

# Database of third sector organisations: schema and process

Fiona Couper Kenney, Alan Duggan

July 24, 2025

## 1 Introduction

This document is a guide to the UK Third Sector and Civil Society Database, a ‘spine’ of all registered organisations comprising the third sector and civil society in the United Kingdom, along with linkages between organisations which appear in more than one register. The dataset is available at the project website<sup>1</sup>.

Ten data sources are combined in order to build the organisational spine, listed in Table 1, and further detailed in appendices. This document briefly introduces the processes of data collection, pre-processing of data per source, and combining all sources to create one master list. The spine files are first introduced, followed by a description of the process of matching organisations between sources. Preprocessing of each source is detailed in the appendix. The project source code is available on GitHub<sup>2</sup>.

## 2 Building the organisational spine

Ten data sources are combined to create the organisational spine. All data in the spine is open and publicly available. The sources and their online location are listed in Table 1.

Data Source	URL
Charity Commission, England and Wales	<a href="http://www.gov.uk/government/organisations/charity-commission">www.gov.uk/government/organisations/charity-commission</a>
Scottish Charity Register	<a href="http://www.oscr.org.uk/">www.oscr.org.uk/</a>
Charity Commission, Northern Ireland	<a href="http://www.charitycommissionni.org.uk/">www.charitycommissionni.org.uk/</a>
Companies House	<a href="http://www.gov.uk/government/organisations/companies-house">www.gov.uk/government/organisations/companies-house</a>
Co-operatives	<a href="http://www.uk.coop/">www.uk.coop/</a>
Mutuals Public Register	<a href="http://mutuals.fca.org.uk">mutuals.fca.org.uk</a>
Scottish Housing Register	<a href="http://www.housingregulator.gov.scot/">www.housingregulator.gov.scot/</a>
Social Housing England	<a href="http://www.gov.uk/government/publications/registered-providers-of-social-housing">www.gov.uk/government/publications/registered-providers-of-social-housing</a>
Care Inspectorate Scotland	<a href="http://www.careinspectorate.com">www.careinspectorate.com</a>
Care Quality Commission	<a href="http://www.cqc.org.uk">www.cqc.org.uk</a>

Table 1: Data sources for the organisational spine

The organisational spine consists of three csv files:

- TSCS\_spine.spine.csv contains a row for each organisation considered to make up the UK’s third sector, with name, address and date details. The data dictionary is found in Table 2.

---

<sup>1</sup><https://uk-third-sector-database.github.io/>

<sup>2</sup><https://github.com/uk-third-sector-database/tso-database-builder>

- TSCS\_spine.supplementary.csv contains any name, address and date information associated with an organisation which is additional to what is in the spine, such as 'Also Known As' names, previous addresses. See Table 3 for data dictionary
- TSCS\_spine.matches.csv lists the organisations which are linked or matched according to our searches (see 2.1), along with the type of match. Data dictionary in Table 4

Fieldname	Notes
uid	UID created for each source, e.g. GB-CHC-1234
organisationname	Name as primary in source
normalisedname	Name normalised to capitalise, remove extra spaces and punctuation - to aid matching
fulladdress	Full address as found in source with highest precedence across any matched organisations
city	City matching full address as defined above
postcode	Postcode matching full address as defined above
registerdate	Earliest registration date found in records for this organisation across all sources
removeddate	Latest dissolution date found in records for this organisation across all sources
source_register	Source of the data, e.g. Charity Commission for England and Wales
is_cic	Boolean indicating if organisation is a Community Interest Company

Table 2: Public spine csv fields

Fieldname	Notes
uid	UID created for each source using source and id from original source
organisationname	Alternative name
normalisedname	(Alt name normalised)
fulladdress	Alternative address
city	Alternative address
postcode	Alternative address
registerdate	Alternative registration date
removeddate	Alternative removed / dissolved date
source_register	Source of the data, e.g. 'Companies House', 'Charity Commission for England and Wales'

Table 3: Supplementary csv fields

Fieldname	Notes
uid	UID created for each source, matches Table 2 - if match is unambiguous; blank otherwise
orgA_id_in_source	ID in original source, for matched organisation A
orgA_source	original source of organisation A
orgA_uid	UID created for organisation A
orgB_id_in_source	ID in original source, for matched organisation B
orgB_source	original source of organisation B
orgB_uid	UID created for organisation B
match_type	Type of match (e.g. by name, by companyid)

Table 4: Matched organisations fields

## 2.1 What constitutes a match?

The main dataset has one row per organisation. If an organisation appears in two datasets, the linked organisation is noted by a row in the matches lookup table. We prioritise those organisations on the charity register, and add new organisations to the main dataset from other sources only if they do not match one already present in the set.

The following routes lead to a match:

- Matching by Company ID:

The charity regulator in England and Wales, CCEW, collects information about organisations' associated registration at Companies House. If two charities share the same companyid they match the same companies house record and are therefore linked.

