Q 2016
European Conference on Quality in Official Statistics

Quality implications of the use of big data in tourism statistics

Session 27

3.6.2016

F. Cortina, INE, fernando.cortina.garcia@ine.es
M. Izquierdo, INE, maria.izquierdo.valverde@ine.es
J. Prado, INE, jesus.prado.mascunano@ine.es
M. Velasco, INE, maria.velasco.gimeno @ine.es

INE Instituto Nacional de Estadística eurostat

Madrid, May 31 - June 3

Q 2016
European Conference on Quality in Official Statistics

Why tourism?

Q 2016
European Conference on Quality in Official Statistics

A need for comparability

Traditional sources

DEMAND

- Residents travel surveys
- Border surveys

SUPPLY

- Surveys to accommodation establishments

Tourism Satellite Account

Balance of payments

New sources

Sensors: traffic loops and traffic control cameras

Mobile phones

Credit cards

Others: searches in Wikipedia, number of access to tourism websites, booking engines,….

Traffic loops and traffic cameras (I)

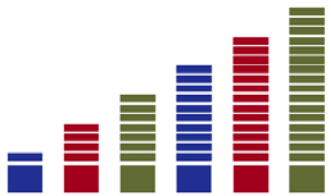
Traffic loops:

- Nº vehicles crossing the border:
  - Intervals of 15 min
  - By lenght
  - And direction

Traffic loops and traffic cameras (II)

time interval entry/exit small/medium/large

	ETD	DIA	HORA	MINUTO	PEQ_ENT	MED_ENT	LAR_ENT	TOT_ENT	PEQ_SAL	MED_SAL	LAR_SAL	TOT_SAL
1	E001	1	22	15	0	3	0	3	0	0	0	0
2	E001	1	22	30	0	0	0	0	0	0	0	0
3	E001	1	22	45	0	3	0	3	0	0	0	0
4	E001	1	23	0	0	3	0	3	0	0	0	0
5	E001	1	23	15	6	0	0	6	0	3	0	3
6	E001	1	23	30	6	3	0	9	6	3	0	9
7	E001	1	23	45	0	1	0	1	0	0	0	0
8	E001	2	0	0	21	9	0	30	6	15	0	21
9	E001	2	0	15	6	0	0	6	3	0	0	3
10	E001	2	0	30	0	9	0	9	0	0	0	0
11	E001	2	0	45	12	0	0	12	6	3	0	9
12	E001	2	1	0	3	0	0	3	0	0	0	0
13	E001	2	1	15	0	0	0	0	0	0	0	0
14	E001	2	1	30	0	3	0	3	3	0	0	3
15	E001	2	1	45	0	3	0	3	3	0	0	3
16	E001	2	2	0	0	0	0	0	0	0	0	0
17	E001	2	2	15	9	0	0	9	6	0	0	6
18	E001	2	2	30	0	0	0	0	0	0	0	0
19	E001	2	2	45	15	0	0	15	6	3	0	9
20	E001	2	3	0	6	0	0	6	0	0	0	0
21	E001	2	3	15	12	0	0	12	9	0	0	9
22	E001	2	3	30	12	0	0	12	6	9	0	15
23	E001	2	3	45	6	0	0	6	0	6	0	6
24	E001	2	4	0	0	0	0	0	0	0	0	0
25	E001	2	4	15	6	0	0	6	0	0	0	0



Traffic loops and traffic cameras (III)

Traffic cameras:

- Who is coming?
  - Vehicle nationality (residents/non residents)

Traffic loops and traffic cameras (IV)

