# Notes on methods used for matching data on grants made as part of the 360 Degree Giving Initiative

# Data sources

The main source of data from grant makers is data released as part of the 360 Giving initiative, which provides:

“support for grant makers to publish their grants data openly, to understand their data, and to use the data to create online tools that make grant-making more effective.”

The 360 giving initiative is led by grant makers, with the aim of encouraging their peers to also release data. It is a long term initiative which has from late 2015 is now a charity in its own right.

Data was also sourced from funders themselves where they have not taken part in the 360 giving initiative. This was largely an ad-hoc process based on funders responding to calls for data or via personal contact with officers at the foundation.

The data sourced falls into four categories:

1. National Lottery funders, including:
	1. Arts Councils in England, Northern Ireland and Wales, and Creative Scotland
	2. Heritage Lottery Fund
	3. Nesta
	4. Sport England, Sport Northern Ireland and Sport Wales
	5. The Big Lottery Fund

In some cases the data involves funds distributed by these bodies that are originally from government funds, but the majority is distributed from Lottery funds.

1. Grant making foundations publishing as part of 360 Giving. Examples include Northern Rock Foundation, Wellcome Trust and Esmee Fairburn Foundation. As an evolving project the 360 giving data has gone through a number of iterations, and so some of the data is no longer made available in the main 360 giving website.
2. Grant making foundations providing data directly to the project, including Lloyds Bank Foundation for England and Wales.
3. Data from grant making foundations sourced from the organisation’s websites or published reports. This category includes Garfield Weston Foundation and Arcadia.

The 360 giving initiative also includes some additional data from public sector funders such as Trafford Council and the Technology Strategy Board, which have not been included here.

The source of data presents a significant dataset from some of the most important non-government institutional grant makers in the UK. However, the nature of the data means it is a partial picture based on those organisations that have chosen to come forward, and any analysis of the data needs to bear this in mind.

The Big Lottery Fund provides a special case within the dataset. The size of the fund (over £7 billion in this dataset distributed, covering 2004 to 2015) and its significance as a funder to the sector, means that it has been looked at separately in our analysis.

In all, data from 32 funders (including 10 lottery funders) is included in the dataset, representing 291,000 grants worth £22.8 billion (265,000 and £16.7 billion of which are lottery funds). The data includes grants from 1995 to 2015, with most grants appearing in the year 2014. The largest single grant is for £214 million, with the smallest at less than £100. The mean grant size is £76,000 and the median £8,300.

In the course of the project we constructed two files - one is called “grantmakers-data” and contains line by line information about individual grants made by funders; the other is “grantmakers-reconciliation” and embodies our best efforts to match the lists of organisations who have received funding against lists of organisations on the register of chariites and other registers. We have linked the two and created one file, through the common ID field which provides a unique reference to each grant.

# Data import and cleaning

The data import process started by downloading or otherwise receiving the data from the source, and storing in a folder. In general the data was received in CSV format, with some files in .xlsx (Microsoft Excel) format. 360 Giving defines a data standard[[1]](#footnote-1) which is recommended for use by grant makers. This standard allows for two data formats – a complex relational format consisting of linked data tables, and a simpler “flat file” format with a condensed data structure. All the files followed the latter format, although adherence to the standard (in terms of field names, data types, etc) were not always followed.

After the files were downloaded and assembled the csvstat[[2]](#footnote-2) utility was used to examine the files and produce basic descriptions of the field names and types, and the scope of the data in each. These statistics then informed the data import process. Data was imported into a database table from each of the files. The database table is specified to cover most of the commonly used (and needed for analysis) fields, although not all fields are equally covered by all datasets.

For some datasets more extensive work was needed before they were suitable for import. This includes extracting data from PDF reports and web scraping techniques.

Data import was then performed using a series of iPython notebooks[[3]](#footnote-3) using the Pandas data analysis library[[4]](#footnote-4). Each notebook imports the data from the CSV file, maps the fields from the CSV file to the format of the database table, and runs any corrections needed to get the data into the correct format. These corrections include manipulating date formats, correcting character encoding and removing commas from numeric values.

