

MRS Census Geodemographics Group

Can Big Data replace the Census?
What does Big Data give us now?

**Hunting better data – what are the
priorities of commercial users?**

7 March 2016

Keith Dugmore
Demographic Decisions Ltd.

Big Data – a preamble

- Gartner (2010): High Volume / High Velocity / High Variety
- McKinsey (2011): “Big data: The next frontier for innovation, competition, and productivity”
- House of Commons Science & Technology Select Committee report (Feb, 2016): “The big data dilemma” – opportunities & ethics
- Mike Whitelegge, M&S: “Big Data? More data!” – transactions
- Professor David Hand (Imperial); Dr. Harvey Lewis (Deloitte)
- MRS’s questions (reversed): What does Big Data give us now? Can Big Data replace the Census?

Outline

- Commercial users – members of the DUG
- Why? Starting with business questions
- Hunting better data (theory, & in practice)
- We live in exciting times: current initiatives to mobilise big data
- A replacement for the Census?
- + some useful links



Making Data Work

YOUR M&S



The **co-operative**



centrica

John Lewis



WHITBREAD



Informed by DUG

- 14 member Business-To-Consumer companies
- Diverse sectors
- Millions of customers (& transactions) – analysing behaviour
- Members' quarterly meetings discussions (since 1998)
- Conference, e.g. 2012: “Retail Issues, Big Data, and Research”
- Posing questions (curiosity) / hunting data / driving decisions

“What’s rattling my cage” at recent DUG meetings

Traditional questions (Best customers? Retention? Finding more? New stores? Closing stores? Targeting advertising?) plus.....

- How are online sales changing?
- What’s the effect of digital trading on physical stores?
- How should we attribute sales?
- How should we organise click & collect?
- How can we track intra-shopping centre footfall?
- Will cash continue to be used?
- What’s the best source of parking data?
- What will be the impact of Crossrail on the South East economy?

Masters' research projects with ESRC, 2012+ – and hunting data (some Big, but not all)

Company	Project	Student	Supervisor	University
Barclays	How are changes in consumer behaviour impacting banking provision?	Harman Sandhu	Abhinay Muthoo	Warwick
Boots	The delivery person as face of the company, and its impact on customer care	Shahzar Anwar	Mark Passera	Kingston
Boots	Defining retail conurbations	Matthew Pratt	Jim Wright	Southampton
Camelot	Wealth and Deprivation - developing a consistent indicator across the entire UK	Alistair Leak	James Cheshire	UCL
Co-op	Shopping missions - Main, Top-up, Distress	Katherine Ryan	Paul Longley	UCL
Co-op	Pedestrian flows - Agent-based modelling of footfall in town centres and retail parks	Menghai Lan	Alison Heppenstall	Leeds
Co-op	Competitor evaluation	Remi Martins-Tonks	Gregory Crawford	Warwick
Co-op	Membership and Loyalty	Samuel Tony Brown	David Reinstein	Essex
DUG	Store Location data -what are the best sources?	Benjamin Bond	Martin Clarke	Leeds
Dixon's	Internet killed the Radio Store	Samuel Jones	Nigel Walford	Kingston
Dixon's	What happens when a competitor enters the market?	Guy Lansley	Mark Birkin	Leeds
John Lewis	Transport trends and Retail centres	Thomas Murphy	John Stillwell	Leeds
M&S	High Streets - Classification and predictions	Benjamin Clark	Mark Birkin	Leeds
Tesco	The Olympics and its possible effects on Tesco.com's business	Helen Parker	Mark Birkin	Leeds
Tesco	Green initiatives and their impacts on consumers	Emily Eldridge	Tim Butler	KCL
Whitbread	Weather histories for regions, since 2000	Yasmin Dogan	Paul Norman	Leeds



DUG Conference 2015

ESRC Prize: Anastasia Ushakova, UCL

Presented by Bruce Jackson, Senior Knowledge Exchange Manager, ESRC

British Gas project:

“Can we identify vulnerable energy customers in the UK using smart meter data?”