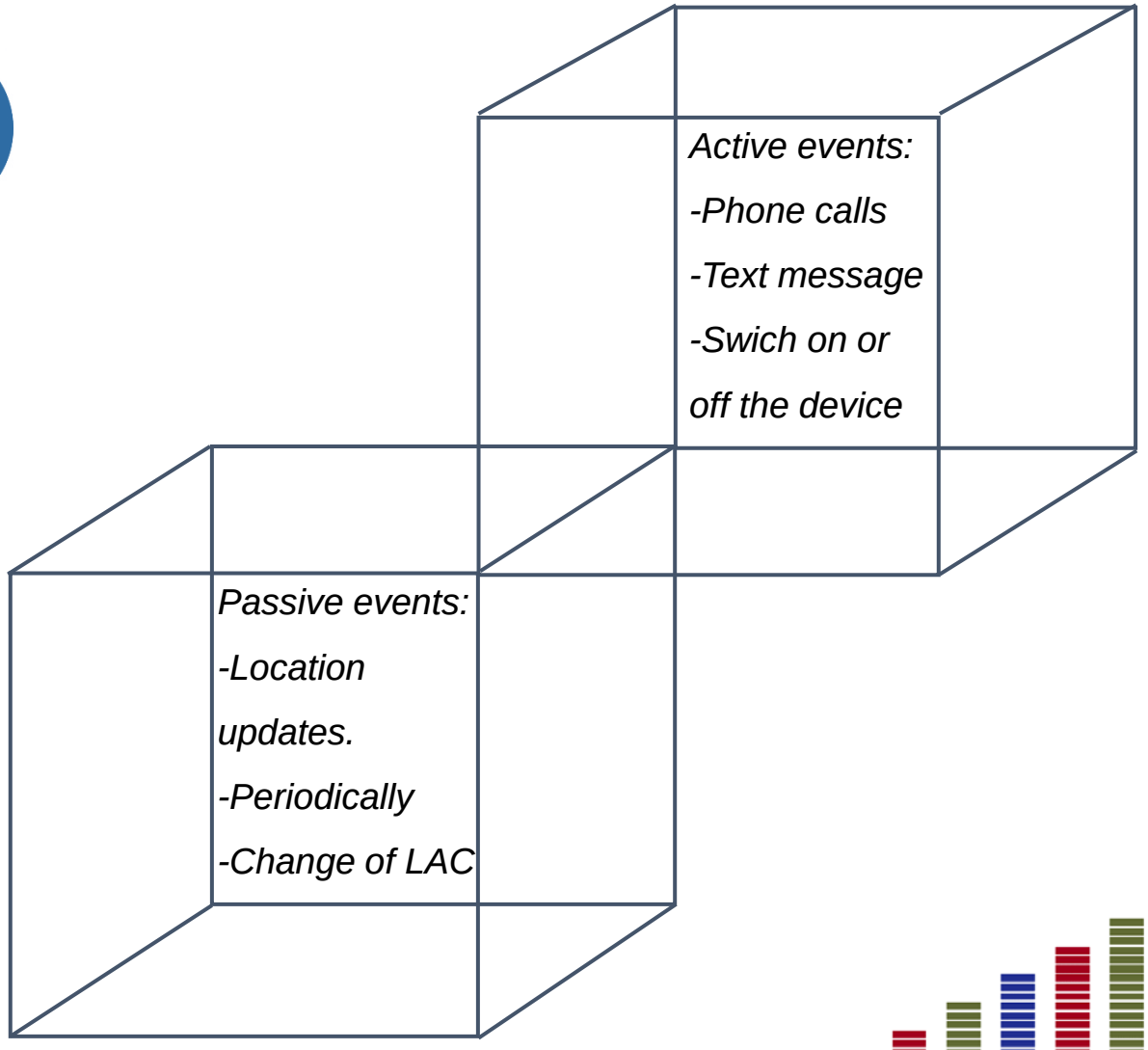
Sesion	ID	Dia	Mes	Año	Hora	Matricula	Nacionalidad	Punto	Sentido	Carril
PF17(09/03/2014)	0000472	09	03	2014	12:42:27		ESP	PF17	S	C1
PF17(09/03/2014)	0000473	09	03	2014	12:44:06		PRT	PF17	S	C1
PF17(09/03/2014)	0000474	09	03	2014	12:44:12		PRT	PF17	S	C1
PF17(09/03/2014)	0000475	09	03	2014	12:44:22		PRT	PF17	S	C1
PF17(09/03/2014)	0000476	09	03	2014	12:44:44		ESP	PF17	S	C1
PF17(09/03/2014)	0000477	09	03	2014	12:45:53		PRT	PF17	S	C1
PF17(09/03/2014)	0000478	09	03	2014	12:47:14		PRT	PF17	S	C1
PF17(09/03/2014)	0000479	09	03	2014	12:49:02		PRT	PF17	S	C1
PF17(09/03/2014)	0000480	09	03	2014	12:49:20		PRT	PF17	S	C1
PF17(09/03/2014)	0000481	09	03	2014	12:49:32		PRT	PF17	S	C1
PF17(09/03/2014)	0000482	09	03	2014	12:49:37		PRT	PF17	S	C1
PF17(09/03/2014)	0000483	09	03	2014	12:51:00		PRT	PF17	S	C1
PF17(09/03/2014)	0000484	09	03	2014	12:51:08		ESP	PF17	S	C1
PF17(09/03/2014)	0000485	09	03	2014	12:51:39		PRT	PF17	S	C1
PF17(09/03/2014)	0000486	09	03	2014	12:51:51		PRT	PF17	S	C1
PF17(09/03/2014)	0000487	09	03	2014	12:51:56		FRA	PF17	S	C1
PF17(09/03/2014)	0000488	09	03	2014	12:52:48		PRT	PF17	S	C1
PF17(09/03/2014)	0000489	09	03	2014	12:53:46		ESP	PF17	S	C1
PF17(09/03/2014)	0000490	09	03	2014	12:53:50		ESP	PF17	S	C1
PF17(09/03/2014)	0000491	09	03	2014	12:53:55		ITA	PF17	S	C1
PF17(09/03/2014)	0000492	09	03	2014	12:56:02		PRT	PF17	S	C1
PF17(09/03/2014)	0000493	09	03	2014	12:56:17		PRT	PF17	S	C1
PF17(09/03/2014)	0000494	09	03	2014	12:57:07		ESP	PF17	S	C1
PF17(09/03/2014)	0000495	09	03	2014	12:57:22		PRT	PF17	S	C1
PF17(09/03/2014)	0000496	09	03	2014	12:58:32		ESP	PF17	S	C1

