# Administrative Data on Civil Society Organisations in Scotland

File: Civil\_society\_organisations\_in\_Scotland.dta/civil\_society\_organisations\_in\_Scotland.csv

# Contents

This file consists of data on civil society organisations operating in Scotland. It originates from two sources, the Scottish Charity Register, which is a public Register of all charities in Scotland, and the Milo database, which is managed by the Scottish Council for Voluntary Organisations on behalf of the network of 32 Third Sector Interfaces in Scotland. The files have been linked to one another, and to area classifications such as the Rural Urban Indicator, and organisations have been assigned to the International Classification of Non-Profit Organisations (ICNPO). The data sources, processes for merging and linking the files, and assigning ICNPO codes, are described here. The syntax files for use in Stata 12 or later, which were used to create the merged and coded file, and could be reused to updated them, or to code other lists of organisations, are also supplied: here (see readme.txt for how to use them).

# Background

Measuring civil society is a perennial problem for nonprofit researchers. Even setting aside definitional issues of the boundaries of the sector, collecting good data on the number, activity and spread of organisations can be challenging. Extensive use has been made of administrative lists of organisational data, usually drawn from charity regulator records. This produces good quality data on organisations that have registered as charities – a significant component of civil society organisations – but does not necessarily capture smaller, more informal, unregistered organisations. Furthermore, it introduces a problem of ‘head office’ bias, where analysis of the registration records identify charity head offices but do not capture the geographic spread of civil society activity.

All charities operating in Scotland are required to register with the Office of the Scottish Charity Regulator (OSC, who maintain and release the Scottish Charity Register. This ensures good coverage of even small organisations operating as charities, but does not record third sector organisations who do not operate as charities and does not capture the spread of activity. Each of the 32 Scottish local authority areas has a nominated Third Sector Interface (TSI) (usually a third sector organisation or partnership). TSIs provide a single point of access for support and advice for the third sector within the area. Each TSI collects considerable local data about CSOs, services, and volunteers. This data is brought together in Milo, a national online database, knowledge management and reporting platform for Scotland managed by the Scottish Council for Voluntary Organisations (SCVO). Milo currently contains around 40,000 organisational records, covering a range of data from basic organisation profile information to volunteering opportunities and financial information. Although the data is primarily gathered for internal administrative purposes, it is also increasingly used to feed public-facing civil society website listings: thus the quality of the data is increasingly important, and it begins to play a role in representing the sector to the public, and policymakers, in Scotland.

As part of the ESRC-funded Scottish Civil Society Data Partnership project we have gathered, cleaned and coded the Milo data, for use by researchers in measuring and understanding the scope and distribution of civil society organisations in Scotland.

# Data Sources

## Milo

Milo is the Scotland-wide database of voluntary action. Staff at Scotland's 32 third sector interfaces (TSIs) use Milo to record up to date intelligence on organisations, contacts, volunteers, and volunteering opportunities. Milo currently contains records for around 40,000 third sector organisations across Scotland, and around 10,000 live volunteering opportunities.

General public data on voluntary organisations is published to the Get Involved website: [www.getinvolved.org.uk](http://www.getinvolved.org.uk). Volunteering opportunities are input and managed by local TSIs in conjunction with local volunteer-involving organisations. Live volunteering opportunities are published to the Volunteer Scotland website: [www.volunteerscotland.net](http://www.volunteerscotland.net)

The data on Milo is neither recorded for research purposes nor as a public record, but as an administrative tool for the TSIs. Therefore the completeness of an organisation’s record will vary according to the purpose for which it was recorded – if it was to advertise a volunteering opportunity, the record would tend to be complete, but if it was to request a phone call for some advice there may be more limited information entered. Moreover, the way in which Milo is used does vary between TSIs, so that some fields are used is some areas and not others. However, according to a recent user survey with 173 respondents from all but one of the TSIs, over 70% of users access Milo several times a week, and over 96% access it at least once a month. Milo is used most for “recording interactions with organisations or volunteers” (67% of users), “entering information about my work” (65%) and “managing contact records” (52%). Over 72% of users feel their TSI is generally using the system well, or at an advanced level.