This same company id is used to generate matches between CCEW and Companies House directly.

Our historic data from Scotland's charity regulator, OSCR, contains a company identifier for some organisations. This is used for a deterministic match between OSCR and Companies House.

Organisations in the Cooperatives register may have an additional id, 'Registered Number', which represents its linkage to the Mutuels Public Register. Matches are thus sought between CoOps and Mutuels using this id.

- Matching by Normalised Name:

If two organisations have the same normalised name they are tentatively matched, but require an additional check. Any of the following being true will create a match:

1. Organisation is in Care Quality Commission (CQC) data, and the matched core organisation is flagged as having a CQC registration, which is a flag in CCEW data.
2. Organisation is in the Scottish charity regular (OSCR) register and has a 'cross border' flag, indicating that a match is expected between itself and an organisation in CCEW.
3. Organisation is in Scottish Housing Regulator or Care Inspectorate Scotland, and normalised name match is with an organisation in OSCR.
4. Organisation is in Social Housing England Regulator or Care Quality Commission, and matches an organisation in CCEW.

- Matching using work by Find That Charity:

A lookup table provided by David Kane<sup>3</sup> is also used to generate matches. These links include historic reregistrations, allowing this dataset to be further deduplicated.

The order in Table 5 shows the precedence given to sources where matches occur: the higher precedence source will provide the data and unique identifier (uid) in the main spine table. Matches will be found in the matches table, and any data found in addition to the primary fields will be in the supplementary table, indexed by the uid given at source. The first eight sources are processed such that any organisations not already present in the dataset are added to it. Sources CIS and CQC are only noted in the matches table where they link to a source present.

---

<sup>3</sup>[https://raw.githubusercontent.com/drkane/charity-lookups/master/relationships/\\_sameas.csv](https://raw.githubusercontent.com/drkane/charity-lookups/master/relationships/_sameas.csv)

Order	Source	Code
1	Charity Commission, England and Wales	CHC
2	Scottish Charity Register	SC
3	Charity Commission, Northern Ireland	NIC
4	Companies House	COH
5	Co-operatives	COOP
6	Mutuals Public Register	MPR
7	Scottish Housing Register	SHR
8	Social Housing England	SHPE
-	Care Inspectorate Scotland	CIS
-	Care Quality Commission	CQC

Table 5: Precedence order of sources for matched organisations

## 2.2 Match types visualised

The plots in Figures 1, 2 and 3 show the number of matches between sources in the public spine. Figure 1 is an UpSetPlot, which is representation of intersection between all registers (like a venn diagram), shows how many matches are made between the different registers. Figure 2 is the same plot, without organisations which did not match with any other organisation, allowing for greater visualisation of the size of the matches on each axis. The bar chart in Figure 3 shows the number of matches by match type, showing the distribution of the types of linkages.