Around 5 millions of data each month → % vehicles per nationality

Future: tracking the ID

Mobile positioning data (I)

Mobile phones connected to the network generate events that are recorded in a database associated to the cell phone ID.



Mobile positioning data (II)

Key points

- First approximation: extraction from the events database of one of the most important MNO operators in Spain.
- Goal: measuring the **number of tourists** both residents and non-residents and their **average stay**, broken down by region of destination (NUTS 2) and region/country of origin. Comparing the results obtained with those derived from official statistics.

Mobile positioning data (III)

Comparing mobile positioning data (MPD) with survey data (FAMILITUR).

Table 1. Distribution of trips by destination (residents) August 2014

	% trips MPD	% trips FAMILITUR
TOTAL	100,0	100,0
Andalucía	15,8	19,1
Illes Balears	2,8	2,2
Canarias	3,1	3,6
Castilla y León	8,7	10,1
Cataluña	14,8	13,4
Madrid, Comunidad de	10,7	3,4
Comunitat Valenciana	12,7	15,6
Rest of Regions	31,5	32,6

Table shows a quite similar distribution of trips among the regions of destination

Main differences are found in Madrid, where the percentage of tourists identified from MPD is three times bigger than in the survey

Mobile positioning data (IV)

Tourism trips:

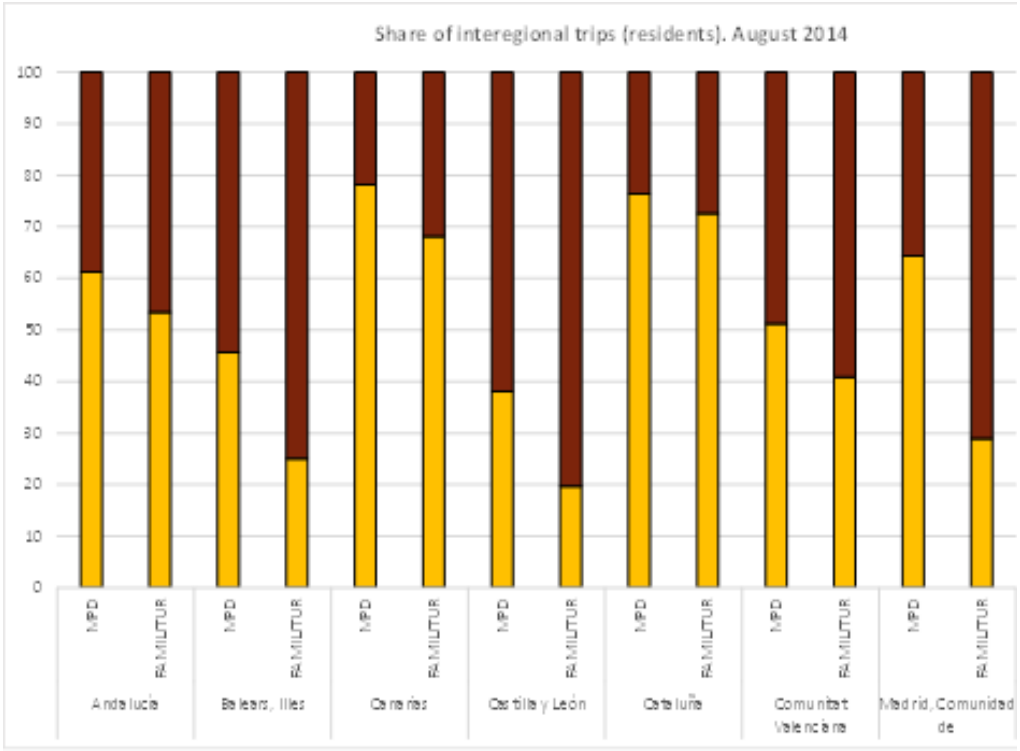
Trips to a main destination outsiede the usual environment, for less tan a year, for any main purpose, other tan to be employed by a resident entity in the place visited.