While the fields were generally consistent (and based on the 360 giving standard), in some cases they varied substantially from the standard, and some organisations provided a minimum of information. Figure X below shows the coverage of fields in the data, with the funded organisations shown from those with most fields covered to least, and the fields ordered from the most used on the left to least on the right. A coloured square indicates that there are non-blank/null values present for that fields for that funder. Note that the presence of a field does not mean it is comprehensively filled in for every field in the data.



The figure shows that data for a core set of fields is available for all funders – including the most important information for data analysis (recipient organisation, amount, award date).

The data also varies in the award dates covered. The figure below shows the years that each grant maker has provided data for, with years highlighted where a grant maker has more than one award labelled with that year. Data for 2014 is available for 29 of the 32 funders.



# Matching of organisations

One key part of cleaning the data was to match grant recipients to organisations. This was done through a process of “reconciliation” with registers of organisations, particularly the Charity Commission for England and Wales. The first part of the process is to look at instances where a recipient has an existing registration number (for example a charity number) which could be checked against a list of valid numbers. Most numbers matched correctly. Invalid charity numbers tended to happen for a number of reasons:

* The number was incorrectly recorded
* The number related to a register which was not available for validation – for example the Northern Irish register of charities, or HMRC charity numbers.

The process then turned to matching based on the names of recipient organisations. The problem could sound simple, but is made much more complicated by organisations using more than one name, through trading names, acronyms, old names, misspellings, etc.

In general, the solution to this problem takes two parts:

* First, identifying which are the relevant entities on the list of names to be matched. In this case that would mean identifying whether each organisation is a charity or not. While for charities this step is not likely to be separated from the second step (as there aren’t generic patterns to charity names that separate them from other types of organisations), it could be used for other sorts of organisations (for example if you were matching public sector bodies you could identify likely organisations as those with “County Council” at the end of the name, without matching them to a specific entity.
* Second, matching each record to a specific entity. This usually involves a lookup to an external database, along with some normalisation of the records on both sides. A confidence score can be attached to the match, and it can be automatically accepted or checked manually.

In part the difficulty of matching and the approach taken depends on the nature of the data. A dataset which is known to contain only records of a certain type, where we know that every record represents a charity and all can be matched, is a different proposition and requires a different approach to a dataset where charities are the majority of records but not all of them, on one where only a small proportion of records are likely to match.

When looking at large numbers of records there are also tradeoffs in terms of performance and expectations which determine the approach used. It is often computationally expensive to perform lookups on a large number of records, so the code used needs to be optimised to make this run as quickly as possible. It’s also not usually possible to check large numbers of matched records for success. A project that depends on every match being correct (more likely used for administration purposes) would have to place a high threshold on matching success and so would risk not matching some records, whereas one that wanted to match a large number of records but was more willing to cope with a level of incorrect matches (such as research work) wold have a lower threshold for a match to be accepted.

## Potential methods for matching

There are a great number of potential methods for performing matching between two lists. A less computationally intensive approach would take two lists, normalise the names in some way, and then directly compare them. This either leads to a match or no match for each record. An approach that looks in more detail at each record, perhaps using a number of different ways of matching strings, or looks at the similarity between records, or includes other variables (postcode, website) in the matching process will take longer, but potentially cover more matches. It also has the potential to include a matching score, with higher scores indicating a greater likelihood of a correct match.

## Name normalisation

A name normalisation function is applied to both sets of names to improve the matching process. The function applies the following transformation to each name:

* Convert to lowercase
* Replace “&” with “and”
* Remove any non-alphanumeric characters
* Remove some common words from the string – “the” from the beginning or the end, “limited” and “ltd”, “trust”, “charity”.
* Trim any leading or trailing spaces
* Make sure only single spaces are used to separate words.

The function used also contains some further options that can aid matching:

* Remove any text in brackets
* Reorder the words into alphabetical order
* Replace any spaces

## Example of the problems of matching

To see some of the problems inherent in matching name strings we can look at an example dataset which has already been matched. In this case we will look at the Big Lottery Fund grant records, which includes both an “Organisation Name” string and a charity number for some of the records. This analysis assumes that the charity number given is the correct one, which in reality is not always the case.

We start with 74,539 records which include a valid charity number (one that is present in a list of registered charities), which falls to 44,668 records once duplicates are removed. Of these records, around 48% of records (21,298) match exactly to the name given in their charity record, with no modification or normalisation. A further 17% of records match based on a normalised version of both names, (7,728 records).

