

Third Sector Research Centre

Working Paper 93

Collecting and classifying data from charity accounts for England and Wales

David Kane (NCVO)

Jenny Clark (NCVO)

David Clifford (TSRC)

John Mohan (TSRC)

Joy Dobbs (NCVO Associate)

Pete Bass (NCVO Associate)

January 2013

Abstract

A key research need for the voluntary sector is consistent, reliable and useful data on the finances of voluntary organisations. This paper discusses challenges which arise in collecting data from charity accounts, and describes a new process, developed by the Third Sector Research Centre and the National Council for Voluntary Organisations, for collecting data from that source. The process aims to create a shared, consistent dataset that will provide a rigorous base for future study of the voluntary sector.

We devised a sample of 10,000 charities, designed to permit totals to be estimated with accuracy (for example, total statutory income) by size, region and sub-sector, while also giving some insight into proportions (for example, the proportion of organisations in receipt of statutory income). Data was then captured from annual audited accounts for charities, and entered via a web-based form. Procedures were then developed for the classification of text strings describing income sources into a limited number of categories that could then be used to analyse the data in more detail.

The paper then discusses a number of data quality issues found in the process of collecting data; these relate to the source material, the development of checks to ensure the accuracy of data, and challenges entailed by the accurate classification of account data.

Keywords

Voluntary sector, charity, statistics.

Acknowledgements

TSRC acknowledges the financial support of ESRC, the Barrow Cadbury Trust, and the Office for Civil Society. We are also grateful for statistical advice from Peter Smith (University of Southampton). Data entry was carried out by the Centre for Data Digitisation and Analysis at Queen's University Belfast (www.qub.ac.uk/cdda/CDDA_2011/Welcome/) where we acknowledge the help and advice of Elaine Yeates in particular.

Contents

- 1. Introduction and context 3**
- 2. Data sources and sampling design 5**
- 3. Data collection 6**
- 4. Data classification 7**
- 5. Quality issues 9**
 - Source material 10
 - Data collection process 10
 - Data classification 10
- 6. Next steps and future issues 11**
- Notes 12**
- References 13**

1. Introduction and context

There is a growing interest in the characteristics and activities of voluntary organisations, associated with recognition of the importance of their role in delivering public services (Musick and Wilson, 2008), and providing opportunities to participate in associational life (Putnam, 2000). In the UK, voluntary action has had, since the 1990s, a profile to rival that at any point in its history (Alcock, 2010). However, the evidence base on the voluntary sector has its limitations. Salamon et al. (2012) argue that the voluntary sector has ‘long been hidden in official statistical mapping’, while Burke (2001) once argued that the voluntary sector was a ‘statistical stepchild of the most neglected sort’. He was referring to the USA in the 1970s, but Tarling (2000) also noted the absence of reliable statistical evidence for the UK. Although the situation has been improving, there is an acknowledged need for better-quality information (UK Statistics Authority, 2012).

One of the principal reasons for this is that economic statistics tend to be classified by sector of activity (for example, primary production (agriculture, mining); secondary production (for example, manufacturing) and tertiary activity (the delivery of services)). Within these there are subsectors such as education, health and social care. Organisations are allocated to these categories and subcategories, and no distinction is formally made as regards the legal form (for example, not-for-profit, for profit, government) of the entities carrying out the activity. In the guidelines developed by the United Nations Statistical Division for national accounting, the *System of National Accounts* (SNA), voluntary organisations have tended to be classified within other sectors. For this reason, the scale of the voluntary sector has been systematically underestimated in official statistics. Efforts are underway to rectify this situation, notably through the Comparative Nonprofit sector project, led by Johns Hopkins University in the United States, which seeks to co-ordinate a more comprehensive, reliable, and internationally comparative approach to the measurement of the voluntary sector.¹

Such initiatives are important. Valid and reliable statistics will enhance the visibility and credibility of voluntary organisations, contribute to improved transparency and accountability of the sector and of government, and provide empirical context for decision making and policy development (Salamon et al., 2012). Such data are important, too, for our understanding of the sector: the development and testing of relevant theory is hindered without a basic understanding of its scale and structure. In other papers we will describe efforts being made by TSRC and also by others to contribute to the evidence base. These will include work on the classification of organisations (which organisations are genuinely part of the third sector or not, and why?), on the construction of panel data allowing more sophisticated analysis of change over time, on the establishment of datasets from regulators and other sources covering the broader third sector (that is, expanding beyond the population of registered charities which forms the core of the sector), and the development of UK-wide statistics.

