

New Data: Sources, Governance, Infrastructure, Analysis

Tim Holt
UK Data Forum

- What do we mean by new forms of data?
- What the UK has done so far?
- What do we mean by infrastructure?
- Access
- Going Forward: Key aims and challenges

New Data Sources

Broad Category	Detailed Category	Examples
A: Government Transactions	Ind Tax Recs Corp Tax Recs Property Tax Recs Social Security Payments Import/export Recs	Income tax, tax credits Corp tax, sales tax, VAT Property sales and/or value taxes State pensions, hardship payments, unemployment benefits, child benefits Border control records, import/export licensing records

New Data Sources

Broad Category	Detailed Registrations	Examples
B: Government And other Registration Records	Housing and Land Use Education Criminal Justice Social Security Electoral Population Health System Vehicle/driver	Ownership Pupil attainment, school inspection Police and court records Register of eligible persons Voter registration records Births, marriages, deaths etc etc Personal medical records, hospital records Driver and vehicle license registers

New Data Sources

<p>C: Commercial Transactions</p>	<p>Store Cards Customer Accounts Other customer records</p>	<p>Supermarket loyalty cards Utilities, financial institutions, mobile phone usage Product purchases, service agreements</p>
<p>D: Internet usage</p>	<p>Search terms Website interactions Downloads Social Networks Blogs, New Sites</p>	<p>Google, Bing, Yahoo search activity Visit statistics, user generated content Music, Films, TV, i-Books Facebook, Twitter, LinkedIn Reddit</p>

New Data Sources

<p>E: Tracking Data</p>	<p>CCTV Images Traffic sensors Mobile phone locations: GPS data</p>	<p>Security/Safety Camera recordings Vehicle tracking records; vehicle movement records</p>
<p>F: Satellite and aerial imagery</p>	<p>Visible light spectrum Night-time visible emissions Infrared, radar mapping</p>	<p>Google earth © Landsat</p>

What problems do the data sources pose for research purposes?

- Access: How to get it for research purposes?
- Provenance: How were the data created, coverage?
- Permanence: How durable, can research be replicated?
- Comparability: How suitable for research that compares and contrasts groups?
- Comparability: time series?
- Legality: What laws govern their use for research?
- Ethics: Should the data be used for research purposes?
- Linkage: Can different types of data be linked and, if so, how?
- Structure: Often complex, non-standard and linkage creates more gaps
- Skills and Capacity: Do we have the necessary skills, knowledge, tools to unlock the research potential?

UK Actions so far

ESRC, National Statistician, Statistics Authority,
Government Departments

Big Data: Three Phases

- Administrative Data (Government)
 - Running since late 2014
- Business and Local Government
 - Centres launched late 2014, being established
- Social Media & Third Sector
 - Call out for demonstrator projects

Phase 1: ADRN

- Centres in E,S,W,NI plus overarching AD Research Service (Essex)
- Not just single data sets
- Strong Governance Arrangements
 - Board
 - Accountable to Parliament
 - Public benefit condition and published results
 - Accredited researcher
 - Review process: ethics, feasibility, public benefit
 - Elaborate access arrangements: safe setting, trusted third party

Phase 2: Business and Local Government

Urban Big Data (Glasgow)

Business and Local Government (Essex)

Consumer Data Research Centre (UCL & Leeds)

Phase 3: Social Media and Third Sector

- Call for Civil Society Data Partnership Projects
- Social Media (de Roure – Oxford)

Infrastructure: Everything not project specific

- **Centres:** creation, governance, promotion, supporting researchers: applications, access and use
- **Methodological Research**
- **Training: Capacity building**
 - Main focus on PhD and early career research staff
- **Access**
 - Breaking down barriers
 - Legal (and resource) impediments

Access

- Government statements are helpful but not enough
- Government is a complex arrangement of semi-independent entities
- Legal arrangements (+interpretation) are very data-source and Department specific
- Need for:
 - Joined up Governmental approach
 - New legislation to remove some impediments
- Other impediments: resource implications

Key aims and challenges

- Key aims
 - Access to different data sources in safe settings
 - Support researchers in using the data
 - Centres of expertise
 - Developing relationships with private, public, civil society organisations
 - Catalyse innovative research
 - Operate as a network – Big Data and wider
- Key Challenges
 - Access
 - Exploitation
 - Capacity – including methods
 - Cross-disciplinary working
 - Combining and linking further
 - Privacy, ethics and security
 - Public engagement
 - Sustainability

Useful Links

- www.adrn.ac.uk
- www.statisticsauthority.gov.uk/national-statistician/administrative-data-research-network
- www.esrc.ac.uk/research/major-investments/Big-Data
- www.esrc.ac.uk/research/data/NDS/index.aspx

- timholt999@btinternet.com