This leaves 35% - 15,642 records - where there is a big enough discrepancy between name shown on the list and the official name that an exact match is not found. For these records a score was calculated using the “difflib” python library[[5]](#footnote-5). This score gives a number between 0 and 1 indicating the distance between the two strings: 1 is when the strings are exactly the same, 0 is where there was no similarity at all.

The median similarity score for unmatched records was 0.67 while the mean was 0.63. Looking at records in each of the quartiles gives an indication of the types of problems a matching algorithm would need to overcome. Note that these scores are given based on the raw names, a different score would be produced for the normalised names.

### Upper quartile (score between 0.84 and 0.99)

These records have a very close match, and usually indicate a small typo or spacing issue. They also can contain a one-word addition to a name, for example an acronym in brackets at the end of the name. They may indicate refinements to the name normalisation function that could improve matching there.

| regno | giving | cc\_name | score |
| --- | --- | --- | --- |
| 1068224 | Yeovil Visually Impaired Bowl Club | Yeovil Visually Impaired Bowls Club | 0.99 |
| 1087617 | Clwb Plant Pentraeth Kids Club | CLWB Plant-Pentraeth- Kids' Club | 0.94 |
| 1076890 | 12th Royal Eltham Scout Group | Xiith Royal Eltham Scout Group | 0.92 |
| 1019826 | Chimney Tots Pre-School | Chimneytots Pre-School | 0.98 |
| 1046269 | Lancashire County Childminding Association (Lcca) | Lancashire County Childminders Association | 0.86 |

### 3rd Quartile (score between 0.67 and 0.84)

These records have a larger difference, usually the results of additional words in the name. They also may indicate where a local project name is listed as the recipient, but the charity number belongs to a larger national charity.

|  |  |  |  |
| --- | --- | --- | --- |
| regno | giving | cc\_name | score |
| 304858 | Barrow Village Hall - Play Area | Barrow Village Hall | 0.76 |
| 304041 | Potter Heigham Playing Field and Village Hall | Potter Heigham Playing Field | 0.77 |
| 201470 | Plymtree Parish Hall & Recreation Ground Committee | The Plymtree Parish Hall and Recreation Ground Charity | 0.83 |
| 290427 | Ravensbourne Link Sitter Service | Ravensbourne Link | 0.69 |
| 251468 | The Royal British Legion Women's Section Laceby & District | The Royal British Legion Women's Section | 0.82 |

### 2nd Quartile (score between 0.44 and 0.67)

These records seem to be smaller names, where a small difference can produce a much lower similarity score. Some more drastic differences in the name appear now. Even at this level there are some indications of how a normalisation algorithm could be improved.

|  |  |  |  |
| --- | --- | --- | --- |
| regno | giving | cc\_name | score |
| SC014223 | CVS Aberdeenshire - Central And South | Aberdeenshire Voluntary Action | 0.60 |
| 801554 | Priorswood Playgroup | Priorswood Pre-School | 0.63 |
| 518985 | K.I.N.D | K I N D | 0.57 |
| 1094592 | Dallow Pre-School | Dallow Community Nursery | 0.44 |
| 1052319 | Frontline Matlock | Dales Christian Centre - Matlock | 0.49 |

### Lower quartile (score between 0.03 and 0.44)

In this quartile the names are likely to be very different from each other, sharing very few words or letters in common. They may indicate a mis-match, but also where a trading name is very different from the registered name.

|  |  |  |  |
| --- | --- | --- | --- |
| regno | giving | cc\_name | score |
| 300923 | Newquay Zoo | South West Environmental Parks Limited | 0.16 |
| 1149000 | Woodside Primary Academy | Reach2 Limited | 0.26 |
| 523250 | Hornsea Tennis Club | The Hollis Recreation Ground | 0.30 |
| 1008531 | Honeylands Sibling Group | Happy Bees Nursery | 0.29 |
| 1089477 | Karma Nirvana Limited | Karma Nirvana Peace and Enlightenment Project for Asian Men and Women | 0.42 |

# Coverage of data

While the data covers a significant amount of grant making activity, it is not fully comprehensive, and numerically only a small number of funders are included. Taking the year 2014 as a benchmark, the non-lottery funders represent around £500 million of grant making in this dataset. In crude terms, comparing this figure to the estimate from the NCVO UK Civil Society Almanac of total grant making of £4.6 billion would suggest that the data covers just over 10% of total grant making.

