

Taking Account 4

Organisation Mapping Exercise Methodology

Context

The aim of the exercise, undertaken by [Oxford Consultants for Social Inclusion](#) (OCSI) was to assess the 'overlap' of networks between different organisations and assess the extent of duplication between data provided by three local organisations.

The mapping exercise formed part of the larger Taking Account 4 project undertaken by Community Works and funded by Brighton and Hove City Council. The idea for a mapping exercise arose out of the need to understand the applicability of the 'known' list of organisations that Community Works engages with. That is, to what extent can we say that a given list of third sector organisations is 'representative' of the (inherently unknowable) wider third sector as a whole?

While we cannot 'know' with complete certainty the entire ecosystem of charities and third sector organisations, we can seek to understand the data that we do know by comparing it to other organisations acting in the same network, but with different purposes.

Methodology

In order to determine a reasonably accurate combined list of third sector organisations, organisation names were compared using an automated process where possible, and a manual process once automation became more complex.

Lists were taken from contact information held by three third sector infrastructure organisations: Community Works, the Trust for Developing Communities and the Resource Centre.

A fully automated process was not possible, due to inconsistencies between data schemes, naming formats, and missing data. As a single identifying number or code was not available across all organisations (incorporated or unincorporated), organisation names were used to compare between the available lists.

Each organisation's name was reduced to a simplified reference code, which:

- removed all punctuation, spaces, and non-alpha-numeric characters
- removed common words, such as "the", "and", "a", etc.
- removed organisation structure identifiers, such as "Limited", "Group", "Club", "CIC", etc.

Following the automated conversion process and an initial set of matching duplicates (eg where names match between lists exactly, or differ only in punctuation), several rounds of manual checking were then required to identify further words which could be cleaned, and to manually amend organisation names known to be incorrect and/or duplicates.

This revealed certain inconsistencies and ambiguities with names which would require further investigation to automate. Prime examples of these are:

- Simple typos, or differences in singular/plural names
- Use of acronyms, eg "An Example Organisation", "An Example Organisation (AEO)", "AEO", etc.
- Ambiguity between organisations of the same name, but feasibly different organisations. This was particularly the case with churches and schools, which often reference a Saint's name.

- Ordering within names, eg “2nd Scout Group (Brighton)”, “Brighton 2nd Scouts Group”

Duplicates were identified by the resulting reference code, and filtered out. Non-duplicates were then assigned a category for inclusion in counting (“Do include”, “Maybe include”, “Do not include”), based on the source of the organisation and the estimated confidence of these sources.

The process was carried out in a proof-of-concept spreadsheet, which allowed for instant summaries of the number of organisations, along with number of duplicates, incorporated organisations, ambiguous names, etc. This spreadsheet could form the basis of a more general algorithm to identify and analyse overlap between lists in future. As above, the algorithm to match organisations could also then be improved over time.