Hunting relevant data: a) existing customers

- Seeking to better understand own customers – BIG data
- M&S (Mike Whitelegge speaking in 2014) about events per week:
 - 21 million customer visits to stores
 - 12 million baskets
 - 60 million items of clothing & food
 - 17 million payment events
 - 5 million non-M&S baskets (using M&S card)
 - 5 million visits per week online
 - Plus Facebook likes, & Twitter followers
 - All data held for 4 + years
 - Analysis & actions (but nothing “creepy”)

Hunting relevant data: b) external sources

- Unit records (people, addresses, competitor locations, etc.)
- Sample surveys (e.g. Living Costs & Food)
- Aggregate statistics or estimates for Output Areas (e.g. Census residents, workers, & commuters; Neighbourhood Statistics (NB); geodemographic classifications; and estimates of market sizes)
- Postcode directories and look-ups
- Social media
- Map data (boundaries, points, background wallpaper, and networks for travel times)

Detailed mapping + Output Area Stats (London OAC)

DataShine LOAC Geodemographics for London based on the 2011 UK Census

Land Labels Postcode: Go

Intermediate Lifestyles High Density & High Rise Flats Settled Asians Urban Elites City Vibe London Life-Cycle Multi-Ethnic Suburbs Ageing City Fringe

Important note: Classifications are an average across the local area, rather than for individual houses, therefore the colour coding on a building is not necessarily indicative of that building.
Want to find out more about what your local classification means? Don't agree with it?
You can [find out more](#) or [choose a better one](#) at [Open Geodemographics](#).

High Density & High Rise Flats
Students and minority mix

Islington
Area Code: E00013461 City: London Classification Code: B3

2011 OAC
An alternative classification covering the whole of the UK. Switch to it.

ESRC Consumer Data Research Centre
Created by Alex Singleton & Oliver O'Brien

2 km

ENG 09:55
UKX 28/02/2016

Where might such data be found? Sources

- Open Data
 - www.data.gov.uk
 - Benefits: Free; Numerous (c. 28,000); Range of topics; Possible proxies
 - Challenges: Hunting; Formats / Ease of use; Ethics
- Value-Added Resellers
 - Cost – but worth it; specialist expertise
 - Additional proprietary data sources, e.g. pooled databases
 - Speed, convenience

The ideal data.....a checklist

- Relevant topic?
- The population, & subject detail?
- Geographical detail – Output Areas?
- UK-wide?
- Recent?
- Coverage, & quality?
- Any damage by Statistical Disclosure Control?
- Accessible?

But, data & decisions in practice.....

- Perfection?
- “National Statistics”?
- The need for speed of analysis & decisions
- Pareto’s 80:20
- “It is better to be vaguely right than exactly wrong” (Carveth Read, 1898, and J M Keynes too)
- Firmi estimates – beyond piano tuners in Chicago
- The vital importance of good metadata (to assess the checklist)

DUG's priorities

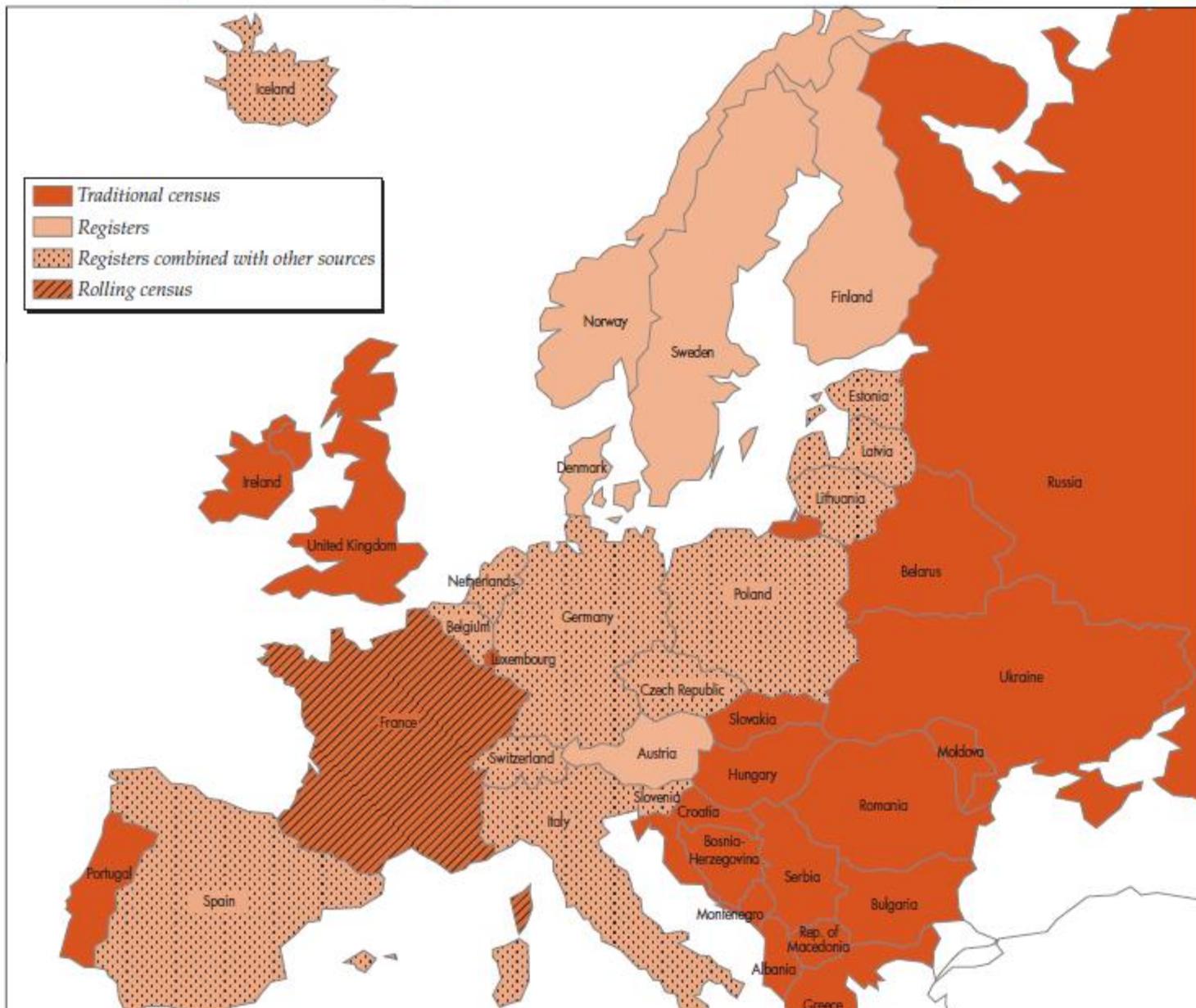
- The (definitive) National Address Gazetteer, and OS Map data enabled by the Public Sector Mapping Agreement
- Counts of people by location, and by time of day, from mobile phone data for Output Areas (not masts)
- Aggregate statistics on Income & Wealth at Output Area level, created from government administrative files, e.g. HMRC, DWP – now!
- ONS to pool companies' transaction data to create timely estimates of prices and growth, and statistics for small areas on market sizes and sales channels (especially online)

