Accessing the UBDC’s Data Service
University of Glasgow
15/8/2016

Andrew McHugh, Senior IT and Data Services Manager
Urban Big Data Centre
The Urban Big Data Centre (UBDC) is a research resource promoting the use of innovative methods and complex urban data to address global city challenges.
Our Objectives:

- Develop novel solutions for using and sharing urban big data
- Provide high quality training and outreach activities
- Deliver cutting-edge research
What we offer

The Urban Big Data Centre aims to support organisations through:

- **Data services** offered to a wide range of users, including academic researchers, policymakers, businesses, third sector organisations and everyday citizens

- **Access to wide spectrum of data**, from open data to more sensitive information

- **Personalised and secure user environments** for data users to curate, manipulate and link data

- **Expert support for data users** who want to use our data resources, including tools and software

- **Consultative support for data owners** to help interrogate their data and extract useful information for policy, business and citizen engagement

- **Training and skills development**, including technical, methodological and policy-focused sessions for both academic and non-academic researchers

- **Knowledge partnerships** with organisations and members of the public to scope and develop solutions for substantive urban issues, both local and global
Partner with us:
We support policymakers, businesses, third sector organisations and everyday citizens to harness the potential of big data to develop solutions for environmentally sustainable, economically resilient and socially just cities.
About: the data service team

- Scientific Computing Officers and Data Scientists
  - GIS, Administrative Data and Research Data Management
  - Proposal development, data sourcing and data management
  - We do not provide research services

- Data Service Officer
  - Curation services

- IT Service Officer
  - Supporting infrastructure for data management and accessibility
Our services

- Providing **access to our data** including **open** and **other licensed** data
- Facilitating access to **controlled administrative data** (via our third party safe haven partners)
- A **data sourcing** and procurement service
- Offering our expertise to facilitate the use of your own data
- **Training** programme

*We do not fund research staff or other resources – we do have limited resources to assist with proposal development, data sourcing and data management*
3 levels of data

- **Open data**: can be placed in the public domain with minimal restrictions on use
- **Safeguarded data**: have licence conditions or other considerations that mean access has to be restricted
- **Controlled data**: could potentially provide information about individuals
Open and safeguarded data

Many of our current data fit into the following categories:

– Open datasets
– iMCD Project data
– Geospatial data
– Transport data
– Housing data
– Education data
Open datasets

- CKAN data portal – [http://ubdc.gla.ac.uk/dataset](http://ubdc.gla.ac.uk/dataset)
  - Population
  - Census
  - Demography
  - Employment
  - Transport
  - Health
  - Education
  - Gender
The UBDC's *Integrated Multimedia City Data* project has gathered a unique and exciting collection of data looking at urban life in Glasgow.

A range of data collection methods including:

- A **household survey** and travel diary
- **Sensor data** collected from individual subjects who wore life-loggining and GPS devices
- A **multimedia information retrieval** exercise from online sources
- A further information retrieval exercise looking at **textual data sources online**
- **Satellite data**

An **integrated** and **linked** picture of urban life in Glasgow, attained from multiple perspectives.
iMCD (2) : household survey

- An anonymised survey covering attitudes, beliefs and lifestyle related to transport, education, technology usage and sustainability
- c. 1500 households, 2095 adults in the Glasgow Clyde Valley
- April to November 2015
- Anyone can use this data (Creative Commons CC-BY licence)
iMCD project data (3) : text retrieval (Twitter)

- Data collection of tweets originating from Glasgow area (12/14 – 11/15)
- Insights into citizens’ behaviour, reaction and moods in certain contexts or at particular times
- Anyone can use this data but Twitter’s terms restrict us to releasing just tweet IDs
Geospatial Data (1) : Spatial Urban Indicators

- Small-area estimates of various economic, social, built environment and physical indicators
- Covering 14 built up areas throughout the UK
- Anyone can use these data – current developments exploring issues of labour market outcomes and transport accessibility and availability
Geospatial data (2) : LiDAR

- Applications for local government planners who need detailed structural plans of the city.
- LiDAR is also regularly used by archaeologists, computer scientists, and civil engineering companies.

LiDAR — Light Detection and Ranging — is a remote sensing method used to examine the surface of the Earth.
Transport data: Strava Metro

- Covering Scotland (13–16) and Tyne and Wear (15-16)
- GIS compatible data with minute-by-minute counts of users at street level and waits at intersections

Strava is a social network for athletes, recording data from its users who upload cycle rides and running activity via smartphone or GPS device.
Housing data: Nestoria and Zoopla

- Zoopla offers current and historical data (to 2010); Nestoria a single year from 3/15-16
- Non-commercial use only

House price and rental advertisement details from these property search aggregators
Education data : UCAS

- Dataset comprises aggregated queried data for full- and part-time students
- Coverage from years of entry to Higher Education study for years 2000 to 2015
- Glasgow, Scotland and whole of UK are represented among the data
User spotlight (1) : open & safeguarded data

