|  |
| --- |
| **Innovation Lab**  **SUMMARY**  **PROTOTYPE** |
| Living healthy and independent lives in our local communities |

|  |  |
| --- | --- |
| **Date** | **Venue** |
| 27th April 2017 | ODI Leeds, Munro House, Leeds |

|  |
| --- |
| **Name of solution** |
| Data Standard |

|  |
| --- |
| **Technologist’s name** |
| Liam Bolton |

|  |
| --- |
| **Description of the solution**  *include group discussions, conclusions, URLs, links to GitHub etc.* |
| Councils and older citizens, for example, have different ways of describing and engaging with council services. This is reflected in city council datasets, which are often highly unstructured, e.g. dataset on neighbourhood network schemes in Leeds have different names for older people (see Fig 1).  So, how can we take this often highly unstructured, textual data and standardise it for councils and the public to use?  **CASE STUDY: BSI are standardising the description of public sector datasets about smart cities in metadata**  **ACTION PLAN:**   * Collect and council datasets. Focus on W. Yorkshire * Clean and format the data * Collaborate on a structured, standardised dataset to show best practice * Analyse the data using the Data Standard Toolkit, add to dataset * Topic Modelling (metadata description) * Entity Extraction * Machine Learning (e.g. clustering documents) * Publish the toolkit and the newly structured, standardised council data on Data Mill North   **ANALYSING COUNCIL DATA:**   * Clean the data and apply advanced data analytics (e.g. text mining and natural language processing) using open source statistical programming language R * This will help to tag datasets and create a structured, standardised dataset |

|  |
| --- |
| **Screenshots** |
| **Fig 1** |