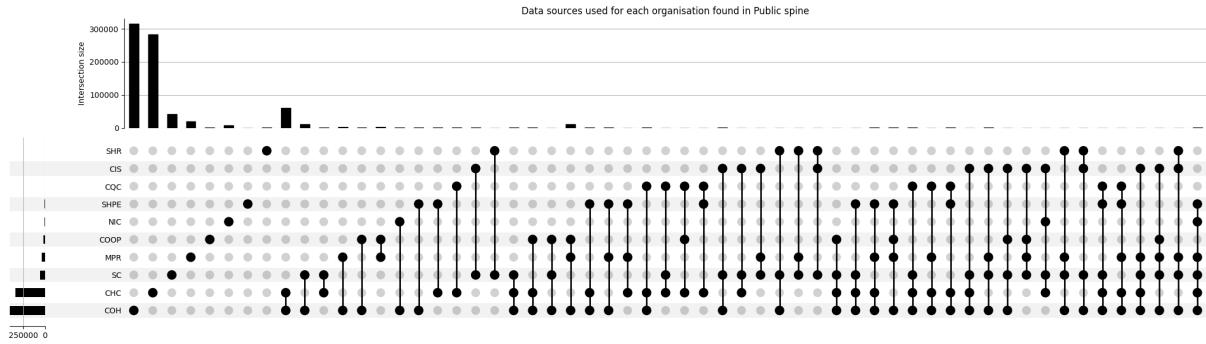


Figure 1: Sources each organisation is registered with

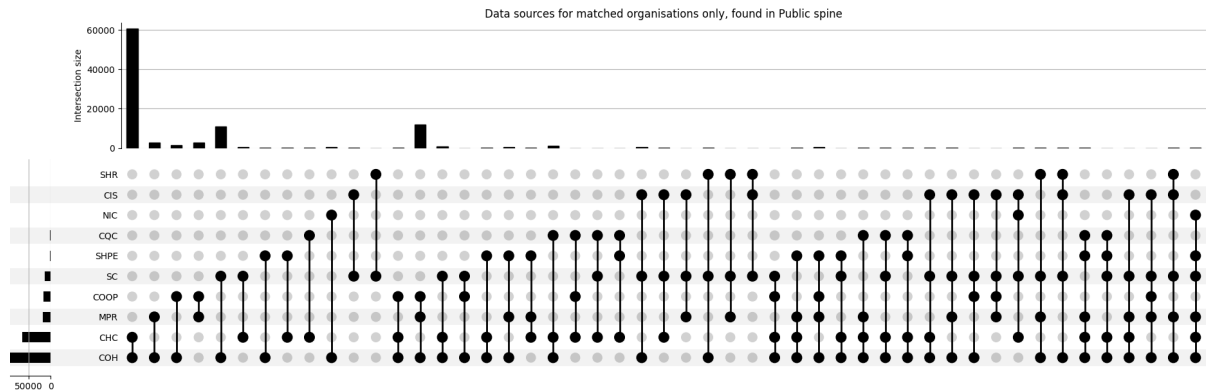


Figure 2: Sources each organisation is registered with, for all organisations in at least 2 registers

## 2.3 Determining primary organisations

The above techniques are used to match organisations between sources. Where a match occurs from within the same source regulator, the primary organisation is determined by the following rules:

- If an organisation from the Charity Commission for England and Wales matches an organisation from the same commission, the primary organisation uses the Find That Charity lookup to identify which of the two is the 'transferee' in a reregistration. The transferor is then linked to the transferee in the matches table, with the transferee being the primary organisation.
- In other cases where there are two organisations from the same regulator which match, the organisation considered primary is the one without a removal date (and is therefore still active), or the one with the most recent removal date (and was therefore active most recently).

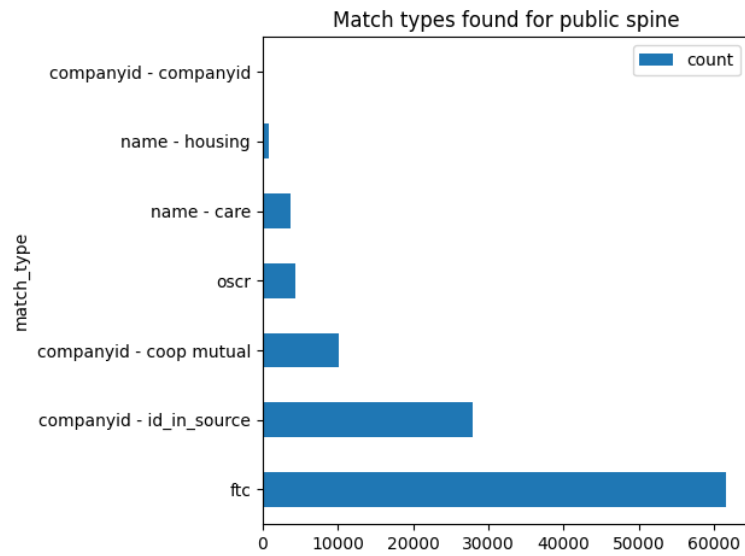


Figure 3: Distribution of match types used to deduplicate the spine

If neither of these conditions lead to a decision, in particular where neither of the matched organisation have a removal date, the organisation with the earliest registration date is considered primary.

### 3 Ingesting data from each source

Each of the ten sources use different conventions for their databases. These needed to be pre-processed individually to accomodate different fieldnames, encodings, systems of marking alternative names, and so on. The pre-process step identified primary names and addresses for each organisation, choosing the most recent where multiple iterations exist, and storing alternatives for completeness.

More information about the processes for each source are found in the appendix, as well as the project source code<sup>4</sup>.

