We already know the potential of these data from international research in the transport field. This could potentially be extended to produce information on Same Day trips and sub-national flows.

There are in Ireland a number of legal issues surrounding the access and use of such data. Some of the data held by mobile-phone operators is done under the Criminal Justice (Terrorist Offences) 2005. It is not clear at this stage whether CSO can obtain access to these data. In Ireland, a mobile-phone number is considered personal information (similar to religion or sexual orientation) so it is not clear whether we can secure access to these data without first obtaining individual permissions. The situation regarding access to inbound “roaming” data is even less straight forward or clear cut. For example, foreign tourists (when in Ireland) may be governed by their own national data protection legislation rather than Irish legislation.

These developments are at an early stage. Provided all the legal access issues can be resolved, CSO must then persuade the mobile-phone operators to provide the data. CSO will also need to prepare and publish a series of clear protocols regarding the treatment of these data (e.g. how data are transmitted, stored, deleted and anticipated data life cycle). Most likely, it will be some years before these developments bear any fruit, but if they do, it will have been worth the effort. Even if such data only supplement existing data, they have the potential to provide very detailed geographic information, which holds enormous possibilities for sub-national and environmental research.
ETHICAL ISSUES

At this point it is probably worth making a few remarks on ethics. As the Central Paper notes, we are all leaving a trail of “electronic fingerprints” as we go about our day to day lives. These fingerprints are resulting in an ever increasing amount of personal data being stored. At the same time, National Statistical Institutes are under increasing pressure to produce more statistics, and often from within a diminishing budget. Globalisation and other trends are continually challenging the accuracy and relevance of statistics. NSIs are increasingly looking towards administrative data and other data holdings, not only as a cheaper and more efficient way of collecting data, but increasingly as the only possible source available. For example, as internet purchases increase in popularity, particularly for durable goods, the robustness and relevance of traditional retail sales indices are increasingly questionable (from the perspective of providing an indicator of consumer confidence and consumer consumption).

So it seems we have a marriage made in heaven. One the one hand NSIs are looking for alternate data sources and on the other hand our electronic fingerprints are producing the required data. However we must be careful. Administrative data sources often hold very detailed personal information and frequently hold universal identifiers that facilitate datasets to be linked. In Ireland, the Data Protection Commissioner understands and is even sympathetic towards the CSOs data requirements. However there is a counter balancing concern that accessing and holding huge volumes of personal data has its dangers. While such a protective or cautious stance is often frustrating to the statistician who can see the huge potential of the data, history has shown there are legitimate reasons for such concern.

Consequently NSIs and international organisations have a responsibility to ensure that data are managed properly. The protection of personal privacy is not a trivial issue and adequate safeguards must be put in place before data are transmitted, stored or disseminated.

CONCLUSION

The demand for sub-national statistics is growing. This poses a number of real challenges for data providers as sub-national data are expensive to compile and requires an alliance between a greater number of stakeholders.

There are potential solutions but all of these solutions will not necessarily work, either immediately or at all. We must be prepared to experiment, to fail in the short run but ultimately to find working solutions. Increased use of administrative data is an important step, both as a way of reducing costs and respondent burden, but also (in some cases) as way of getting superior data to survey data. Harnessing technology is another necessary step.

Cooperation is vitally important. Partnerships with national, regional and local authorities will become increasingly important, to ensure we are working towards common goals. Equally partnerships between universities, NSIs and other specialist organisations will allow us to develop collective solutions beyond our respective narrow expertise. Links with other statistical domains is also very important, both in terms of better contextualising and broaden the user base for tourism statistics but also to learn lessons from other research.

Finally, electronic storage of data has opened up a brave new world for the statistician. But we must be careful not to neglect our responsibilities, these data must be treated with respect. It is the collective responsibility of international organisations, NSIs and other agencies using data, to put in place proper data protocols and protection mechanisms to safeguard information relating to individuals and enterprises.
REFERENCES


Eurostat (2004), Harmonised Index of Consumer Prices (HICPs) - A Short Guide for Users, Luxembourg.


