Converting Bristol Well Aware Data to Open Referral with LGA Extensions

Paul Davidson, iStandUK

September 2019

[**Environment**](#_4p5c1edcm5l3) **1**

[**Input Data**](#_bc3e8krvhuhc) **1**

[Preparing](#_qi540j66ded) 1

[**Conversion**](#_njn3z531wvs5) **2**

[Creating some standard Taxonomy Items](#_h37xhvk8apdl) 2

[Bristol City Council Data Sources](#_c3eirtbobp8i) 2

[Converting Data](#_jj0xpu3s9v1g) 2

[The Organisations table](#_ku5owkgsy9eh) 2

[**Updates to LGA Application Profile of Open Referral**](#_klrjpa8nyocp) **5**

[**To Do**](#_bch997vitlmj) **6**

[**To Solve**](#_cue1ujjy12be) **6**

[Matching Activities to Organisations](#_xuqpvefaa9n1) 6

# Environment

Python script developed and run using PyCharm - see <https://www.jetbrains.com/pycharm/>

# Input Data

An excel spreadsheet is provided with tabs for

* Organisations
* Activities
* Events
* Groups
* Projects

These tabs have some common columns and some that are specific to the tab.

## Preparing

Each tab needs to be converted to a separate csv file, and converted to UTF8, so

* utf8Activities.csv
* utf8Events.csv
* utf8Groups.csv
* utf8Organisations.csv
* utf8Projects.csv

# Conversion

The Organisations tab is processed first to populate the ‘organization’ table.

The Organisations tab is then processed again, to create a ‘service’ entry for each organisation.

Each of the remaining tabs is then processed. These have some common columns that are also in the Organisations tab, and some that are specific to each tab.

## Creating some standard Taxonomy Items

### Bristol City Council Data Sources

These entries are created on the ‘taxonomy’ table. As each tab is processed, the service is tagged with one of these to mark which tab it was generated from.

|  |  |  |
| --- | --- | --- |
| **id** | **name** | **vocabulary** |
| bccsource:Organisation | Organisation | BCC Data Sources |
| bccsource:Project | Project | BCC Data Sources |
| bccsource:Event | Event | BCC Data Sources |
| bccsource:Group | Group | BCC Data Sources |
| bccsource:Activity | Activity | BCC Data Sources |

## Converting Data

### The Organisations table

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Input Table** | **Field** | **Output Table** | **Field** | **Treatment** |
| Organisations | ID | organization | id |  |
|  | Title |  | name |  |
|  | ImageURL |  | logo |  |
|  | Websiteurl |  | url |  |

### The Organisations table is read again to create corresponding entries on ‘service.csv.’

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Input Table** | **Field** | **Output Table** | **Field** | **Treatment** |
| Organisations |  | service | id | next id |
|  |  |  | status | ‘active’ - all entries on this table. |
|  | ID |  | organization\_id |  |
|  |  | service\_taxonomy | id | next id |
|  |  |  | service\_id | The id of the Service |
|  |  |  | taxonomy\_id | bccsource:Organisation |
|  | Content | service | description |  |
|  | Websiteurl | service | url |  |
|  | Telephone | phone | id | The next id for the ‘phone’ table |
|  |  |  | service\_id | The id of the Service |
|  |  |  | number |  |
|  | Email | service | email |  |
|  | PrimaryService | taxonomy | id | This taxonomy item may appear in many services - so it is only written once to the taxonomy table |
|  | Service |  |  | This is a more detailed categorisation of service type - treated the same as way ‘PrimaryService’ |
|  | AgeGroup |  |  | This is a controlled list to describe an Audience Group. Treated the same was as ‘Primary Service’ |
|  | UserGroup |  |  | This is a controlled list to describe an Audience Group. Treated the same was as ‘Primary Service’ |
|  | Address |  |  | This is held a one string delimited with either commas, or newline.  This is converted to address an array of address lines, and the Postcode is extracted and handled as a separate field.  A check is made to see if the postcode and 1st line of address already exists on the physical\_address file, otherwise the address ( and location ) is created. |
|  |  | location | id | The next id for location table |
|  |  | physical\_address | id | Same as location id |
|  |  |  | location\_id | The id from the location table |
|  |  |  | address\_1 | Address lines separated by comma. |
|  |  |  | city | I have put the remainder of the address after the 1st line here. |
|  |  |  | state\_province | !! I have left this blank - can we find an API to take a Post Code and return with City/Town/County. |
|  |  |  | postal\_code |  |
|  |  |  | country | “GB” |
|  | Postcodeareawheretheorganisationworks |  | postal\_code | If the Postcode cannot be extracted from the Address field, this field is used if it has content. |
|  | lat | location | latitude |  |
|  | lng |  | longitude |  |
|  | Accessibility | accessibility\_for\_disabilities | id | The next id |
|  |  |  | location\_id | The Location Id |
|  |  |  | accessibility | This is a list delimited by a pipe character. One row is written for each entry. |
|  |  | servce\_at\_location |  | This connects a service to a location/physical-address |
|  |  |  | id | The next id for this table |
|  |  |  | service\_id |  |
|  |  |  | location\_id |  |
|  | KeySearchTerms | taxonomy |  | This is a controlled list to of search terms. Treated the same was as ‘Primary Service’ |
|  | District |  |  | This is delimited by the pipe character |
|  |  | servce\_area | id | The next id |
|  |  |  | service\_id | The Service Id |
|  |  |  | service\_area | An element from the ‘District’ array. |
|  | AreasServed |  |  | This is a free text version of the same information in the ‘District’ field. |
|  | CostInformation | service | fees | This is free text |
|  | ConcessionsInformation |  | fees | Appended to the fees field. |