- A University of Stirling user is using housing data from the UBDC in their research project, which aims to provide insights that will guide city governments in using ICT to engage citizens in governance of sustainable cities.
- Likewise, another Stirling researcher is currently interested in using UBDC data to compare air pollution and traffic flows across major UK cities.
- A transportation researcher from the University of Aberdeen is using transport data from the UBDC to explore the use of data generated in apps and other self-monitoring systems in active travel planning.
- A researcher from the University of Warwick is using our geospatial SUDS data to map the general quality of life of people with the long-term chronic illness of Coeliac Disease.
- A local authority's Chief Executive Office is using our IMCD project data to understand citizen values and behaviours.
- A private researcher is using housing data to explore the impact of transport infrastructure on housing prices and rents using hedonic regression.
End user licensing

- Prospective users of UBDC’s data services will be expected to sign our End User License Agreement, as well as accept the terms of any given dataset.
- Terms vary, e.g., some data sets can be used for commercial research, and others are more restrictive.
- More details are available on request.
Controlled data service (1)

- We work with the *electronic Data Research and Innovation Service (eDRIS)*, to provide researcher access to controlled data.

- Controllers of **personal data** (which is data that may identify an individual living person, or could do so when combined with other information) may agree to permit others to have access for specific research purposes where there is a public interest case for them to do so.

- We do not acquire such data, but can help **approved researchers** negotiate access to a **secure environment** to work with controlled data.
Controlled data service (2)

In order to gain access to controlled data, researchers' experience and competence for undertaking their proposed research will be assessed, and the following requirements met:

- Researchers must gain approval from our Research Approvals Committee (RAC). Projects by postgraduate students will be considered, but supervisors must be listed on the project as lead researchers.
- Data controllers must provide approval for each project, meaning the data owners may require additional applications.
- Researchers must be accredited by undergoing approved researcher training so they are aware of their legal responsibilities.
Controlled data service (3)

- Project development
- Project approval
- Formal agreements and Research
- Post-research
Controlled data service (4)

- Are there organisational benefits for the data owner? (they may have limited resources to support access to their data, which they are typically not obliged to offer)
- Could the research cast the data owner in a negative light? Can such concerns be allayed?
- How can you engage early with the data owner?
- Is there a ‘social licence’ for the proposed project? Would the public consider it to be a legitimate use of data?
- What are the public benefits of the research, as well as academic merits?
User spotlight (2): controlled data service

- ‘Looked after children’ at home and away from home: how do they contribute to inequalities in health in Scotland
- Renfrewshire Council social care data analysis
- Supporting independent living through the prediction and prevention of falls
Data sourcing service (1) : About

- UBDC may be able to source and acquire data not already held in our data collections
- Research projects for which data is needed must be approved by our independent approvals committee
- Since data sourcing may result in additional costs (UBDC has a limited budget) all data sourcing requests also require approval based on several criteria
Data sourcing service (2) : Criteria

- In scope…
  - Relevant to urban research?
  - Within the definition of, or support working with big data?
  - Is UBDC the most appropriate body to acquire it?
- Strategic priority…
- Usable and reusable data…
- Data availability…
- Data life expectancy…
- Level of required technical support…
- Time sensitivity…
- Legal and ethical issues…
How to Apply to Use UBDC Resources

- Current Call for Expressions of Interest
- Eligibility
- Time scales
- The process explained
Call for Expressions of Interest

- We have two *Calls for Expressions of Interest* each year – **one is currently live!**
- Access to data within our collections can generally be arranged at any time
- Prospective users of our *Data Sourcing* and *Controlled Data Services* must submit an application by **9th September 2016** to be considered within the current call


- Full proposals are evaluated by UBDC’s *Research Approvals Committee* – assessed in terms of **quality, potential for impact** and the **appropriateness of use of UBDC resources**
Eligibility – who can submit an interest form?

- Academics in any discipline (including Masters and PhD students)
- Local authorities, government departments and public agencies
- Third sector organisations and businesses
- Citizens with an interest in using data to understand and improve their local communities

Unfortunately we cannot support non-UK researchers, although our open datasets and some safeguarded data can be accessed by anyone, regardless of where they are based.

We are interested in hearing about other opportunities for international collaboration!
Timelines and Deadlines

- **9th September 2016**: Deadline for submission of expression of interest
- **23rd September 2016**: Guaranteed response date (to provide access to data and/or advise on whether your proposal is in scope and feasible to proceed to the full proposal stage)
- **18th November 2016**: Full proposal due (applicants will be assigned a named UBDC support contact who will assist with the completion of the proposal)
- **13th January 2017**: Notification of outcome of proposal
Questions?

Contact: Andrew.Mchugh@glasgow.ac.uk

Call for Expressions of Interest: Apply to use our Data Services

If you want to access our novel and innovative urban data, or use our other data services, submit your interest by 9 September, 2016!

http://ubdc.ac.uk/data-services/data-services/call-for-expressions-of-interest/