Only fields that were populated by the majority of TSIs are included in this file: potential impact of missing data is described further in the documentation.

## Office of the Scottish Charity Register (OSCR)

The Scottish Charity Register is a public Register of all charities in Scotland, with no income qualification. It contains individual entries for each charity including a charity's name, contact address, charitable purposes and certain financial information. The data is available for public download, the file used here was downloaded on 19 April 2017, containing 24,078 charities, 22,605 of which were active.

Charities tend to take care over their registration, so the data entry is generally complete and high quality. Nonetheless the responses reflect the opinions of the person recording the registration, so that, even variables with precoded categorical responses, such as Purposes, Activities and Beneficiaries, have a wide variety of responses for organisations carrying out the same activities, for example Scout troops. In addition, the extent to which charities record their relationship with larger bodies varies: over 90% of charities with Church of Scotland in their name also record their relationship with their parent organisations, whereas fewer than 40% of Scout troops record their affiliation with the larger Scouting movement.

OSCR data is covered by Crown Copyright and database right 2017. It contains information from the Scottish Charity Register supplied by the Office of the Scottish Charity Regulator and licensed under the Open Government Licence v.2.0.

# Data processing

## Cleaning Milo

Cleaning the Milo database entailed checking the records for duplicates, for organisations that no longer exist or have moved premises, and for incomplete fields/non-standard entries. For this purpose a version of the organisation name was created which was standardised to remove all punctuation, symbols and accents (which were present in some organisations with Scots Gaelic names, in particular). A simple search for duplicates according to organisation name and postcode identified only a very modest number, as did an external data cleaning service (0.5% and 0.6% of voluntary organisations, respectively). The external cleaning service (Experian Intact) also identified 1.1% of organisations that were no longer at the address recorded on Milo (according to information supplied by the Royal Mail). In both cases, while it was recommended that the records should be checked and updated/removed where necessary, these results are not considered to be problematic in terms of the overall integrity of the data. However, this will not include the 4,264 voluntary organisations with no address details (see below).

## Missing address data

Almost 4,300 voluntary organisation records have no address or postcode recorded, meaning that they cannot contribute to a geographic analysis of organisations. In order to understand whether this would introduce bias to the dataset, we analysed which TSI had created the record, as usually TSIs are recording organisations within their own local authority district. Overall, 10.5% of records for voluntary organisations had no address details, with a range of values from 1% to 32% for different TSIs, so the utility of the geographic analysis will be greater for some local authorities than others.

To understand the impact of missing address details Scotland-wide, we categorised the TSIs according to their deprivation and rurality. In the nine most deprived local authority areas, where over 20% of their population are living in areas in the most deprived quintile for Scotland, we found that 10.2% of voluntary organisations had no address details, very similar to the overall average. On the other hand, the 10 most rural council areas, where over 40% of the population live in rural areas, 12.9% of records had no addresses, slightly higher than average. So Scotland-wide the Milo data will be broadly representative of deprived and wealthier areas, but will slightly under-represent rural areas.

## Matching to OSCR

Just over 16,900 Milo records had a charity number recorded. Having cleaned this to ensure correct formatting, 16,248 could be matched to the OSCR register. However, there are a large number of registered charities in Scotland which are not represented on Milo, and further attempts have been made to identify whether there were registered charities present on Milo which had not been recorded as such, including identifying organisations with an identical name and postcode, those with an identical name only, and finally probabilistic matching using a user-written Stata package, reclink2 (Wasi & Flaaen, 2015). This compares combinations of fields (name, address, postcode, and postcode sector) to generate the probability that two records are in fact the same entity. Automated matching was found to be difficult, given the great similarity in names between distinct organisations (eg the 103rd A, B, C and D City of Edinburgh Brownie Units are all separately registered charities). Different combinations of fields were examined manually for their accuracy, and only those criteria that achieved at least 90% accuracy were taken as correct matches.