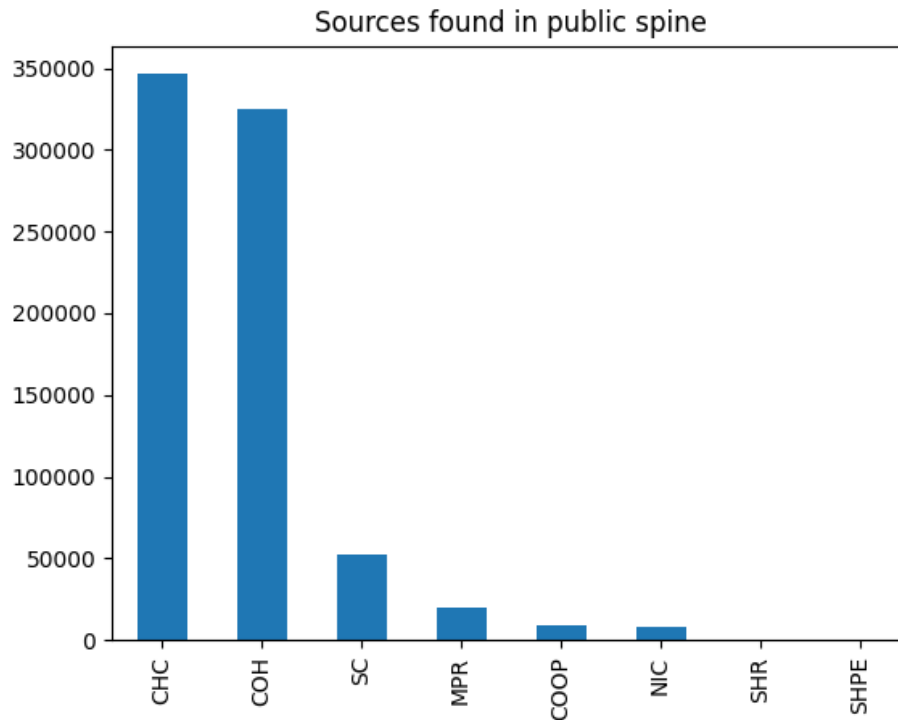


Figure 4: Sources found in public spine

Source	Count of orgs in spine	Count of orgs matched to spine	Total count per source
CHC	346390	5323	351713
SC	52506	1402	53908
NIC	8108	0	8108
COH	325403	74364	399767
COOP	9481	1344	10825
MPR	19903	12018	31921
SHR	109	50	159
SHPE	23	1311	1334
CIS	0	2156	2156
CQC	0	973	973

Table 6: Counts of organisations in the spine, and matched to the spine, from each source register

<sup>4</sup><https://github.com/uk-third-sector-database/tso-database-builder>

## Appendix: Data Sources

All data in the spine comes from publicly available sources. The number of organisations in the database (whether in the list of core organisation or via matching) are shown in Table 6.

Details for each source are given in the sections below.

### .1 Charity Commission for England and Wales (CCEW)

<https://www.gov.uk/government/organisations/charity-commission>

<https://register-of-charities.charitycommission.gov.uk/en/register/full-register-download>

This section outlines the use of CCEW regulatory data in the construction of the third sector organisational spine<sup>5</sup>.

The initial processing of CCEW data for this project comprised 27 individual data files that were obtained between 2001 and 2023, along with updates lat. These files were used to compile a list of unique charities, and to extract as many unique names, addresses, registration dates, and removal dates for each organisation. These files are detailed below:

- CC\_appended\_March23.dta [English and Welsh Charity Register - March 2023]
- CC\_charity\_other\_names.dta [File containing additional names for English and Welsh charities - March 2023]
- ccr2001\_repaired.dta [Historical file from 2001, used to extract charity names, addresses, company numbers, dates of registration and removal]
- 24 historical snapshots from June 2011, March 2012 to 2019, April 2020 to 21, and March 2022 [Used to extract charity names, addresses, dates of registration and removal]

In order to compile all available information from these files, we created separate lists of unique names, addresses, registration dates, and removal dates. These lists were grouped using the identification number assigned to each charity by the CCEW (e.g., Stroud Earth Community in Gloucestershire operates under the charity number 1199586). These lists were then normalised in order to de-duplicate the information. For charity names, this process entailed removing leading spaces, removing trailing spaces, collapsing consecutive internal spaces, and converting all characters to lower case. In the case of addresses, the same process was followed using a concatenated string comprising the first address line and the postcode. For registration and removal dates, we converted all dates to the same format (dd/mm/yyyy).

Once this normalisation was complete, we de-duplicated each list to keep only unique pieces of information. This de-duplication process keeps the first appearance of each piece of information and we provide variables in the dataset to show from which data iteration each value originates. These separate list of unique names, addresses, and dates are then compiled into a single dataset which is later merged with data from all other data sources. Using the historical file from 2001, we also extract 4,700 Companies House identification numbers that are absent from the later files we use (i.e., 2011–2023). Later in the process, we use Companies House IDs to link together the various data sources that produce the third sector organisational spine. The codebook of variables is found in Table 7.

#### .1.1 Limitations

Despite the extensive temporal coverage of these data files (2001–2023), there are 13,243 unique charity numbers in the data for which we do not have address information. A large majority of these organisations were removed

---

<sup>5</sup>For further detail on how this data was used, please see the accompanying `ccew-publicspine-prep.do` file.



from the CCEW register before 1997. Similarly, three of these charity numbers do not have an organisation name in any of the data iterations used.

## .1.2 Notes

For a subset of charity numbers, the 2001-2023 data contain multiple dates of registration and/or removal. The data contains 341,990 unique charity numbers with 12,148 (3.55%) having more than one registration date. Of the 177,706 charity numbers that have a removal date, 4,074 (2.29%) have more than one. Below are some possible explanations for a charity having more than one date of registration or removal, though these are not exhaustive.