However, the picture is more complicated than that. Looking from the perspective of data from grant makers, the Wellcome Trust accounts for 75% of the grants made by amount, and largely makes grants to universities, so the proportion of grants to charities is lower. But the £4.6 billion amount also includes a wide variety of grant making activities, not all of which are relevant to our analysis. In particular in its original form the data includes grants to individuals; we remove these from the data.

Defining a grant making foundation is not a simple exercise. Data from the Charity Commission for England and Wales provides a number of ways of determining whether an organisation is a grant maker. Organisations can choose “makes grants to organisations” or “makes grants to individuals” as a description of their activities. Organisations with income greater than £500,000 also have to return the “part B” of their annual return, which includes a figure for the amount of grants they made in that year. However, these fields cover a wide variety of activity, not all of which meets a traditional definition of grant making. For example, it includes some transfers between branches of federated charities. This project is focused on grants to UK charities, and many grant makers do not have this as a focus – instead they may focus on grants to universities or individual researchers, or grants to individuals, or they may operate entirely overseas.

Perhaps a more appropriate starting point is the 300 grant makers described in the Association of Charitable Foundation’s “Giving Trends” report[[6]](#footnote-6). These organisations are defined in the report as “independent charitable foundations”, and have an income of £2.5 billion. The Wellcome Trust is also present in this definition, accounting for 20% of the total grant making of these organisations. Eleven of these top 300 grant makers are in our dataset, although eight of these appear in the top 25 grant makers. Including the Wellcome Trust 29% of grant making expenditure is held in our dataset, 9% excluding Wellcome. The top 25 organisations from the ACF report are shown below.

| Name | Charity Number | Giving (£m) | In dataset |
| --- | --- | --- | --- |
| Wellcome Trust | 210183 | 487.7 | Yes |
| Comic Relief | 326568 | 103.1 |  |
| Children's Investment Fund Foundation | 1091043 | 68.6 |  |
| Garfield Weston Foundation | 230260 | 53.4 | Yes |
| Leverhulme Trust | 1159154 | 50.5 |  |
| Royal Society | 207043 | 48.4 |  |
| BBC Children in Need Appeal | SC039557 | 43.9 |  |
| Monument Trust | 242575 | 35.2 |  |
| Esmee Fairbairn Foundation | 200051 | 34.5 | Yes |
| Wolfson Foundation | 1156077 | 31.0 | Yes |
| The Lempriere Pringle Trust | 1007102 | 30.8 |  |
| Grace Trust | 257516 | 29.5 |  |
| Clore Duffield Foundation | 1084412 | 29.3 |  |
| Gatsby Charitable Foundation | 251988 | 29.0 | Yes |
| Henry Smith Charity | 254923 | 27.0 |  |
| Nuffield Foundation | 206601 | 23.7 |  |
| Arcadia | [Part of CAF] | 22.5 | Yes |
| Lloyds Bank Foundation for England & Wales | 327114 | 21.9 | Yes |
| Shell Foundation | 1080999 | 20.7 |  |
| City Bridge Trust | 1035628 | 20.0 |  |
| Tudor Trust | 1105580 | 19.8 |  |
| Leukaemia & Lymphoma Research | 216032 | 19.6 |  |
| Vodafone Foundation | 1089625 | 19.5 |  |
| Ahmadiyya Muslim Jamaat International | 1102949 | 19.5 |  |
| Paul Hamlyn Foundation | 1102927 | 19.4 | Yes |

# Eligibility Criteria

When analysing those organisations that have received funding, it is important to have an appropriate comparison set. Comparing organisations who receive funding from a particular source to the charity population as a whole would not be appropriate – in most cases funding is targeted at particular activities or types of organisation. In order to provide an appropriate comparison set, the eligibility criteria of organisations have been researched, and then operationalised to filter a set of organisations.

Researching the eligibility criteria involves visiting organisation’s website and their grant making documents, and identifying statements made by them as to the criteria they apply for their grants. In some cases no such statements were available – particularly for smaller grant makers who don’t solicit applications but instead pick fundees themselves – but the majority do. These statements were gathered into one document.