Following the work described above, 16.585 voluntary organisation records on Milo have been matched to the OSCR register, or 42%. Some registered charities are listed multiple times on Milo, legitimately in that they have multiple sites (for example, the Red Cross has their head office, charity shops and centres offering social care of various kinds, from first aid training to mobility aids). This is an important contribution of Milo as a research resource, and means that the 16,585 records on Milo represent 12,440 records on OSCR, or 52% of the register. While we suspect that there are further records to be successfully matched, it is also clear that at least a large minority of each database does not overlap with the other. The table below shows the method and extent to which Milo and OSCR records were reconciled.

Table 1: Method of linking Milo & OSCR records

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Method | Milo CSO Records | % Milo | OSCR Records | % OSCR |
| Registered number | 16,212 | 40.9% | 12,070 | 50.0% |
| Name & postcode | 39 | 0.1% | 39 | 0.2% |
| Name only | 112 | 0.3% | 111 | 0.5% |
| Probabilistic | 222 | 0.6% | 221 | 0.9% |
| Not matched | 23,044 | 58.2% | 11,637 | 48.2% |
|  |  |  |  |  |
| Total | 39,629 | 100.0% | 24,078 | 100.0% |

## ICNPO Coding

In order to usefully categorise the organisations for analytical purposes, we developed syntax to search for key terms in order to allocate organisations to ICNPO codes. As well as being the standard classification for comparing CSOs internationally, the project was able to build on the work of Mohan & Barnard (2013), which used this method to generate and compare a classification of Scottish Charities registered on OSCR to the English Charity Commission register (which had been coded using this method by NCVO (2010)). As in these previous exercises, it would be noted that this coding has not been generated by manual examination of each record: much care has been taken to check and recheck the allocation of organisations to the coding frame, but it is inevitable that some organisations will be miscoded and we are keen to have feedback on any coding issues.

The ICNPO coding frame consists of a number of top level codes, to which Mohan and Barnard, and NCVO, had added bespoke sub-classifications. In consultation with SCVO we created some further sub-codes which made the analysis more discriminatory for the Scottish sector, which are outlined in the table on page 9. Codes in bold are those specified within the official classification, those with normal weight are not but can be aggregated to the official classification for comparison purposes.

Reasons for the additional codes are varied. In some cases, for example 4110-4150, they provide discrimination where otherwise a large proportion of records would be assigned to one code. Others were created for organisations that perform functions in more than one top level group, so they could perhaps belong elsewhere. An example is 2410 Medical Research, which Mohan and Barnard had separated out as many medical research charities primarily sponsor work by other organisations, rather than carrying it out themselves, thus perhaps belonging to Grantmaking organisations rather than research. Our experience of coding such organisations is that there are many which both fund research into particular conditions, and offer social care to those diagnosed with them, which therefore straddle Medical research and Social care. These were coded on the balance of prominence given to these activities, where possible, so that Cancer Research UK is coded to 2410 Medical Research, and Click Sargent to 4140 Services for people with disabilities, but it should be acknowledged that there is overlap in their activities.

A number of methods were used to allocate organisations to the coding frame, which are listed below in order of the accuracy of the results given:

Table 2: Method of deriving ICNPO Codes

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Method | Milo Records |  | OSCR Records |  |
| 1 Organisation Name | 28,111 | 71% | 16,657 | 69.2% |
| 2 Mohan & Barnard/ NCVO | 4,292 | 11% | 5.663 | 0.0% |
| 3 Generic Keyword | 2,698 | 7% | 776 | 3.2% |
| 4 SCVO | 32 | 0% | 51 | 0.2% |
| 5 Client Group | 834 | 2% | 58 | 0.2% |
| 6 Aims | 1,305 | 3.3% | 852 | 3.5% |
| Not coded (allocated to 12100) | 2,704 | 5.9% | 38 | 0.2% |
| Total | 39,625 |  | 24,078 |  |