1. Data entry errors that are corrected in later data iterations.
2. The CCEW removes organisations that they believe have ceased operating. In some cases, the organisations are still active and the CCEW adds them back to the register once they have confirmed their continued activity. This creates multiple distinct dates for the same charity number.
3. These multiple values show the registration/removal dates for distinct organisations that have since merged into a single organisation with a single charity number. In this case, the component parts of the merger retain their distinct dates in historical data files and these are kept in the final organisational spine dataset. Related to point 3 above, there are charity numbers and organisations in the organisational spine that have since been merged into single entities. For more detail on this, please see data on merged charities held by the CCEW (<https://www.gov.uk/government/publications/register-of-merged-charities>).

Fieldname	Notes
uid	Unique ID assigned to each organisation in compiled data
charitynumber	ID assigned to each organisation by CCEW
organisationname	Name of organisation
normalisedname	
companyid	ID assigned to organisation by Companies House
hounumber	
addressline1	First line of address
addressline2	Second line of address
addressline3	Third line of address
addressline4	Fourth line of address
addressline5	Fifth line of address
city	City
localauthority	Local Authority area
postcode	Postcode
registerdate	Date of registration
removeddate	Date of removal
name_origin	Iteration from which the organisation name originates
address_origin	Iteration from which the organisation address originates
regdate_origin	Iteration from which the registration date originates
remdate_origin	Iteration from which the removal date originates
iteration	Iteration from which all data in a given row originates
source	Variable to indicate the data source (in this case, CCEW)

Table 7: Fields in pre-processing of CCEW data

## .1.3 Updates

Subsequent to the initial dataset created using this process, data downloads were acquired from CCEW as 'public extracts' in 2024 and 2025. These were normalised, pre-processed and deduplicated for updated dataset delivery. The main sections of this document detail the numbers ingested from each source, which include these updates.

## .2 Scottish Charity Regulator (OSCR)

<https://www.oscr.org.uk/>

This section outlines the use of OSCR regulatory data in the construction of the third sector organisational spine<sup>6</sup>.

The OSCR data used in this project comprised eight individual data files that were obtained between 2012 and 2025. These files were used to compile a list of unique charities, and to extract as many unique names, addresses, registration dates, and removal dates for each organisation. These files are detailed below:

- CharityExport-05-Feb-2025.csv : Scottish Charity Register, February 2025
- CharityExport-Removed-05-Feb-2025.csv : List of removed charities, February 2025
- CharityExport-12-Apr-2023.csv : Scottish Charity Register, April 2023
- CharityExport-Removed-12-Apr-2023.csv : List of removed charities, April 2023
- oscr\_20210909.xlsx : Scottish Charity Register and Annual Returns, September 2021
- oscr\_20210203.xlsx : Scottish Charity Register and Annual Returns, February 2021
- ocsr\_scr-and-ar\_20190109.xlsx : Scottish Charity Register and Annual Returns, January 2019
- sc\_chars2012.xlsx [Historical file from 2012]

In order to compile all available information from these files, we created separate lists of unique names, addresses, registration dates, and removal dates. These lists were grouped using the identification number assigned to each charity by OSCR (e.g., Inverness Civic Trust operates under the charity number SC021748). These lists were then normalised in order to de-duplicate the information. For charity names, this process entailed removing leading spaces, removing trailing spaces, collapsing consecutive internal spaces, and converting all characters to lower case. In the case of addresses, the same process was followed using a concatenated string comprising the first address line and the postcode. For registration and removal dates, we converted all dates to the same format (dd/mm/yyyy).

Once this normalisation was complete, we de-duplicated each list to keep only unique pieces of information. This de-duplication process keeps the first appearance of each piece of information and we provide variables in the dataset to show from which data iteration each value originates. These separate list of unique names, addresses, and dates are then compiled into a single dataset which is later merged with data from all other data sources. Finally, the 2012 historical data file includes additional charity regulator and companies house ID numbers. These additional IDs were gathered at the time the data file was created and we include them in the final dataset in case they prove useful for future research.

The codebook of variables is found in Table 8.

### .2.1 Limitations

There are 1,776 unique charity numbers in this data that don't have organisation names or any address information. Similarly, there is no useful data on registration or removal dates for these organisations. OSCR uses default values for registration (01/01/1992) and removal (02/06/2012) dates for many organisations in their data. For these 1,776 charities, all have the default removal date and all but two have the default registration date. Due to the lack of organisation names, addresses, and dates for these organisations, we exclude them from the final dataset.

In addition to the 1,776 charities outlined above, there are 7,591 charities for which we don't have any address data. Of these 7,591 charities, 7,572 were removed from the register on the default date of 02/06/2012. Despite these limitations, these charities are retained in the final data because we do have info on charity name,

---

<sup>6</sup>For further detail on how this data was used, please see the accompanying oscr-publicspine-prep.do file.