The statements were then “operationalised”. This involves turning the textual information into a query that can be applied to a list of charities. Not all criteria can be applied in this way – for example subjective judgements about a charity or a criteria that isn’t represented in the data – but the ones that could be were coded to the most appropriate categories. Examples of criteria generated include:

* Thematic area of the charity (generally coded to the International Classification of Non Profit Organisations).
* Area of operation – either a particular area within the UK or a country/continent overseas.
* Annual income of the organisation.

The process of applying these criteria is a subjective one, requiring a judgement to be made about the intentions of a grant maker and the most appropriate way to represent this in the database. The data available is also not flawless, and provides a simple representation of an complex organisation. In particular the classification of the organisation’s thematic area is subjective. Registration dates were also applied as criteria, with only organisations registered between the dates of the first and last grant in the dataset included.

Bearing those caveats in mind, criteria were developed for all but seven funders. One of the seven funders is the Big Lottery Fund – in this case a single set of criteria were not used as the wide variety of grant programmes undertaken by the fund means they would not apply. Instead criteria were applied to individual grant programmes within the Fund.

Two measures of success are applied to the criteria that have been developed, by testing against the organisations that have received funding from each grant maker. The first is to see the proportion of those organisations that have received funding that meet the eligibility criteria. With perfect data and no subjectivity this would obviously be 100%, but in practice a lower number is expected, and a high figure is likely to indicate that the criteria are too wide and are including unsuitable organisations. The second criteria is the proportion of eligible organisations that have received funding (the reverse of the first). A higher percentage is sought here, although the proportion depends more on the scale of the funder.

The results of these two tests are shown in the table below. Organisations where no criteria were applied are indicated with a star.

| Grantmaker | % of eligible with grants | % of grants eligible | number of eligible orgs |
| --- | --- | --- | --- |
| Staples Trust | 0.0 | 100.0 | 165,983 | \* |
| Tedworth Charitable Trust | 0.0 | 100.0 | 167,236 | \* |
| Three Guineas Trust | 0.0 | 100.0 | 165,033 | \* |
| Wolfson Family Charitable Trust | 0.0 | 100.0 | 161,788 | \* |
| Big Lottery Fund | 12.4 | 98.6 | 221,337 | \* |
| Nominet Trust | 0.0 | 97.7 | 199,653 | \* |
| Nesta | 0.0 | 92.1 | 164,573 | \* |
| ZING | 0.0 | 88.9 | 103,095 |  |
| Macc | 0.3 | 88.2 | 4,746 |  |
| Paul Hamlyn Foundation | 0.1 | 86.0 | 96,889 |  |
| Oxfordshire Community Foundation | 3.2 | 85.9 | 3,488 |  |
| True Colours Trust | 0.0 | 82.1 | 98,994 |  |
| Devon Community Foundation | 4.9 | 77.8 | 3,833 |  |
| Wolfson Foundation | 0.1 | 77.5 | 42,249 |  |
| Lloyds Bank Foundation for England and Wales | 5.5 | 74.0 | 48,545 |  |
| Esmee Fairbairn Foundation | 0.8 | 72.0 | 112,825 |  |
| Northern Rock Foundation | 8.9 | 70.6 | 11,389 |  |
| Arts Council England | 7.0 | 64.9 | 15,356 |  |
| Arts Council Wales | 1.0 | 59.8 | 14,350 |  |
| The Baring Foundation | 0.1 | 53.3 | 18,672 |  |
| Creative Scotland | 0.2 | 51.0 | 13,896 |  |
| Sport England | 1.2 | 49.7 | 6,491 |  |
| Sport Wales | 0.3 | 43.7 | 32,727 |  |
| Heritage Lottery Fund | 3.0 | 43.1 | 58,453 |  |
| The Dulverton Trust | 2.0 | 35.2 | 11,867 |  |
| The Wellcome Trust | 0.1 | 29.2 | 27,353 |  |
| Indigo Trust | 0.2 | 26.9 | 11,274 |  |
| Sport Northern Ireland | 0.0 | 0.0 | 6,216 |  |

In general the results show that the application of these criteria is generally successful. In most cases the proportion of funded organisations that are in the list of eligible organisations is over 50%, with ten organisations having over 70%. There are also positive figures in the proportion of organisations funded – 9% of eligible organisations for the Northern Rock Foundation have received funding, 5% for the Lloyd’s Bank Foundation and 7% for the Arts Council England.