The focus of this paper is narrower. It describes the development of a rich data resource on charities in England and Wales. The specific contribution made by this resource concerns the classification of income obtained by registered charities. One possibility would have been to use readily-available survey data. The *National Survey of Third Sector Organisations 2008* (Cabinet Office et al., 2010), and the follow-up *National Survey of Charities and Social Enterprises 2010*, provide

useful information on financial streams – but data are banded, making the accurate estimation of totals difficult. In addition, data on the sources of income are restricted to information on whether an organisation receives income from a particular source – there is no information on the share of income from particular sources.

An alternative is to draw upon data supplied to the Charity Commission, a non-ministerial government body which registers and regulates charities in England and Wales. Its system of registration is well developed, and the use of administrative data from the Charity Commission for this purpose is well established. In contrast, in many countries where registration is not compulsory, there is no incentive to organisations to register – leaving a less comprehensive listing of organisations (United Nations, 2011). We use the Charity Commission register as a sampling frame for a nationally representative survey of charity accounts. This survey provides more detail than is available in the information provided in the annual returns completed by charities for the Charity Commission. In particular, it provides information on financial streams for a range of sizes of charities (not just those with an income above the £500,000 threshold for the reporting of detailed financial information) and information on the source of income (allowing research, for example, on the extent of funding from government) that is not available in the annual returns.

First we examine the context within which the data is situated. Existing data sources for exploring the size and scope of voluntary organisations are centred on the Charity Commission Register of Charities (though there are others). Various registration and administrative thresholds apply to the information that needs to be submitted to the Charity Commission, which affect the comprehensiveness and coverage of the resulting data.

Next, we look at the sampling methodology needed for the work. With a population of over 160,000 organisations, it is not realistic to obtain detailed information on every organisation. We divided the charities in the sector into different groups ('strata') according to size. We then randomly sampled within each of these groups. The aim was to design a sample which would allow for totals to be estimated with accuracy (for example, total statutory income) by size, region and sub-sector, while also giving some insight into proportions (for example, the proportion of organisations in receipt of statutory income).

The next step in the process is data collection. The annual audited accounts for charities are available from the Charity Commission, and information captured from these was entered via a web-based form. In total over half a million rows were entered from the sampled accounts, each row representing a single data item within each account. Typically, each row contained a description taken from the account, such as 'Big Lottery Fund grant', alongside the amount. These text strings needed to be classified to produce a limited number of categories that could then be used to analyse the data in more detail.

The paper then discusses a number of data quality issues found in the process of collecting data. These issues fall into a number of distinct areas:

1. Issues with the source material. In part, this section forms a response to some of the issues raised by Morgan (2011).

2. Issues with data collection. In general, there were very few issues caused by wrongly-entered data. In part this was due to checks put in place at data collection stage.
3. Potential problems with the classification of account data. The ambiguous nature of some account data, and the scale and range of possible values that need to be classified can make classification very difficult.

Finally, we discuss the next steps and future possibilities for the data. Currently data collection is a manual process involving account documents. We discuss ways the process could be improved, but future processes may involve progressing to digital formats such as XBRL, which would allow much easier and quicker data collection.

2. Data sources and sampling design

The sampling frame for the survey was provided by a list of the population of organisations on the Charity Commission Register. Charities are voluntary organisations which benefit the public in a way the law says is charitable. They should work for the benefit of the public as a whole, or a significant section of the public. Their purposes should be charitable, and these include the four heads of charity: the relief of poverty; the advancement of education; the advancement of religion; and other purposes beneficial to the community.