Fieldname	Notes
uid	Unique ID assigned to each organisation in compiled data
charitynumber	ID assigned to each organisation by OSCR
organisationname	Name of organisation
normalisedname	
companyid	ID assigned to organisation by Companies House
address	Full street address
houenumber	
addressline1	First line of address
addressline2	Second line of address
addressline3	Third line of address
addressline4	Fourth line of address
addressline5	Fifth line of address
addressline6	Sixth line of address
addressline7	Seventh line of address
addressline8	Eighth line of address
city	City
localauthority	Local Authority area
postcode	Postcode
registerdate	Date of registration
removeddate	Date of removal
name_origin	Iteration from which the organisation name originates
iteration	Iteration from which all data in a given row originates
charitynumber_2012	Alternative OSCR ID (originates in 2012 data)
companyid1_2012	Alternative Companies House ID (originates in 2012 data)
companyid2_2012	Alternative Companies House ID (originates in 2012 data)
companyid3_2012	Alternative Companies House ID (originates in 2012 data)
source	Variable to indicate the data source (in this case, OSCR)
crossborder	Boolean variable, reflecting whether 'Regulatory Type' is 'Cross Border' for this organisation

Table 8: Fields in pre-processing of OSCR data

registration date, and removal date. Finally, there are 194 charities for which we have street address information, but the data files do not contain the accompanying postcode. Of these 194 charities, 189 were removed from the charity register on the default date of 02/06/2012. These charities are also retained in the final data because we do have info on charity name, registration date, removal date, and partial address data.

## .2.2 Notes

The 2012 historical data file does not include any information on addresses, registration dates, or removal dates. As such, many of the names in our data file are first observed in 2012 while the earliest point of observation for the other key pieces of information is 2019.

### .3 Charity Commission for Northern Ireland (CCNI)

<https://www.charitycommissionni.org.uk/>

This section outlines the use of CCNI regulatory data in the construction of the third sector organisational spine<sup>7</sup>.

The CCNI data used in this project comprised of open data files obtained in

- April 2023 (ccni-charity-register-2023-04-19.csv [CCNI charity register - 19 April 2023])
- January 2025 (ccni-charitydetails\_2025\_01\_29\_14.23\_47.csv [CCNI charity details - 29 January 2025])

These files were used to compile a list of unique charities with accompanying names, addresses, and registration dates for each organisation.

Unlike the preceding sections on CCEW and OSCR data, the research team did not have historical files to supplement the open data provided by the CCNI. Additionally, the CCNI data provide only one organisation name, address, and registration date per CCNI charity number (e.g., Sólás in Belfast operates under the charity number 100114). For organisations whose details differed between the two files, the most recent data was prioritised for the spine, and older data retained in the 'supplementary' table. This dataset is later merged with data from all other data sources.

The codebook of variables is found in Table 9.

#### .3.1 Limitations

Unlike the other charity regulator data sources, our data for CCNI does not alternative or historic unique names, addresses, or registration dates. At this point in time, the CCNI data is less comprehensive than its counterparts in England, Wales, and Scotland. Additionally, the CCNI does not provide information on the date that a charity is removed from the charity register. This further limits the CCNI data at this time.

Fieldname	Notes
uid	Unique ID assigned to each organisation in compiled data
charitynumber	ID assigned to each organisation by CCNI
organisationname	Name of organisation
normalisedname	
companyid	ID assigned to organisation by Companies House
hounumber	
address	Full street address and postcode
city	City
localauthority	Local Authority area
postcode	Postcode
registerdate	Date of registration
source	Variable to indicate the data source (in this case - CCNI)

Table 9: Fields in pre-processing of CCNI data

---

<sup>7</sup>For further detail on how this data was used, please see the accompanying ccni-publicspine\_prep.do file.

## .4 Companies House

<https://www.gov.uk/government/organisations/companies-house>

Companies House release a bulk download every quarter of all currently live companies<sup>8</sup>. The Alpha public spine incorporates these downloads for April, June and October 2023, March, May and September 2024, and January 2025.

Using bulk downloads alone would mean organisations that were dissolved earlier than 2023 would not be present in the database. Using the Companies House API we found all companies registered and dissolved which fit our inclusion criteria (see Table 10 for included company types). The code used to search the API is found in our github repository [https://github.com/uk-third-sector-database/ch\\_adv\\_scraper](https://github.com/uk-third-sector-database/ch_adv_scraper).

### .4.1 Company types

Companies house data includes a 'company category' for each organisation. Table 10 shows the full list of these, indicating which have been included in the public spine data.