# Organisation Types

Each grant was also assigned an organisation type. For some funders there was an organisation type supplied with the data. In the remaining cases each organisation needed to be assigned an organisation type.

The categories used were based on those used by the Big Lottery Fund. As the funder with the largest amount of grants, and whose data was already categorised, they provide a useful based for classifying other organisations. The classification categories used are as follows:

| Type | Subtype | Grant count |
| --- | --- | --- |
| Charity | Charitable Incorporated Organisation | 1,183 |
| Charitable Trust | 49 |
| Charitable Unincorporated Association | 44,562 |
| Charity (Royal Charter or Act of Parliament) | 20 |
| Excepted Charity | 7 |
| Exempt Charity | 793 |
| Registered Charity | 109,991 |
| Company/Mutual Society | CIC - Limited by Guarantee | 1,415 |
| CIC - Limited by Shares | 32 |
| CIC - Listed Publicly | 51 |
| Company - Limited by Guarantee | 2,873 |
| Company - Limited by Shares | 189 |
| Company - Listed Publicly | 17 |
| Co-operative - unincorporated | 4 |
| Credit Union | 10 |
| Friendly Society | 5 |
| Housing Association | 120 |
| Industrial & Provident Society | 271 |
| Limited Liability Partnership | 2 |
| Public Sector | Community Council | 298 |
| Fire Service | 1 |
| Health Authority | 7 |
| Local Authority | 12,327 |
| NHS Trust - Foundation | 127 |
| Non-Departmental Public Body | 478 |
| Other | 537 |
| Parish Council | 2,614 |
| Police Authority | 2 |
| Town Council | 544 |
| School | Academy | 444 |
| City Technology College | 1 |
| Community School | 1,562 |
| Foundation or Trust School | 248 |
| State School | 16,239 |
| Voluntary Aided School | 566 |
| Voluntary Controlled School | 268 |
| Other | Church-based faith organisation | 42 |
| Community Amateur Sports Club | 1,379 |
| Further / Higher Education | 515 |
| Independent School | 55 |
| Individual | 38,301 |
| Non charitable unincorporated organisation | 8,285 |
| Other | 1,124 |
| Parochial Church Council | 3,610 |
| Partnership | 12 |
| Sole Trader | 1 |
| Trade Union | 2 |
| University | 14,070 |

These categories are not perfect, as categorising organisations in this way is always difficult due to the myriad overlapping forms that organisations can choose to take. While the categories are mainly based on legal statuses, there are some that are more “thematic” – for example a church-based faith organisation. Some categories overlap – registered charities can also be registered as companies limited by guarantee, many Parochial Church Councils are also registered charities.

Some funders provided data that showed the organisation type of their recipients. While this data did not use the same categories as above, it was generally possible to provide a crosswalk between the categories in order to re-classify them into these categories.

The remaining organisations needed to be identified, generally through keyword searching. A particular difficult was identifying individuals within the data so that they could be excluded. Arts funders in particular make lots of grants to individuals, and it is useful to have these separately identified. Individuals were identified through a manual process of looking line-by-line at grants made by Arts funders. We developed techniques to increase the likelihood of developing a match – for example looking in particular at those grants with two words in the recipient name. Once this process was complete the grants that were confirmed as either to an institution or an individual could be used to construct a training dataset that has been turned into a naïve Bayesian classifier. This classifier can now be run automatically against a dataset to produce a useful result.

The remaining grants were classified according to keyword searches on a variety of terms to produce a result for many of the grants. The reconciliation process was also used – where a grant recipient had been reconciled to a charity record it was put into the “Registered Charity” category.

The final result was that around 35,000 grants out of 300,000 were without an organisation type. The organisations that were classified could then be used as training data for another naïve Bayesian classifier. However, this classifier was less successful than the one which separates individuals and institutions. Because of the wide range of terminology, and because organisation type is often not related to the words in an organisations’ name, the results produced were often not satisfactory.

NOTE: 22,000 records for individual named grant recipients were removed as they named the individuals.