1. Organisation name: specific combinations of keywords were searched for, which had been identified as accurately allocating an organisation to a category
2. Mohan & Barnard/NCVO: the coding generated by this previous exercise was used, if the organisation had been identified as a registered charity – this existing coding was also an important means of checking the quality of the new coding we were creating.
3. Generic keyword: keywords were searched for which were likely to allocate them to the correct code, for example, “community”, “castle”, but where this was recognised to be less accurate.
4. SCVO: Milo records were matched to a separate, internal database of a sample of registered charities which SCVO compiles, to which they have allocated ICNPO codes
5. Client Group: a code entered when volunteering opportunities are entered onto Milo, which we mapped onto ICNPO codes.
6. Aims: the original search for specific keywords used in the organisation name was then run on the stated aims of the organisation, where these are recorded: the text field is longer and thus more likely to contain conflicting information, so is less accurate than the organisation name for coding purposes.
7. Finally, 6% of organisations could not be coded by this process and were added to the “not otherwise classified” code.

The coding was run only for all 38,000 records on Milo: c3,000 records on Milo represent private companies, statutory organisations and others. Where the aims of the organisation were found to match an ICNPO code they were allocated to it, otherwise they were given the code 99999 to denote that they were outside the scope of ICNPO coding.

Table 3: ICNPO coding of OSCR records, and Milo records representing voluntary organisations