Company category	Included / Excluded
Charitable Incorporated Organisation	Included
Community Interest Company	Included
Industrial and Provident Society	Included
PRI/LBG/NSC (Private, Limited by guarantee, no share capital, use of 'Limited' exemption)	Included
PRI/LTD BY GUAR/NSC (Private, limited by guarantee, no share capital)	Included
Registered Society	Included
Scottish Charitable Incorporated Organisation	Included
Converted/Closed	Included
Private Limited Company	Excluded
Limited Partnership	Excluded
Limited Liability Partnership	Excluded
Public Limited Company	Excluded
Private Unlimited Company	Excluded
Scottish Partnership	Excluded
Private Unlimited	Excluded
Investment Company with Variable Capital(Umbrella)	Excluded
PRIV LTD SECT. 30 (Private limited company, section 30 of the Companies Act)	Excluded
Investment Company with Variable Capital (Securities)	Excluded
Investment Company with Variable Capital	Excluded
Overseas Entity	Excluded
United Kingdom Economic Interest Grouping	Excluded
Old Public Company	Excluded
United Kingdom Societas	Excluded
Other Company Type	Excluded
Protected Cell Company	Excluded
Royal Charter Company	Excluded
Further Education and Sixth Form College Corps	Excluded
Other company type	Excluded

Table 10: Companies house 'company category' field

Each of the companies house data sources were pre-processed to identify primary name and address for each included organisation, with any alternative names and addresses also stored. The sources were then also combined, since there is significant overlap between them, and the most recent name and address per organisation

<sup>8</sup>[https://download.companieshouse.gov.uk/en\\_output.html](https://download.companieshouse.gov.uk/en_output.html)

was included in the companies house spine data, to be used for matching with other registries to form the public spine.

Compressing the company type categories such that all Community Interest Companies are grouped, and all remaining companies which are limited by guarantee (with or without 'limited exemption', with or without 'private'), gives the counts shown in Table 11, and in Figure 5.

Company type	Count
Company Limited by Guarantee	436872
Community Interest Company	93216
Charitable Incorporated Organisation	38318
Registered Society	10957
Scottish Charitable Incorporated Organisation	7002
Industrial and Provident Society	162

Table 11: Distribution of company types in dataset (compressed company types).

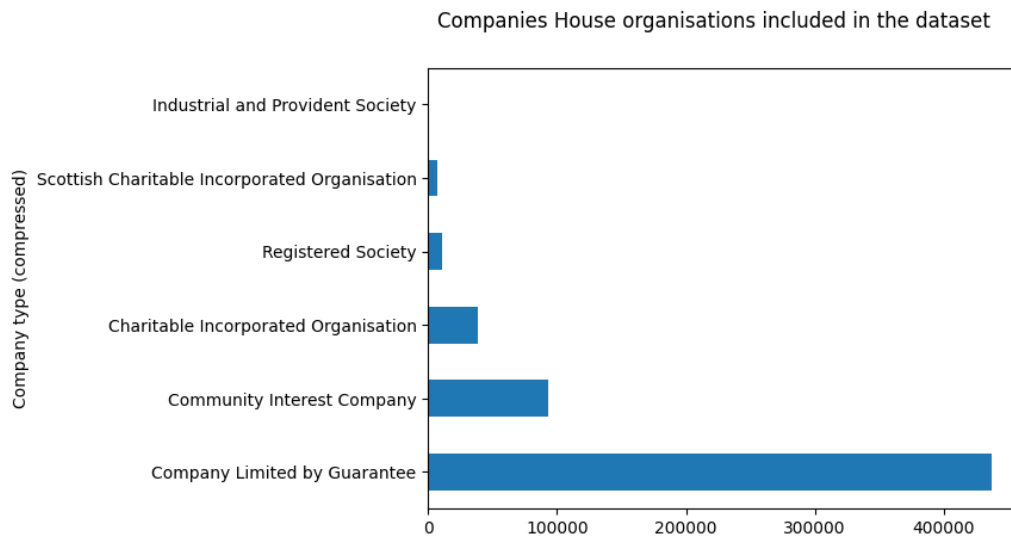


Figure 5: Companies house organisations by company type (compressed).

## .5 Co-operatives

<https://www.uk.coop/resources/open-data>

The Co-operatives UK organisations used in this project were from open data downloads in April 2022, April 2024 and January 2025. Open data from the Co-operatives UK organisation was mapped to the format required by the public spine according to the rules shown in Table 12 below.

Fieldname	Notes
uid	Unique ID assigned to each organisation using 'GB-COOP-' + <b>CUK Organisation ID</b>
organisationname	<b>Registered Name</b>
normalisedname	Normalised version of organisationname
companyid	<b>Registered Number</b> (COOPs data links to Mutuals data using this field)
fulladdress	<b>Registered Street</b>
city	<b>Registered City</b>
postcode	<b>Registered Postcode</b>
registerdate	<b>Incorporation Date</b>
removeddate	<b>Dissolved Date</b>
source	Variable to indicate the data source (in this case - COOP)
organisationname (in supplementary data)	<b>Trading Name</b>

Table 12: Fields in pre-processing of COOP data (bold text indicates fieldname in co-op open data)

Where an organisation has multiple options for name and address, the most recent is used in the primary data, and alternatives are stored in supplementary data.