The aim is to obtain reliable information on the population of registered charities in England and Wales (around 160,000 organisations). Since it would be very costly to examine the accounts of every charity, we collect data on a sample of the population (around 10,000 organisations). The principles of sampling are well established and are foundational to much empirical work in social sciences: we can make inferences about the population based on a well-designed sample.

The inference will be reliable only if the sample is representative of the 'target population' of registered charities about which we want to make conclusions. An important principle is that the sample should be randomly selected from the population – so that each organisation has a known, non-zero probability of being included in the sample.

We also stratify the population of charities into groups. We know that on average (in a hypothetical scenario in which we take a number of different samples), a random sample will give unbiased estimates of the population as a whole. However, we only take one sample. Therefore we ensure our sample is representative of the population in certain ways by dividing it into certain groups and then taking a random sample within each of these groups. This ensures that we get a representative spread of these groups in our sample. So, to ensure that we're including charities of all sizes and right across the country in our sample, we stratify by size.

We also want to minimise 'sampling variability' – which refers to the tendency for different samples, even if they are random, to give slightly different answers. The aim was to design a sample which would allow for totals to be estimated with some precision (for example, total statutory income across the charitable sector), while also giving some insight into proportions (for example, the proportion of organisations in receipt of statutory income). To minimise sampling variability for totals, we would design our sample in such a way that we oversample the big charities heavily (then 'weight' our totals at the end to ensure our estimates are 'unbiased'.) To minimise sampling variability for the 'proportion',

we wouldn't oversample the bigger ones. Therefore, we decided on a compromise between these two different scenarios. We oversampled the big charities to ensure that totals were well estimated, but we also ensure that we have enough charities of all sizes in our sample so that we get good estimates of income and expenditure for the typical charity. Table 1 presents the sampling fractions (the percentage of the population that are included in our sample) according to headline income.

Table 1: Sampling fractions by size

Income Band	Population	Sample	Sample %
Under £10k	85,961	251	0.3%
£10k–£100k	49,885	1,452	2.9%
£100k–£1m	20,586	2,947	14.3%
£1m–£10m	5,125	4,539	88.6%
Over £10m	912	880	96.5%
Total	162,469	10,069	6.2%

3. Data collection

The aim of the data collection process is to take information that is held in charities' annual accounts, and turn it into structured, machine-readable data that can be aggregated and analysed. The annual accounts of charities are prepared by the charities themselves to agreed standards, audited or independently examined (depending on the size of the charity). They must then be submitted to the Charity Commission within ten months of the financial year-end to which they refer.

The accounts submitted must conform to agreed standards. Large charities should follow the Statements of Recommended Practice (SORP) 2005 (Charity Commission, 2005); there is a simpler standard for small charities which prepare receipts and payments accounts. However, despite these standards there is a wide range of variability in the structure, content and quality of the accounts. Accounts are displayed by the Charity Commission as portable document format (PDF) files on their website, either after being scanned from posted documents or, increasingly, directly submitted as PDFs through a web interface. While these documents exist as downloadable files that can be viewed on most computers, the nature of PDF files, and variations in the content and structure of the accounts, mean that it is not possible to automatically extract information from them.

Instead, manual data entry was necessary, and this was carried out by the Centre for Data Digitisation and Analysis (CDDA), based at Queen's University Belfast. A custom web-based form was developed which allowed the data entry staff to enter data in a way that replicated the hierarchy of data in the accounts. The process was designed to minimise the need for detailed knowledge of the intricacies of charity accounts by taking much of the decision-making out of the process while also retaining as much written data as possible in its raw form so that it would be possible classify the information and then provide an audit trail.

The data entry team were provided with a wide range of helpful information to aid the data entry process. This included templates for certain types of organisation (that is, a separate template for receipts and payments accounts) and prepopulating the form, where possible, with data from the Charity Commission register or information from accounts previously submitted by organisations. The web-based form allows for each line from the account to be entered separately by recording the text and amounts given in the accounts. The median number of account lines collected per account was 51, with a lower quartile of 42 and an upper quartile of 62. Taken together, around 500,000 lines were collected from accounts for the 10,000 accounts in the sample.