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **ICNPO** | **OSCR database** | | **Milo** | | **All voluntary orgs** | |
| **1100. Culture and Arts** | 2,507 | 10.4% | 3,741 | 9.4% | 4,812 | 12.1% |
| **1200. Sports** | 690 | 2.9% | 3,337 | 8.4% | 3,640 | 9.2% |
| **1300. Other Rec & Soc Clubs** | 571 | 2.4% | 1,610 | 4.1% | 1,922 | 4.9% |
| **2100. Primary and Secondary Education** | 161 | 0.7% | 380 | 1.0% | 465 | 1.2% |
| 2110. Parent-Teacher Associations | 233 | 1.0% | 252 | 0.6% | 389 | 1.0% |
| 2120. Educational Foundations | 133 | 0.6% | 56 | 0.1% | 145 | 0.4% |
| **2200. Higher Education** | 182 | 0.8% | 239 | 0.6% | 316 | 0.8% |
| **2300. Other Education** | 265 | 1.1% | 364 | 0.9% | 493 | 1.2% |
| **2400. Research** | 179 | 0.7% | 132 | 0.3% | 234 | 0.6% |
| 2410. Medical Research | 165 | 0.7% | 310 | 0.8% | 383 | 1.0% |
| **3100. Hospitals and Rehabilitation** | 184 | 0.8% | 355 | 0.9% | 432 | 1.1% |
| **3200. Nursing Homes** | 84 | 0.3% | 164 | 0.4% | 189 | 0.5% |
| **3300. Mental Health** | 180 | 0.7% | 470 | 1.2% | 510 | 1.3% |
| **3400. Other Health Services** | 411 | 1.7% | 850 | 2.1% | 1,015 | 2.6% |
| 3410. Addictions support | 57 | 0.2% | 227 | 0.6% | 244 | 0.6% |
| **4100. Social services (general)** | 1,061 | 4.4% | 1,730 | 4.4% | 2,043 | 5.2% |
| **4110. Services for Children and families** | 700 | 2.9% | 1,511 | 3.8% | 1,814 | 4.6% |
| **4120. Pre-school daycare** | 183 | 0.8% | 360 | 0.9% | 421 | 1.1% |
| **4130. Services for young people** | 614 | 2.5% | 1,507 | 3.8% | 1,736 | 4.4% |
| 4131. Scouts, Guides etc | 2,919 | 12.1% | 1,807 | 4.6% | 3,873 | 9.8% |
| 4140. Services for people with disabilities | 366 | 1.5% | 1,063 | 2.7% | 1,157 | 2.9% |
| 4150. Services for elderly people | 212 | 0.9% | 1035 | 2.6% | 1,083 | 2.7% |
| 4160. Services for women | 55 | 0.2% | 132 | 0.3% | 142 | 0.4% |
| 4170. Carers Organisations | 87 | 0.4% | 264 | 0.7% | 282 | 0.7% |
| 4180. Refugee services | 7 | 0.0% | 26 | 0.1% | 28 | 0.1% |
| **4200. Emergency and Relief** | 130 | 0.5% | 197 | 0.5% | 245 | 0.6% |
| 4220. Homelessness support | 29 | 0.1% | 123 | 0.3% | 132 | 0.3% |
| **4300. Income Support and Maintenance** | 684 | 2.8% | 541 | 1.4% | 995 | 2.5% |
| **5100. Environment** | 623 | 2.6% | 1,242 | 3.1% | 1,477 | 3.7% |
| **5200. Animal Protection** | 308 | 1.3% | 494 | 1.2% | 622 | 1.6% |
| **6100. Social and Community Development** | 1,179 | 4.9% | 2,359 | 6.0% | 2,782 | 7.0% |
| 6110. Village Halls | 748 | 3.1% | 1,066 | 2.7% | 1,284 | 3.2% |
| 6120. Community Transport | 54 | 0.2% | 103 | 0.3% | 118 | 0.3% |
| 6140. Credit Unions | 3 | 0.0% | 111 | 0.3% | 112 | 0.3% |
| 6150. Social enterprise | 10 | 0.0% | 51 | 0.1% | 54 | 0.1% |
| **6200. Housing** | 259 | 1.1% | 609 | 1.5% | 703 | 1.8% |
| 6210. Housing Associations | 125 | 0.5% | 268 | 0.7% | 275 | 0.7% |
| **6300. Employment and Training** | 258 | 1.1% | 508 | 1.3% | 621 | 1.6% |
| **7100. Civic and Advocacy Orgs** | 369 | 1.5% | 739 | 1.9% | 895 | 2.3% |
| 7110. Ethnic minority organisations | 24 | 0.1% | 39 | 0.1% | 48 | 0.1% |
| 7120. Sexuality/LGBT | 11 | 0.0% | 31 | 0.1% | 36 | 0.1% |
| **7200. Law and Legal Services** | 149 | 0.6% | 367 | 0.9% | 447 | 1.1% |
| 7210. Advice Services | 68 | 0.3% | 120 | 0.3% | 127 | 0.3% |
| 7230. Offenders/ex-offenders support | 0 | 0.0% | 7 | 0.0% | 7 | 0.0% |
| **7300. Political Organisations** | 2 | 0.0% | 36 | 0.1% | 38 | 0.1% |
| **8100. Grantmaking Foundations** | 1,000 | 4.2% | 508 | 1.3% | 1,208 | 3.0% |
| **8200. Volunt Promotion, Fund-raising** | 156 | 0.6% | 390 | 1.0% | 426 | 1.1% |
| **9100. International Activities** | 1019 | 4.2% | 1,009 | 2.5% | 1,545 | 3.9% |
| **10100. Religious Congs & Assocs** | 3,761 | 15.6% | 2,988 | 7.5% | 4,954 | 12.5% |
| **11100. Business Associations** | 207 | 0.9% | 417 | 1.1% | 508 | 1.3% |
| **11200. Professional Associations** | 52 | 0.2% | 60 | 0.2% | 77 | 0.2% |
| **11300. Trade Unions** | 12 | 0.0% | 32 | 0.1% | 42 | 0.1% |
| **12100. Not elsewhere classified** | 46 | 0.2% | 2,353 | 5.9% | 2,382 | 6.0% |
| **12200. Gen or Mult Charitable Purps** | 572 | 2.4% | 404 | 1.0% | 813 | 2.1% |
| 12300. Charity shops | 40 | 0.2% | 464 | 1.2% | 470 | 1.2% |
| 12400. Cafes | 9 | 0.0% | 67 | 0.2% | 71 | 0.2% |
| Total | 24079 |  | 39625 |  | 39,625 |  |

# Data File

Following is a definition for each field in the analysis. The source defines whether the field is from the Milo database, OSCR, or has been derived from other fields (for example the ICNPO coding, or by linking to area-level data using the postcode). In some cases the Milo fields have been edited a little, for example the orgtype has been cross-checked for consistency with whether the organisation was matched with a registered charity, non-standard entries were corrected, and additional records were allocated to the Statutory code based on the organisation name.

