The Consumer Data Research Centre and its Data Service

Sarah Sheppard
Project Manager
Who we are...

• Bringing together world-class researchers from Leeds, UCL, Liverpool & Oxford.
• Funded by the Economic and Social Research Council as part of the Big Data Network Phase 2 Business and Local Government Data
  • Urban Big Data Centre
  • ESRC Business and Local Government Data Research Centre
  • Consumer Data Research Centre
What we do...

- Multi-institution laboratory that discovers, mines, analyses and synthesises consumer-related datasets from around the UK
- Our objective is to create a service for social science researchers, to harness the potential of consumer-related Big Data for the benefit of society.
- Creates, supplies, maintains and delivers consumer-related data to a range of end users, alongside a programme of research training and outreach activities.
Datastore aims...

- Catalogue all CDRC data deposits
- Attributes are searchable (Metadata)
- Provide access to CDRC Data Products
- Make life simple...
  - Visual and interactive interfaces
  - Training & Materials
  - Local Authority Datastores
Browse by Product
Catalogue...

Browse by Topic
Value Added

• Cleaned data
• Spatially referenced
• Added attributes
• More usable formats
• Attribute descriptions
• Available for national, regional or local extents
Data Products...

Data description

Attribute descriptions

Shapefiles

Cleaned CSV files
- Easting / northing
- Lat / Lon
Data Products...

Retailer / Consumer Data Partners

Open Data + Consumer Data Research Centre + CDRC Academic/Researcher/Postgraduate student = Internet User Classification
CDRC Maps...
Local Authority Datastores
Coming very soon...

Bricks Versus Clicks: Mapping Retail Centre Exposure to Online Shoppers and Material Deprivation

1 Introduction
1.1 Learning Aims and Objectives

The Introduction to QGIS 2016 course provides an introduction to QGIS, an open source Geographic Information System software package. You will be introduced to the various components of the software and learn how to work with both vector and raster data layers. You will learn how to run simple geospatial analyses, how to edit existing maps and how to create new ones using your own data.

https://www.cdrc.ac.uk/events/category/training-session/
UKDS 5 safe principles:

- **Safe Projects**: Research Approvals Process
- **Safe People**: National Accredited Researcher Training
- **Safe Setting**: Secure Labs
- **Safe Outputs**: Stringent output checks
- Resulting in - **Safe Data**
Safe People

- Safe Projects

Safe Setting

Safe Outputs

Safe Data

Open Data Service

- Register to download
- Download open data
- Undertake research

Safeguarded Data Service

- Researcher registration
- Proposal development, assessment & approval
- Safe Researcher Training
- Safeguarded data available for secure download
- Create summary of findings
- Publish & archive as appropriate
- Send to CDRC
- Inform CDRC

Secure Data Service

- Researcher registration
- Proposal development, assessment & approval
- Safe Researcher Training
- Access to agreed controlled data at secure CDRC sites
- Perform analysis at secure site and submit to CDRC approver pool
- CDRC review analysis and confirm it is in line with project approval
- Analysis released to researcher
- Create summary of findings
- Publish as appropriate
- Send to CDRC
- Inform CDRC
The CDRC supports a wide range of users to carry out research projects that provide fresh perspectives on the dynamics of everyday life, problems of economic well-being and social interactions in cities.

- Health
- Retailing
- Policing
- Residential & Social Mobility
- Ethical & Sustainable Consumption
• Data Acquisition Strategy
  o Strategic Approach
  o Reactive Approach
  o Passive Approach
  o Co-producing data

• Ladder of Engagement
  o Research Masters Dissertation Programme
  o PhD/Internships
  o Internal research projects
  o Data acquisition for Centre under Data Licence Agreement
<table>
<thead>
<tr>
<th>Project Description</th>
<th>Sponsor</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>Using Big Data to understand the drivers behind store footfall.</td>
<td>The co-operative</td>
<td>Open</td>
</tr>
<tr>
<td>Understanding the Grocery Click &amp; Collect Customer.</td>
<td>Sainsbury's</td>
<td>Open</td>
</tr>
<tr>
<td>Exploratory and Interactive Visualisation of Data</td>
<td>Sainsbury's</td>
<td>Open</td>
</tr>
<tr>
<td>Dark Data</td>
<td>Sainsbury's</td>
<td>Open</td>
</tr>
<tr>
<td>Effect of store locations and other factors on geographic sales density.</td>
<td>Thomas Cook</td>
<td>Closed</td>
</tr>
<tr>
<td>The size and placing of promotional banners on a website.</td>
<td>SHOP DIRECT</td>
<td>Open</td>
</tr>
<tr>
<td>An investigation of what triggers customer activation of credit facilities.</td>
<td>SHOP DIRECT</td>
<td>Open</td>
</tr>
<tr>
<td>Generating a linguistic algorithm to interpret shoppers' comments and how this affects their shopping experiences and behaviour.</td>
<td>CACI</td>
<td>Closed</td>
</tr>
</tbody>
</table>
Data Licence Agreements...

• Data Licence Agreement template
  - Specific restrictions including publication conditions
  - Data set variables description – level of aggregation, time etc.
  - Use of partner name

• Data Sharing Agreement
Importing Data

- Encrypted hard drive
- Asset management
- Data Protection Registration (if required)
- Importing to Datastore
Asset Management

- Date
- CDRC Owner
- Designation – controlled, safeguarded, open
- Personal data or not
- Dataset reference
- Dataset name
- Partner name
- Partner contact
- Original/revised
- Version
- Metadata
- Physical location
- **Digital Object Identifier**
- DLA restrictions
- Institutional Data Protection registration

DOI:
- CDRC unique prefix
- Datacite Mandatory Properties
- Anonymous creators
Data Preservation

• Duration of CDRC, CDRC datasets will be curated, catalogued, maintained and preserved by the CDRC.

• After the CDRC those datasets created by the CDRC and that permission has been given will be archived at the UKDS UK Data Archive.

• Consumer organisation data – the raw data, if not representative of the population, do we want to be preserving this data? And who decides this?
Three-tier Data Access
Secure Facilities
Trusted Researchers
Governance
Safe results

www.cdrc.ac.uk