**User Centred Methods for Measuring the Value of Open Data**

Appendix 1: Details of the Workshops

# Preparation

All participants should be aware of the aims of the workshop and roughly what they will be asked to do before attending.

# Workshop 1

The times are approximately right for the UK for three participating organisations. We found we needed longer in India and would need longer if there were more organisations.

|  |  |  |  |
| --- | --- | --- | --- |
| Activity | Description | Results | Approximate Duration |
| Introductions | Introduce the participants to each other and recapitulate the structure and purpose of the workshops. | Participants are confident of their role and the significance of the project. | 10 mins |
| Examples of Open Data | Present some leading examples of how Open Data has been used. | Participants are aware of what is possible. | 10 mins |
| List the Biggest Problems | Brainstorm the most important problems facing this community. There are many techniques for doing this. We used Posts-Its. Stress this is not about data at this point of the process. | List of the most pressing problems facing the community. | 20 – 30 mins |
| Prioritise Problems (may not be necessary) | Try to prioritise the problems. In the UK we tried using a Difficulty, Frequency, Impact matrix but this was abandoned as all problems scored high on all counts and the list of problems was not long. As a consequence we did not do this in India. | Prioritised list of problems. | 10 mins |
| BREAK |  |  |  |
| Information requirements | Ask participants what information that they currently do not have would help them solve these problems. This might be addressed through a structured technique such as Post-Its but we found that it was quite sufficient to hold a facilitated discussion. | List of major information needs for each problem. | 30 mins |
| Wrap Up | Summarise what has been achieved and confirm it refects their opinion. Confirm date and objectives for the next workshop and invite communication between the workshops. | Confirmed list of information needs and engaged participants. | 10 mins |

# Workshop 2

The researchers used the period between the workshops to identify open datasets that came as close as possible to supplying the required information.

|  |  |  |  |
| --- | --- | --- | --- |
| Activity | Description | Results | Approximate Duration |
| Welcome back | Reminder of what happened at the previous workshop and what will happen in this one. Summary of what we have discovered. | Participants aware of possible datasets to use and how they address their problems. | 30 mins |
| Choice of datasets | Allow participants to study the datasets – this was done using paper copies of excerpts in the UK and by displaying the datasets on a projector in India. Then ask them which dataset is the most likely to be useful.  The discussion may reveal key attributes of the datasets which should be noted. | Choose dataset for study.  Possibly some attributes. | 20 minutes. |
| Using the dataset | Ask the participants to explain in as much detail as possible how they might use that dataset in practice. Prompt with questions about why, where, when, how often, with what tools, with what expected results.  Note any key attributes of the dataset. | Attributes of the dataset. | 30 mins. |
| If there is time repeat preceding step for another dataset. | | | |
| Value to participants | Discussion of how (if at all) open data can be used to help participants. This is a practical step mainly to reward participants for their contribution although it may also confirm or refine the list of attributes. | Actions to help participants. | 20 mins |
| Wrap-up | Thanks to participants and follow up arrangements e.g. opportunities to make further comment, learn about the result of the research. |  | 10 mins |

# **Notes**

* There were only 3 days between the workshops in India. Although we were able to identify 5 sets of applicable data, we discovered just before the workshop started that excel formats of some data that we had only found in PDF format were in fact available.
* Although the organisations were broadly similar, the participants in the two Indian workshops were different. This made little difference to the value of the second workshop as the members accepted the results of the first workshop.
* We planned originally to include a DIF analysis. However, the Winchester workshop experience suggested this was futile so abandoned it and did not include in Ahmedabad design
* It was initially decided to use a grounded theory approach in the second workshop and allow data issues to be derived organically without prompting. Given the short list of concerns that emerged in the first workshop, we decided to provide a list of issues commonly found in the literature (see below) in the second workshop. However, this approach did not generate any relevant new points and those attributes identified by the participants remained very much in line with those identified by the UK participants.
* We initially hoped to present the data in two ways, raw and in a ‘cleaned’, perhaps visualised version in order to explore whether the two formats would elicit different responses. However, due to the onerous nature of finding the data, the variety of existing formats and the fact that most of the data could not be found in a completely raw state, we simply provided each dataset in the format that it was located.