.....All are BIG

Current initiatives to mobilise big data

- Administrative Data Research Network (ADRN) – ESRC
- Consumer Data Research Centre (CDRC) – ESRC
- Easing access to microdata (& creating another role for resellers)
- HoC Science & Technology Select Committee report
- Sir Charlie Bean's Independent review of UK statistics
- Cabinet Office proposals for Data Sharing – consultation
- So, can Big Data replace the Census?.....

Map - Methods used by European countries in 2010 and 2011 to enumerate their population



Can the UK create good small area statistics with Big Data?

- Geography – Output Areas
- Frequency – e.g. annual
- Topics
 - Additions (e.g. Income); also proxies
 - Omissions (e.g. Language)
- All UK?
- Coverage, & Quality?
- Change / Instability?
- Multivariate analysis, & Government Data Sharing
- Non-Government Big Data

Some useful links (1)

- Gartner (2010): High Volume / High Velocity / High Variety
<http://www.gartner.com/it-glossary/big-data/>
- McKinsey (2011) report: “Big data: The next frontier for innovation, competition, and productivity” <http://www.mckinsey.com/business-functions/business-technology/our-insights/big-data-the-next-frontier-for-innovation>
- House of Commons Science & Technology Select Committee report (Feb, 2016): “The big data dilemma”
<http://www.parliament.uk/business/committees/committees-a-z/commons-select/science-and-technology-committee/news-parliament-2015/big-data-dilemma-report-published-15-16/>

Some useful links (2)

- Mike Whitelegge (M&S): Why big data is a big deal
<https://www.youtube.com/watch?v=3LtcVY63vjQ>
- DUG Conference 2012 “Retail Issues, Big Data, and Research”
<http://www.demographicsusergroup.co.uk/22.html>
- Professor David Hand (Imperial College): “Big Data”
<http://www.demographicsusergroup.co.uk/resources/David+Hand.pdf>
- Dr Harvey Lewis (Deloitte) “Open, Mobile, Gigantic – OMG”
<http://www.demographicsusergroup.co.uk/resources/Harvey+Lewis.pdf>
- Datashine: <http://blog.datashine.org.uk/>

Some useful links (3)

- Administrative Data Research Network (ADRN) – ESRC
<http://adrn.ac.uk/>
- Consumer Data Research Centre (CDRC) – ESRC
<https://www.cdrc.ac.uk/>
- Sir Charlie Bean's Independent review of UK statistics
<https://www.gov.uk/government/publications/independent-review-of-uk-economic-statistics-interim-report>
- Cabinet Office proposals for Data Sharing – consultation
<https://www.gov.uk/government/consultations/better-use-of-data-in-government>

Keith Dugmore
Demographic Decisions Ltd.

Tel: (0044) 020 7834 0966

Email: dugmore@demographic.co.uk

Web: www.demographic.co.uk