Note that many of the fields will have missing values: records which are not on Milo will have all Milo fields empty, and records which are not on OSCR will have all those fields empty. Only the “derived” fields have values in all fields, even if to denote a missing value (for example council area will be missing if postcode is missing). Note that the postcode given here is only that found on the OSCR data, which is already published for all registered charities on their website. Organisations found only on Milo will have datazone as the most detailed geographic identifier.

|  |  |  |  |
| --- | --- | --- | --- |
| Variable name | Format | Source | Description |
| combname | str126 | Derived | Organisation name from Milo, if present, if not, Charity Name from OSCR |
| Lists | float | Derived | Whether organisation is on Milo, OSCR, or both |
| match\_how | str11 | Derived | How Milo was linked to OSCR: see table 1 for coding |
| icnpo\_comb | long | Derived | ICNPO code assigned: see table 3 for coding |
| icnpo\_comb\_how | str17 | Derived | How ICNPO Code was derived: see table 2 for coding |
| voluntary | float | Derived | From orgtype - statutory or private = 0, others = 1 |
| milo\_orgid | str15 | Milo | Organisation ID |
| milo\_par\_orgid | str15 | Milo | Parent Organisation ID on Milo |
| milo\_branchoflargerorg | byte | Milo | Branch of larger organisation: 1=Yes 0=No |
| milo\_largerorg | str72 | Milo | Parent Organisation Name on Milo |
| milo\_citycounty | str61 | Milo | City/County |
| milo\_aims | strL | Milo | Aims And Objectives |
| milo\_ownertsi | str21 | Milo | The TSI that is responsible for the record on Milo |
| milo\_mainactivities | str1190 | Milo | Main Activities |
| milo\_par\_aims | strL | Milo | Parent Aims And Objectives |
| milo\_par\_mainacts | str1082 | Milo | Parent Main Activities |
| milo\_par\_regchar | str9 | Milo | Parent Registered Charity Number |
| milo\_par\_oname | str72 | Milo | Parent Organisation Name |
| milo\_orgtypecl | str25 | Derived | Organisation Type adapted from Milo: OSCR records assigned to "Registered Charity" |
| oscr\_charitynumber | str26 | OSCR | Charity Number |
| oscr\_charityname | str126 | OSCR | Charity Name |
| oscr\_registerdate | str16 | OSCR | Registered Date |
| oscr\_knownas | str139 | OSCR | Known As |
| oscr\_charitystatus | str13 | OSCR | Charity Status |
| oscr\_constitutionalform | str79 | OSCR | Constitutional Form |
| oscr\_geographicalspread | str74 | OSCR | Geographical Spread |
| oscr\_mainoperatinglocation | str19 | OSCR | Main Operating Location |
| oscr\_purposes | str1008 | OSCR | Purposes |
| oscr\_beneficiaries | str252 | OSCR | Beneficiaries |
| oscr\_activities | str195 | OSCR | Activities |
| oscr\_objectives | strL | OSCR | Objectives |
| oscr\_address | str240 | OSCR | Principal Office/Trustees Address |
| oscr\_website | str113 | OSCR | Website |
| oscr\_lastyearincome | long | OSCR | Most recent year income |
| oscr\_lastyearexpenditure | long | OSCR | Most recent year expenditure |
| oscr\_yearend | str10 | OSCR | Year End |
| oscr\_parentcharityname | str86 | OSCR | Parent charity name |
| oscr\_parentcharitynumber | str28 | OSCR | Parent charity number |
| oscr\_parentcountry | str16 | OSCR | Parent charity country of registration |
| oscr\_designatatedreligiousbody | str3 | OSCR | Designated religious body |
| oscr\_postcode | str11 | OSCR | Postcode |
| geog\_LAD15CD | str9 | Derived | Local Authority District code, linked from postcode |
| geog\_LAD15NM | str25 | Derived | Local Authority District name, linked from postcode |
| geog\_osward | str9 | Derived | 2013 Electoral Ward code, linked from postcode |
| geog\_oshlthau | str9 | Derived | 2012 Strategic Health Authority code, linked from postcode |
| geog\_GOR10CD | str9 | Derived | Government Office Region code linked from postcode |
| geog\_GOR10NM | str24 | Derived | Government Office Region name, linked from postcode |
| geog\_teclec | str9 | Derived | Local Learning and Skills Council (LLSC)/ Dept. of Children, Education, Lifelong Learning and Skills (DCELLS)/ Enterprise Region (ER) |
| geog\_lsoa11 | str9 | Derived | 2011 Census Lower Layer Super Output Area (LSOA)/ Data Zone (DZ)/ SOA linked from postcode |
| geog\_ru11ind | str2 | Derived | Rural/Urban indicator 2011, 8-categories, linked from postcode |
| geog\_simd2012\_sc\_decile | str1 | Derived | Scottish Index of Multiple Deprivation decile linked from postcode |
| geog\_simd2012\_sc\_quintile | str2 | Derived | Scottish Index of Multiple Deprivation quintile, linked from postcode |