# **Data issues used to prompt participants in Workshop 2**

* Size of data file
* Aggregation/granularity
* Information about what it is
* Accuracy
* Completeness/precision – is everything you need there?
* Consistency
* Timeliness
* Credibility/provenance – did this come up at all in Winchester? That it was NHS data? Do the Ahmedabad people believe this?
* Freshness
* Currency
* Lag (Tau of data)
* Uniqueness (data is free of redundancy)

# 

# Problems and Information

These are direct transcripts of the charts summarising the key problems identified in the first workshop in each country and the information the participants felt would most help with those problems.

## UK

|  |  |
| --- | --- |
| Problems | Information that would help |
| Handling legal highs | Knowing the effects  Diagnosing |
| Problems with client finances  universal credit  benefit sanctions  EU nationals  Payment access due to… | Who is sanctioned in practice  why  how long  what conditions  is there an appeal  Who are they? |
| Evidence  persistent attempts  comparative monitoring  outcomes for tender submission  stock levels  accessing multiple services | Financial data on cost of public sector support  How many services to they use  Cost of non-intervention  Comparing  - progress ????  - proportion moving to planned |
| Moving On  landlords won’t accept benefits  lack of places  lack of supported housing  outside local area  getting work with no address | Where can funding come from ??  How to persuade a landlord  Rates of acceptance by landlord  Profile eg arrears history, risk eg police, health |
| Alcohol/mental health - safety | A way of categorising people so as to quickly know how to deal with them |
| Contractual side of delivering | No information identified |
| Volatility of regulations | No information identified |

### India

|  |  |
| --- | --- |
| Problems | Information that would help |
| Entitlements and Benefits  Access to basic services (utilities)  Children dropping out of school  Migrant labour entitlement | Service availability  Migrants need to register  Map of who has moved round the country  Sharing information about benefits with other districts and states  What health, education, programmes are they eligible for?  ICDS scheme (child nutrition)  Availability of school places  Where should new schools be built/old schools removed  What is current school enrolment, What is birth rate  What is net migration (this should tell you how many have dropped out, there is a project on this)  BPL (Below poverty line) Card data |
| Evidence Base  Sources  Level/country/state/urban/rural  Informal sector  Summary representation | Want information to be presented in easy to understand and use manner |
| Outcomes not outputs  Who receives monies/where spent  Subsidies  Social/religious classes  Where? geography | Where are the houses/number of houses, not just total spent  List of beneficiaries (this is identified before the housing is built in the application process)  What is socio-economic grouping of beneficiaries  Can the proposed socio-economic benefits actually be delivered in the chosen area? Did it happen?  Is possession clear in schemes where woman receives home in scheme but property is in man’s name? |
| Rights and vulnerabilities (Problems stopping people from moving into more permanent housing situations)  Abusive Employers  Unrecognised slum dwellers  Non-working youths without skills  Vulnerable living conditions  No money to move on | Have employers been inspected  Are they maintaining lists of employees (have to supply some of these to labour department and licensing agencies)  No. of youths working status, education levels  Sector gaps, job availability by type  Industry level data and household data  Skilled/semi-skilled  Lists of recognized slums  Planning notifications |

Datasets selected for Workshop 2

### UK

|  |  |  |
| --- | --- | --- |
| **Description** | **Source** | **Link** |
| Housing Stock - Empty and Council Owned Houses by District | Shelter Housing DataBank | http://england.shelter.org.uk/professional\_resources/housing\_databank |
| Housing Need – Number of people accepted as homeless by district authority and their status (from Shelter data portal) | Shelter Housing Databank | http://england.shelter.org.uk/professional\_resources/housing\_databank |
| Costs of different types of crime | Home Office | <https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/118042/IOM-phase2-costs-multipliers.pdf> |
| Costs of different types of medical intervention | NHS | https://www.gov.uk/government/publications/nhs-reference-costs-2013-to-2014 |