### The Activities Table

The Activities Table is processed in the same way that the Organisations table is processed in its 2nd pass above, - except that ...

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Input Table** | **Field** | **Output Table** | **Field** | **Treatment** |
| Activities |  | service | id | next id |
|  | OrganisationID / OrganisationLink |  | organisation\_id | The OrgId field should be a lookup to the input Organisations.csv table - this is present for about half of the entries.  The OrgLink field contains the name of the organisation. . Sometimes, this is present but the OrgId is missing.  Sometimes both of these fields are blank.  OrgId is Present: Use that for the organisation\_id  otherwise  OrgLink is present: Loop through all existing entries on the organisation.csv looking for a match - if found use that ‘id’ - if not found, create a new entry on organisation.csv  otherwise  same as above but using the Title field instead of the OrgLink field. |
|  | Cost | service | fees | As per CostInformation |
|  |  |  |  |  |
|  | StartDate | regular\_schedule | valid\_from |  |
|  | EndDate |  | valid\_to |  |
|  | Recurringoption |  | freq  interval | * weekly = WEEKLY * monthly = MONTHLY * Fortnightly = WEEKLY and interval = 2 |
|  | StartDate |  | byday | Get the day of the week for WEEKLY frequency by finding the day of the week of the StartDate and converting that to one of MO, TU, WE etc |
|  |  |  |  |  |
|  | Whichoccurrenceofthedayofthemonth |  | freq | If this is present, and the freq is WEEKLY, change it to MONTHLY. |
|  |  |  | byday | ‘first’ = 1  ‘second’ = 2  ‘third’=3  ‘fourth’=4  ‘last’=-1  Prefix byday with this number e.g.  -1FR means the last Friday of each Month. |
|  | StartDate |  | bymonthday | Where the frequency is MONTHLY, use the Start Date to find the day number in the month. |
|  | Whichoccurrenceofthedayofthemonth |  | interval | Where the frequency is MONTHLY, use this as the interval.  E.g. 2 means every other month. |

# Updates to LGA Application Profile of Open Referral

|  |  |  |
| --- | --- | --- |
| Service Area / Coverage | The OR service\_area table defines the ‘coverage’ for a single ‘service’. It is possible for a single service to have more than one ‘service\_area’ entry.  OR has no persistent identifiers for coverage.  My original proposal has a new LGA extension table for ‘Coverage’. However, we can now make that an extension of ‘service\_area’.  Fields are a combination of existing OR fields and the fields in the Coverage extension table. |  |
| Accessibility | The documentation defines permitted values. The Bristol data has different values.  This would be better as a taxonomy. |  |
| fees | The documentation says that this is an LGA extension. In fact, this is an ‘open referral’ field, for free text. We have the extra cost\_option table where costs are recorded in a structured way, and related to circumstance. |  |
| email |  | Should email be allowed on organization |

# To Do

|  |  |
| --- | --- |
| Update documentation for | Contentdetails  Is implemented in SetupService |
|  | Document ‘Cost’ in tables other than Organisations |
| Activities/Venue | Should be location name |
| Activities/address\_2 | Check how Activities addressing is dones |
| City / State | Improve how city and state\_province are handled by looking for ‘Bristol’ in the address. |
| freq | Check Recurringoption - daily |

# To Solve

# To Fix

Make sure that phone always uses contact.

Add ‘en’ and language for contact.