4. Data classification

The next task was to classify the data that was collected. The classification process involves taking each account line in turn and coding it into a category for the financial information it contains. For example, 'Income from Charitable Activities' would be coded as 'IC'. The categories are split into six groups depending on the type of financial information recorded. The six groups are:

- Incoming resources (I) – resources the charity receives throughout the year. Sub-categories in this group record the type of income – so whether it is earned, voluntary or investment, rather than the source of the income;
- Expenditure (E) – the charities' spending throughout the year;
- Assets (A) – a snapshot of the charities' assets and liabilities at year end;
- Funds (F) – a snapshot of the charities' funds at year end (total funds should be equal to net assets (A), as they are two ways of splitting the same resources);
- Other financial (O) – other information that is recorded by monetary value, but does not fit into the four categories above. This includes staff costs, reserves, depreciation and others;
- Other non-financial (N) – this includes a range of information that is collected which is not recorded as monetary values. This includes items such as the number of employees and the numbers of volunteers.

Additionally, income items are classified into another set of categories which relate to the source of the income. This allows us to analyse whether the income has come from government, individuals or elsewhere. In this way we develop a two-dimensional picture of the incoming resources of the sector.

Classifying these rows is challenging. The text found in account items is not consistent; there can be many variations of spelling and phrasing between items in the same category, as well as variations in abbreviations or contractions used. The classification system therefore needed to be able to allow for these variations. There is often ambiguity in the text of many items. Text taken from the accounts does not always easily match to one of the categories, and some text strings do not contain enough information by themselves to match to a category. The same text string could be categorised in two different ways across two accounts depending on the characteristics of the charity concerned, and upon where the information is found within the charity's accounts. The final challenge is the sheer volume of data – an annual throughput of 500,000 lines means that manual classification is simply not

feasible, regardless of the level of knowledge of the voluntary sector and charity accounting held by those doing the classification.

The aim of the classification process is therefore to address these challenges as much as possible, and minimise the number of misclassified items. There will always be some degree of error in this process, and some of the subjective judgements needed mean that even with perfect information two experts may well disagree on whether a classification is correct or not.

A range of techniques were developed as part of the process of classifying accounts data. The methods are based around the principle of building up layers of classification. Automatic 'brute force' classification methods are used to ensure that every account item has a classification assigned (even if the chance of that classification being wrong is higher). Account items that are more important for future analysis (such as larger amounts of money, or those attached to the largest organisations) were classified by hand to ensure that they are correct. The judgement of what constitutes a 'large' item or organisation is a subjective one, and is in part driven by the time and resources available for classification. A series of semi-automatic methods were then used to cover as many items as possible using keywords.

Classification was initially done using a web tool which allowed account lines to be manually classified. After a sufficient number of lines had been classified the manual classifications could then be applied to the unclassified item. This process was accomplished using a custom-made PHP² script which went through unclassified items and automatically classified those lines that exactly match an existing classification. If a match was not exact then the script would make a guess at an appropriate classification and this would need to be confirmed by the user. This script was therefore to some extent self-learning, particularly when looking for spelling mistakes and very similar phrasing. While some context was provided for the line using the classification of its parent item, the script largely took each line by itself.

The second technique was manual classification of account items. This was usually done on the largest items by monetary value and those that belong to the accounts of the largest charities. Account rows were identified and exported to Microsoft Excel, and additional contextual information was also extracted for each account line in order to facilitate interpretation. In some cases additional information was required in order to classify ambiguous lines, including looking at narrative text in the account PDF or even directly contacting the charity itself.

The third technique used was based on using Google Refine, a program designed for analysing and cleaning large data sets. Account lines were imported into Google Refine, and keyword searching used to find groups of items that could be classified. To address the problem of classifying items without the context around them, items were displayed with their parent items also included in the text. So where an item was laid out within the hierarchy:

- Expenditure
 - Spending on charitable activities
 - Supporting our members

It would be presented in Google Refine as a single line:

Expenditure – Spending on charitable activities – Supporting our members

Keyword searches would therefore match any word throughout the text string, rather than just the item itself.