## .6 Mutuals Public Register

<https://mutuals.fca.org.uk/>

The Mutuals Public Register is a public record of mutual societies registered by the Financial Conduct Authority. Openly available register data has information for societies currently registered, and those no longer registered. For this project, downloads from June 2023, April 2024 and January 2025 are used. The field mapping is shown in Table 13.

Fieldname	Notes
uid	Unique ID assigned to each organisation using 'GB-MPR-' + <b>Full Registration Number</b>
organisationname	<b>Society Name</b>
normalisedname	Normalised version of organisationname
companyid	
fulladdress	<b>Society Address</b>
city	
postcode	Postcode taken from <b>Society Address</b> : = last two 'words' of address if their lengths are at most 4 characters each
registerdate	<b>Registration Date</b> mapped from yyyy-mm-dd format to dd/mm/yyyy
removeddate	<b>Deregistration Date</b> mapped from yyyy-mm-dd format to dd/mm/yyyy
source	Variable to indicate the data source (in this case - MPR)

Table 13: Fields in pre-processing of Mutuals data (bold text indicates fieldname in Mutuals Public Register data)

Where an organisation has multiple options for name and address, the most recent is used in the primary data, and alternatives are stored in supplementary data.



## .7 Scottish Housing Register

<https://www.housingregulator.gov.scot/landlord-performance/statistical-information/>

The Scottish Housing Regulator for social landlords publishes organisational data which has been included in the spine. Datasets used in this project were annual downloads collected in 2018, 2019, 2020, 2021 and 2022, and their most recent publication at the time of creating our data product: 'All Social Landlords Complete Dataset 2019-20 to 2023-24'.

Fieldname	Notes
uid	Unique ID assigned to each organisation using 'GB-SHR-' + <b>Reg No</b>
organisationname	<b>Social Landlord</b>
normalisedname	Normalised version of organisationname
companyid	
fulladdress	
city	
postcode	
registerdate	
removeddate	
source	Variable to indicate the data source (in this case - SHR)

Table 14: Fields in pre-processing of Scottish Housing Regulator data (bold text indicates fieldname in register)

Where an organisation has multiple options for name and address, the most recent is used in the primary data, and alternatives are stored in supplementary data.

## .8 Social Housing England

<https://www.gov.uk/government/publications/registered-providers-of-social-housing>

Two bulk downloads were used from Social Housing England's register, from 2023 and 2025.

Fieldname	Notes
uid	Unique ID assigned to each organisation using 'GB-SHPE-' + <b>Registration number</b>
organisationname	<b>Organisation name</b>
normalisedname	Normalised version of organisationname
companyid	
fulladdress	
city	
postcode	
registerdate	<b>Registration date</b>
removeddate	
source	Variable to indicate the data source (in this case - SHPE)

Table 15: Fields in pre-processing of Social Housing Providers England data (bold text indicates fieldname in register)

Where an organisation has multiple options for name and address, the most recent is used in the primary data, and alternatives are stored in supplementary data.

## .9 Care Inspectorate Scotland

<https://www.careinspectorate.com/index.php/statistics-and-analysis/data-and-analysis>

Care Inspectorate Scotland release data at their financial year end (March 31st). This dataset uses data published in the years 2015 to 2025 inclusive. Organisations are included where field "Service Type" = 'Voluntary or Not for Profit'. Organisations have both a 'Provider Name' and a 'Service Name' in the Care Inspectorate data. We have retained Service Names, however the primary name for matching is the Provider Name.

Fieldname	Notes
uid	Unique ID assigned to each organisation using 'GB-CIS-' + <b>CSNumber</b> or <b>CaseNumber</b>
organisationname	<b>ServiceProvider</b>
normalisedname	Normalised version of organisationname
companyid	
fulladdress	Concatenation of <b>Address_line_1</b> , <b>Address_line_2</b> , <b>Address_line_3</b> , and <b>Address_line_4</b>
city	<b>Service_town</b>
postcode	<b>Service_Postcode</b>
registerdate	<b>DateReg</b>
removeddate	
source	Variable to indicate the data source (in this case - CIS)
organisationname (in supplementary data)	<b>ServiceName</b>

Table 16: Fields in pre-processing of Care Inspectorate Scotland data (bold text indicates fieldname in register)

## .10 Care Quality Commission

<https://www.cqc.org.uk/about-us/transparency/using-cqc-data>

Fieldname	Notes
uid	Unique ID assigned to each organisation using 'GB-CQC-' + <b>CQC Provider ID (for office use only)</b>
organisationname	<b>Provider name</b>
normalisedname	Normalised version of organisationname
companyid	
fulladdress	<b>Address</b>
city	
postcode	<b>Postcode</b>
registerdate	
removeddate	
source	Variable to indicate the data source (in this case - CQC)
organisationname (in supplementary data)	<b>Name</b>
organisationname (in supplementary data)	<b>Also known as</b>

Table 17: Fields in pre-processing of Care Inspectorate Scotland data (bold text indicates fieldname in register)