Usual environment:

Geographical área, not necessarily a contiguous one, within which an individual conducts his regular life routines.

Mobile positioning data (V)

Comparing interregional trips by destination

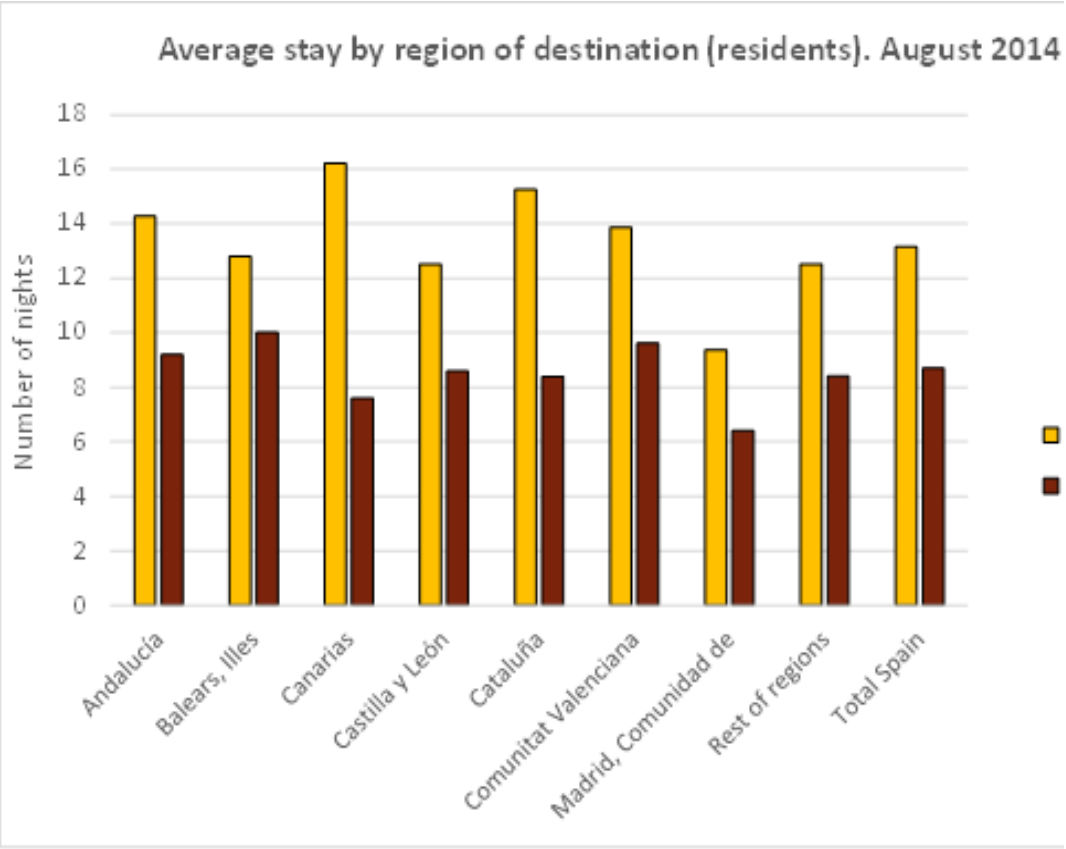


Data from MPD present a higher proportion of intraregional trips than the survey in all the regions represented

Madrid with the highest difference.