These three methods classified over 80% of the account items, leaving around 84,000 items unclassified. The final classification method involved applying a naive Bayesian filter PHP script to the existing classifications, which generated a series of probabilities for each word that was found in a category. These probabilities can then be applied to a new unclassified item, and the probabilities combined to find the most likely category for each unclassified item. The principle is often applied to catching spam emails, by training the filter on a set of emails already classified as spam or not spam. The technique can then be used to determine whether incoming email is spam based on the words in that email. Existing PHP scripts made available on the internet were adapted to produce this script.³

This final method requires no explicit manual input from the user and therefore avoids subjectivity and inconsistency: keywords are generated automatically rather than by hand, and therefore can include words and patterns that would not be considered if they were manually generated. Another advantage of this method is that it will apply a classification to every item no matter what the probability is of it being correct (although it always applies the most likely candidate). This means that this method is useful to provide a base classification for use when no other classification has found an appropriate category, although there is a higher probability that this classification will be incorrect.

The Bayesian classification method is not perfect, however. The nature of the process and the algorithm means that it also can have a natural skew towards particular categories if they are overrepresented in the source material. This means special attention should be given to categories that may be underrepresented.

5. Quality issues

In general, we feel that this process produces a large and good-quality dataset which is an asset to future investigation of the economy of the voluntary sector. Some support for this verdict is provided by work we have done to compare the results of this exercise with information derived from other sources. For example, we have found that there is a substantial degree of consistency between the estimates of the numbers of organisations which received public funding from the 2008 NSTSO, and the estimates generated by the work reported here for the 2008–9 financial year. This arises from the ability of our classification methods to identify public funding streams. We would nevertheless identify three possible data quality issues which need to be acknowledged and mitigated. The first relates to the source material, the accounts of registered charities, and in part our comments form a response to some points raised by Morgan (2011). The second concerns data collection, although, in practice, there were very few issues caused by wrongly-entered data. In part this was due to checks put in place at data collection stage. The third issue relates to the classification of account data. The ambiguous nature of some account data, and the scale and range of possible values that need to be classified can make classification very difficult.

Source material

Charity accounts are designed to ensure that the reader can obtain a true picture of a charity's activities in the relevant financial year. Information can be recorded in different ways for different charities, and therefore consistent data may not be available. Many of the terms used in charity accounts can be unfamiliar and require careful reading to produce a true picture across many accounts. Morgan (2011) examined the suitability of accounts for large scale data analysis. Because of variations in the presentation of charity accounts, he concluded that large scale analysis of data from charity accounts was problematic, and recommended that it should be restricted to a subset of account categories and year-on-year comparisons of the same charity.

The methods presented in this paper do produce a dataset that goes further than Morgan recommends. However, for a number of reasons, we believe that it is still possible to produce useful data while acknowledging the difficulties of producing it. First, the data is often used to produce year-on-year comparisons. These are less problematic if it is assumed that any errors remain consistent across years, as change could be measured. One issue that Morgan does mention, however, is whether there are variations in accounting standards over the years. Second, some of the more technical issues outlined by Morgan, such as the timing for recording grants, might be expected to cancel out in a dataset containing observations from over 10,000 charities. Nevertheless the issues raised by Morgan are important, and they provide both a useful guide to the complexities of this exercise, and pointers to where further development should concentrate. One final point to note is that problems with the source material were particularly found in smaller charities, where accounts were more freeform, with very sparse or strangely laid out accounts.

Data collection process

Data entry is carried out by non-experts, with limited time spent on each individual set of accounts. On a data entry exercise of this scale input errors are impossible to avoid. It is also difficult to replicate the varied hierarchy of account data in a flat database form.