# Appendix I – database table specification: file “grantmakers-data.csv”

| Column Name | Data type | Description |
| --- | --- | --- |
| id | Text | Unique Identifier for column |
| title | Text | Title of the grant |
| description | Text | Description of the project |
| fundingorganization\_id | Text | Unique identifier for funder |
| fundingorganization\_name | Text | Name of the funder |
| recipientorganization\_id | Text | Unique identifier for recipient |
| recipientorganization\_name | Text | Name of recipient |
| recipientorganization\_type | Text | Organisation type of the recipient |
| currency | Text | Currency of the amount fields |
| totalamountappliedfor | Number |  |
| totalamountawarded | Number |  |
| recipientorganization\_charitynumber | Text | Registered charity number provided with the data |
| recipientorganization\_companynumber | Text | Registered company number provided with the data |
| applicationdate\_startdate | Date |  |
| awarddate\_startdate | Date | Date that grant award was made |
| planneddates\_startdate | Date | Date of start of grant activity |
| planneddates\_enddate | Date | End date of grant activity |
| recipientorganisation\_streetaddress | Text |  |
| recipientorganisation\_city | Text |  |
| recipientorganisation\_county | Text |  |
| recipientorganisation\_postalcode | Text |  |
| recipientorganisation\_country | Text |  |
| recipientorganisation\_alternatename | Text |  |
| planneddates\_duration\_months | Number | Duration (in months) of grant activity |
| beneficiarylocation\_name | Text | Location of grant beneficiaries |
| webaddress | Text | Web address of recipient organization |
| charity\_matched | Text | Matched registered charity number |
| company\_matched | Text | Matched registered company number |
| grantprogramme | Text | Programme/stream within funder's activities |

# Appendix II – Data sources

|  |  |
| --- | --- |
| Funder name | Data source |
| Arcadia | Scraped data from <http://www.arcadiafund.org.uk/grants/grant-directory.aspx>  |
| Arts Council England | 360 Giving |
| Arts Council of Northern Ireland | 360 Giving |
| Arts Council Wales | 360 Giving |
| Creative Scotland | Website list of funding organisations: <http://www.creativescotland.com/funding/latest-information/awards-listings>  |
| Devon Community Foundation | Directly provided by the foundation |
| Esmee Fairburn Foundation | 360 Giving |
| Garfield Weston | Extracted from 2014/15 Accounts |
| Gatsby Charitable Foundation | 360 Giving |
| Heritage Lottery Fund | 360 Giving |
| Indigo Trust | 360 Giving |
| Lloyds Bank Foundation for England and Wales | Directly provided by the foundation |
| Macc | 360 Giving |
| Nesta | Available on Nesta website: <http://www.nesta.org.uk/about-us/how-we-spend-our-money>  |
| Nominet Trust | 360 Giving |
| Northern Rock Foundation | 360 Giving |
| Oxfordshire Community Foundation | 360 Giving |
| Paul Hamlyn Foundation | 360 Giving |
| Sport England | 360 Giving |
| Sport England [Lottery] | 360 Giving |
| Sport Northern Ireland | 360 Giving (original data) |
| Sport Wales | 360 Giving (original data) |
| Staples Trust | 360 Giving |
| Tedworth Charitable Trust | 360 Giving |
| The Baring Foundation | 360 Giving |
| The Big Lottery Fund | Directly provided by the foundation (also available on 360 Giving) |
| The Dulverton Trust | 360 Giving |
| The Wellcome Trust | 360 Giving |
| Three Guineas Trust | 360 Giving |
| True Colours Trust | 360 Giving |
| Wolfson Family Charitable Trust | 360 Giving |
| Wolfson Foundation | 360 Giving |
| ZING | 360 Giving |

1. <http://www.threesixtygiving.org/standard/> [↑](#footnote-ref-1)
2. <http://csvkit.readthedocs.org/en/540/scripts/csvstat.html> [↑](#footnote-ref-2)
3. An iPython notebook is a collection of code and explanatory text in an easy-to-use format <http://ipython.org/notebook.html> [↑](#footnote-ref-3)
4. <http://pandas.pydata.org/> [↑](#footnote-ref-4)
5. <https://docs.python.org/2/library/difflib.html#difflib.SequenceMatcher> [↑](#footnote-ref-5)
6. <http://www.acf.org.uk/policy-practice/research-publications/giving-trends-top-300-foundation-grant-makers-2015-report> [↑](#footnote-ref-6)