### India

|  |  |  |
| --- | --- | --- |
| **Description** | **Source** | **Link** |
| School numbers and enrolment | Gujarat State Economic Review | <http://gujecostat.gujarat.gov.in/wp-content/uploads/Publication/34%20-%20Socio%20Economic%20Review%20English.pdf> (section 11) |
| Total school age children 6-13, total never enrolled, total drop out rate | Gujarat State Economic Review | <http://gujecostat.gujarat.gov.in/wp-content/uploads/Publication/34%20-%20Socio%20Economic%20Review%20English.pdf> (section 14) |
| Total school age children 6-13, total never enrolled, total drop out rate | Gujarat State - survey | <http://ssa.nic.in/pabminutes-documents/NS.pdf> (Table A2) |
| Sample of non-working people by age, what work they are seeking, how long they have been out of work | National Survey | <http://mospi.nic.in/Mospi_New/site/inner.aspx?status=3&menu_id=31> |
| Population statistics:  Age group 0-6  Total workers  Total non-workers | National census | http://www.censusindia.gov.in/datagov/CDB\_PCA\_Census/PCA\_CDB\_2407\_F\_Census.xls |

Appendix 2: Results of metrics pilot

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Description | Data owner | Granularity (1) | | Granularity (2) | | Intelligibility | | Trustworthiness | | Discoverability | |
|  |  | Level | Confident | Level | Confident | Level | Confident | Level | Confident | Level | Confident |
| Affordable housing supply | UK DCLG / HCA | National | High | 1 - Agg only | High | 3 | Medium | 6 | Medium | 4 | High |
| Building Price and Cost Indices | UK BIS | Region | High | 1 - Agg only | High | 1 | Low | 1 | Medium | 4 | Medium |
| Central Government Property and Land including Welsh Ministers estate | UK Cabinet Office | Named building | High | 3 - Post code | High | 1 | Low | 1 | Medium | 4 | High |
| House building statistics | UK DCLG | Country | High | 1 - Agg only | High | 3 | Medium | 6 | High | 4 | High |
| Housing Energy Fact File 2012: energy use in homes | UK DECC | National | High | 1 - Agg only | High | 3 | Medium | 4 | Low | 4 | High |
| House Price Index background data | UK Land Registry | Region | High | 1 - Agg only | High | 5 | High | 5 | Medium | 4 | High |
| Housing benefit recoveries & fraud data | UK DWP | Primary Local Authority (county level) | High | 1 - Agg only | Medium | 3 | Medium | 6 | Medium | 4 | High |
| Shelter Databank | UK Shelter | City | High | 1 - Agg only | Medium | 3 | High | 2 | Medium | 2 | Medium |
| Statistical Data Release - Social Housing Stock by Local Authority and Private Registered Provider | UK Homes and Communities Agency | City | High | 2 - Individual | High | 4 | High | 6 | High | 4 | High |
| Tenure dwelling by accommodation type (of dwelling) | UK Office of National Statistics | City | High | 1 - Agg only | High | 1 | Low | 1 | Low | 4 | High |
| School numbers and enrollment (Ahmedabad) (section 11) | Directorate Of Economics And Statistics - Government Of Gujarat | District | Medium | 1 - Agg only | High | 1 | Low | 2 | Low | 2 | Medium |
| Employment figures (Ahmedabad) (section 14) | Directorate Of Economics And Statistics - Government Of Gujarat | State | Medium | 1 - Agg only | High | 1 | Low | 2 | Low | 2 | Medium |
| Children dropping out of school (table A2) | Gujurat Government, Social and Rural Research Institute | State | High | 1 - Agg only | High | 3 | High | 7 | Medium | 0 | Low |
| Sample of non-working people by age | Government of India, NSSO | City/Village | Low | 3 - Geographical identifiers | Low | 3 | High | 7 | High | 2 | Medium |
| Population statistics including employment status | Census Government of India | City/Village | High | 1 - Agg only | High | 1. | Low | 2 | Low | 2 | Medium |

Efficiency was high for all datasets for all metrics except for Trustworthiness were is was medium for all datasets and Discoverability where it was not possible to rate efficiency for individual datasets as significant effort went into discovering the characteristics of a portal which was then shared among all datasets using that portal.