Mobile positioning data (VI)

Comparing average stay



In aggregate terms, mobile data shows an average of 13.5 nights for residents' trips to a destination in Spain

Survey data estimate is 8.7 nights

Mobile positioning data (VII)

Comparing for non-resident mobiles.

Table 2. Distribution of international tourist August 2014

	% tourist MPD	% tourist FRONTUR
TOTAL	**100,00**	**100,00**
Germany	16,24	14,22
France	18,52	21,88
Holland	8,69	3,69
Italy	6,34	7,18
Portugal	4,24	3,41
United Kingdom	17,09	23,52
USA	1,29	1,62
Rest of the world	27,59	24,48

The distributions of tourist by country of origin present slight differences

The most important countries are United Kingdom, France and Germany. In both cases Germany gets the third position

Mobile positioning data (VIII)

Comparing average stay for non-residents

Table 3. Average stay by country of origin August 2014

	MPD	EGATUR
TOTAL	**5,8**	**9,8**
Germany	7,0	10,2
France	4,9	9,4
Holland	5,6	11,3
Italy	6,1	8,1
Portugal	4,4	6,3
United Kingdom	6,2	9,8
USA	5,3	11,7

The differences are significant, always much higher the estimation of the survey.

Low averages from MPD could be explained by the fact that different legs of the same trip for the survey are considered as different trips in MPD

Data from credit cards: tourism expenditure

Information

- For residents, registers of all payments made by the bank clients in every point of sales terminal (POS) and ATM extractions with an entity card.
- For non-residents, information comes from payments or extractions in POSs or ATMs in the BBVA network

Comparison

- Average expenditures per trip shows higher values for the credit cards' data (CCD) series.
- Seasonality seems to be softer in the CCD series.

Unbundling package tours (I)

Package tour

- Transport + accommodation + (other travel services)

Tourism Satellite Account/ Balance of Payments

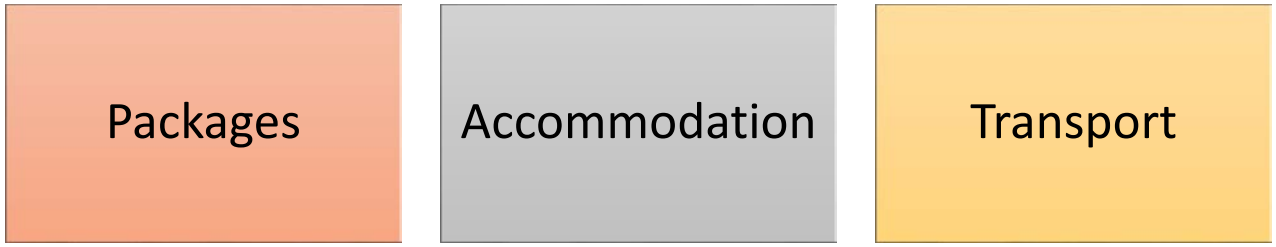
- Each component must be treated individually

Sources

- Similar non-package trips: no margins
- Statistics on Products in the Services Sector: only one structure

Unbundling package tours (II)