A number of measures were put in place to mitigate these problems. The data collection process was carefully designed to build in knowledge of accounting practices where required. Initial guidance for data entry staff was developed and applied, with any issues found escalated to those more familiar with charity accounts. Where general rules could be derived from issues found, these were incorporated into the guidance. Technological solutions were used as far as possible to guide the entry staff, including the use of account templates and automatic totals that ensured that the accounts add up. Finally the database itself was designed to be as flexible as possible, allowing for the hierarchy of data to be represented.

Data classification

The final category of possible errors is in the data classification process. Again, the large volume of data collected means that it is not possible to manually classify each individual item; automated keyword searches and matching techniques are essential. These keyword techniques were carefully designed and refined based on their success, but there will always be errors created by ambiguous and unclear text.

To combat this the manual process was targeted at the most significant items – those with a large monetary value or attached to a large charity. The keyword processes were tested and refined by manually checking a sample of rows classified by keywords. This checking allowed for ambiguous or incorrect keywords to be identified and changed before rerunning the process.

One final problem with the data classification process was the prevalence of ‘other’ items within the accounts, that is, those items where a large amount of money was grouped under one item labelled ‘other’. An example might be a list of grants received by a charity, with an ‘other grants’ item at the end. Often this last item will contain the majority of the monetary value, either because grants were only listed when grant conditions required them to be, or because a large amount of small grants was received. Where the ‘other’ item was very large it was examined manually, with context provided from other parts of the accounts. Otherwise these items were processed as part of the Bayesian classification method, with the most likely category determined by being based on existing ‘other’ items.

6. Next steps and future issues

This paper has outlined the process undertaken in the first year of data collection for charity accounts data. The data collection exercise will be repeated in future years, with lessons learnt applied to improve the process. The second year of data collection has already taken place. The guidance for data entry operatives was revised based on lessons learnt from the first year and the data entry system was evaluated and changed where necessary to provide the most efficient environment for entering data.

The classification process has also been improved for the second year. A key challenge for the automatic classification methods is to use intelligently the context of an item to aid classification, rather than looking at it in isolation. The second year of classification should also build on the first by matching items across the years for the same charity. We also believe that there are opportunities for improving the Bayesian classification process, in particular by iterating the classification a number of times, checking agreement and refining the keywords used as it progresses.

Over the medium term, we believe there are opportunities to design a better way of producing accounts which aids exercises such as this one. Annual accounts are currently thought of as a standalone document which is published as a PDF or paper document. While the principle of accounts as an accurate record of one year’s activity is important to maintain, however, this can mean that the data within the accounts is locked up (and cannot be easily accessed by machines) and takes time and effort to extract. Emerging data formats and standards, such as the XML-based iXBRL standard,⁴ offer an opportunity for accounts to be presented with data in a more easily accessible format. These formats offer advantages for regulators, researchers and others who use accounts. Their use is at an early stage at the moment, but we believe that there are significant opportunities if takeup were to be accelerated. The challenges are particularly applicable to small charities, who also may not be able to take advantage of developments like iXBRL.

More generally, the value of an exercise of this kind is that it will permit accurate tracking of a consistent set of organisations, and their finances, over time. This is a representative sample of

organisations, in contrast to databases generated by the Charities Aid Foundation, which focus on the largest organisations in the sector. It has also been established at what is a critical time for the sector – and because we began to collect financial information for 2008 onwards, our dataset will pick up the effects on particular income streams both of recessionary conditions and also of changes in the amount and distribution of public funding after 2010. As a consequence, we believe that this research initiative will be of considerable value for research on the sector, both in terms of improving our knowledge of the amount and distribution of funding streams within the sector, and enhancing our ability to track change in funding streams over time.

Notes

¹ <http://ccss.jhu.edu/research-projects/comparative-nonprofit-sector/about-cnp>

² PHP is a language used primarily for web development. In this case it was used as a simple way to connect with the database which holds the data, and to provide a user interface for interacting with it.

³ See <http://mathforum.org/~ken/bayes/bayes.html> and <http://xhtml.net/scripts/PHPNaiveBayesianFilter>.

⁴ This standard describes an electronic format for accounts which can be read easily by both humans and machines.