# Data construction

The Stata 12 do files used to merge and code the MILO and OSCR data are also included here, for the purposes of updating this dataset or developing ICNPO coding for separate datasets. The files and their actions are as follows (and should be run in this order:

|  |  |
| --- | --- |
| **Filename** | **Action** |
| milo share step 1 clean.do | Imports Milo and OSCR files to Stata, attempts to merge them on charity number, then cleans unmatched charity numbers and merges again. Adds ICNPO coding of from NCVO and variables from parent organisations if missing |
| milo share step 2 merge geography.do | Imports postcode directory and adds area-level identifiers using lookups based on either Milo or OSCR postcode. Also adds deprivation data from ISD. |
| milo share step 3 finalised matching.do | Uses organisation/charity name and postcode to match records on Milo and OSRC not already matched through charity number. |
| code icnpo comb 0 prep.do | Defines ICNPO codes and labels, and creates cleaned, lower-case, alpha-numeric organisation name field |
| code icnpo comb 1 orgname.do | Assigns ICNPO code based on words or phrases contained in the cleaned organisation name |
| code icnpo comb 2 ncvo.do | Assigns ICNPO code using the NCVO coding, if available. |
| code icnpo comb 3 generic.do | Assigns probable ICNPO code based on generic keywords |
| code icnpo comb 4 scvo.do | Assigns ICNPO code from SCVO Panel database |
| code icnpo comb 5 clients.do | Assigns ICNPO code using client field from volunteering opportunities recorded on MILO |
| code icnpo comb 6 aims.do | Assigns ICNPO code searching for the name words or phrases as used in the organisation name, but this time in the Aims field |
| code icnpo comb 7 orgtype.do | Checks and assigns organisation type to ensure that council services etc are correctly coded |
| create file for deposit.do | Selects and renames fields, saves dataset as a new .dta file and exports data as a .csv |

# References

Mohan, J. (2012) Entering the lists: what can we learn about the voluntary sector in England from listings produced by local infrastructure bodies? Voluntary Sector Review Vol 3 No 2

Mohan, J., & Barnard, S. (2013). Comparisons between the characteristics of charities in Scotland and those of England and Wales. <http://www.cgap.org.uk/uploads/Mohan%20Barnard%20Characteristics%20of%20charities%20in%20Scotland%20and%20those%20of%20England%20and%20Wales.pdf>

NCVO (2010) Classification of charities in England and Wales. Available online at https://data.ncvo.org.uk/datastore/datasets/dataset-4-icnpo-classification-of-charities/

Scottish Council for Voluntary Organisations (2014) State of the Sector Report. SCVO, Edinburgh

UN (United Nations) (2003) Handbook on nonprofit institutions in the system of national accounts, Studies in Methods, Series F 91, New York, NY: United Nations Department of Economic and Social Affairs, Statistics Division, <http://unstats.un.org/unsd/publication/seriesf/seriesf_91e.pdf>

Wasi, N., & Flaaen, A. (2015). Record linkage using Stata: Preprocessing, linking, and reviewing utilities. *Stata Journal*, *15*(3), 672-697.