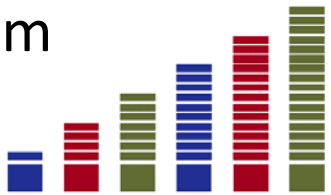
Pilot project to collect prices on



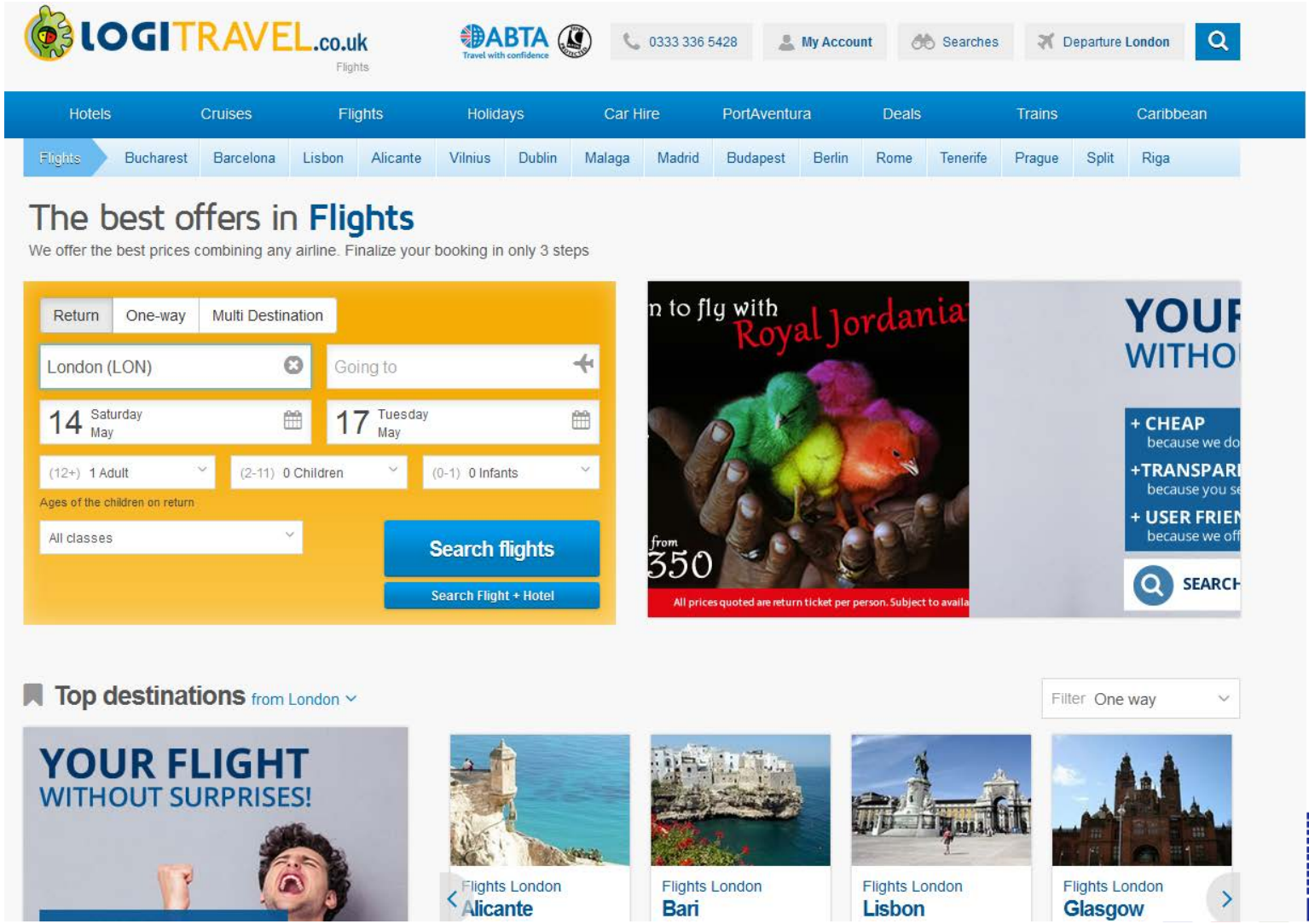
using webscraping techniques…

Origin	Destination	Departure	Nights	Number of room	Hotel name	Adress	Category	Board basis	…	Airline	Stops	…
Madrid	Canary Islands	Saturdays	7 nights	1 room (2 adults)								
Barcelona	Balearic Islands	Thursdays	3 nights									
	European capitals											

…and obtain a more specific structure combining them



Unbundling package tours (III)



Unbundling package tours (IV)

Conclusions

Quality standards

1 Efforts have to be made by statistical authorities and data providers to obtain results with the quality standards actually achieved by current official statistics.

Conceptualization

2 An in - depth conceptualization exercise should be made in first place to identify the phenomenon to be measured and second to assure comparability over time and between countries.

Control

3 Official statisticians need to have big control of the original databases, how they are processed, every assumption made, etc. with the highest level of detail.

General

4 The new sources of information are really promising as they can provide accurate information about the tourism phenomenon, allowing a more detailed geographical analysis.