References

- Alcock, P. (2010) 'Partnership and mainstreaming: voluntary action under New Labour', *TSRC Working Paper 32*, Birmingham: University of Birmingham.
- Burke, C. (2001) 'Nonprofit histories new numbers (and the need for more)', *Nonprofit and Voluntary Sector Quarterly*, 30 (2): 174–203.
- Cabinet Office, Office for Civil Society and Ipsos MORI (2011) *National Survey of Charities and Social Enterprises 2010*, available at www.nscsesurvey.com.
- Cabinet Office, Office of the Third Sector, Ipsos MORI, Social Research Institute and GuideStar UK (2010) *National Survey of Third Sector Organisations, 2008* [computer file]. Colchester, Essex: UK Data Archive [distributor], March 2010. SN: 6381, <http://dx.doi.org/10.5255/UKDA-SN-6381-1>.
- Charity Commission (2005) *Accounting and Reporting by Charities: Statement of Recommended Practice*, available at: www.charitycommission.gov.uk/charity_requirements_guidance/Accounting_and_reporting/Preparing_charity_accounts/sorp05docs.aspx.
- Morgan, G. (2011) 'The Use of UK charity accounts data for researching the performance of voluntary organisations', *Voluntary Sector Review*, 2 (2): 213–30.
- Musick, M. A. and Wilson, J. (2008) *Volunteers: A Social Profile*. Bloomington, IN: Indiana University Press.
- Putnam, R. (2000) *Bowling Alone: the collapse and revival of American community*. New York: Simon & Schuster.
- Salamon, L. M., Sokolowski, S. W. and Haddock, M. (2012) 'Measuring civil society and volunteering: new comparative findings from the implementation of the UN Nonprofit Handbook in 17 countries', Paper presented at the 10th International Society for Third Sector Research Conference, Siena, 2012.
- Tarling, R. (2000). 'Statistics on the voluntary sector', *Journal of the Royal Statistical Society*, Series A, 163 (3): 255–61.
- UK Statistics Authority (2012) Official Statistics and the Voluntary Sector, available at: www.statisticsauthority.gov.uk/reports---correspondence/correspondence/letter-from-andrew-dilnot-to-nick-hurd-03052012.pdf.
- United Nations (2011) *State of the World's Volunteerism Report*. Bonn: United Nations Volunteers.

About the Centre

The third sector provides support and services to millions of people. Whether providing front-line services, making policy or campaigning for change, good quality research is vital for organisations to achieve the best possible impact. The Third Sector Research Centre exists to develop the evidence base on, for and with the third sector in the UK. Working closely with practitioners, policy-makers and other academics, TSRC is undertaking and reviewing research, and making this research widely available. The Centre works in collaboration with the third sector, ensuring its research reflects the realities of those working within it, and helping to build the sector's capacity to use and conduct research.

Third Sector Research Centre, Park House, 40 Edgbaston Park Road,
University of Birmingham, Birmingham, B15 2RT

Tel: 0121 414 3086

Email: info@tsrc.ac.uk

www.tsrc.ac.uk

Quantitative Analysis

This research stream is designed to improve our understanding of the third sector through a large-scale programme of quantitative work. It is designed to help us better explain the distribution of third sector organisations, analyse their contribution to society and the economy and understand their dynamics. We are interested in data not just on third sector organisations and their resources, but also on both financial inputs to the sector (funding flows from various sources) and human inputs (e.g. the paid workforce and volunteers).

Contact the author

www.ncvo-vol.org.uk

ncvo@ncvo-vol.org.uk

020 7713 6161

David Clifford

d.clifford@tsrc.ac.uk

023 80594442



This work is licensed under the Creative Commons [Attribution-NonCommercial-NoDerivs 3.0 Unported License](https://creativecommons.org/licenses/by-nc-nd/3.0/).

© TSRC 2013

The support of the Economic and Social Research Council (ESRC), the Office for Civil Society (OCS) and the Barrow Cadbury UK Trust is gratefully acknowledged. The work was part of the programme of the joint ESRC, OCS Barrow Cadbury Third Sector Research Centre.