

Procurement Payments

A Guide to the Dataset

August 2026

1. Overview

This dataset provides records of public procurement payments made to civil society organisations — including registered charities, community interest companies, co-operatives, and other third sector bodies — across the United Kingdom. It was produced by the [UK Third and Civil Society Sector Database \(https://uk-third-sector-database.github.io/\)](https://uk-third-sector-database.github.io/) project, which collects, processes, and links public administrative data on civil society organisations throughout the UK.

The data is drawn from three open data sources: central government transparency spending data, NHS payment records, and Contracts Finder awarded contract notices. Each payment record links a public sector funder to a civil society supplier, providing a detailed picture of how public money flows to the third sector through procurement and commissioning.

The dataset contains 883,816 payment records covering 11,886 organisations and 1,503 public funders, spanning the period from 2010 to 2025 and totalling £148.4 billion. Together, these records offer a comprehensive view of the scale, distribution, and evolution of public procurement relationships with civil society across the UK.

883,816

Payment Records

11,886

Organisations

1,503

Public Funders

2010–2025

Years Covered

Related dataset. This guide covers the **payments** dataset (`cso-procurement-dataset-refined.csv`): one row per payment or award made to a register-confirmed civil society organisation. The project also publishes a companion **Contracts Finder** dataset (`contractsfinder-enriched-refined.csv`) — one row per awarded contract notice on Contracts Finder, covering contracts awarded to *all* suppliers with civil society organisations identified by a register match, plus the contract title, description, sector codes, estimated and awarded values and dates. Use the payments dataset to measure money paid to civil society over time; use the Contracts Finder dataset to study the contract market (what was tendered, who won, civil society's share) and to test the buyer's voluntary-sector flags. The Contracts Finder dataset has its own guide: [Contracts Finder Records — A Guide to the Dataset](https://uk-third-sector-database.github.io/guidance/tcss-contracts-finder-guidance.html) (<https://uk-third-sector-database.github.io/guidance/tcss-contracts-finder-guidance.html>).

2. What are Procurement Records?

UK public bodies are required to publish details of their spending over certain thresholds as part of government transparency commitments. This dataset draws from three open data initiatives that collect and standardise these spending records, filtered to include only payments made to civil society organisations identified in the TCSS Organisation Register.

Central Government Spending

UK government departments publish monthly CSV files of all transactions exceeding £25,000 as part of HM Government's spending transparency commitments. These files are collected and harmonised by the [centgovspend](https://github.com/crahal/centgovspend) (<https://github.com/crahal/centgovspend>) project, which aggregates and cleans thousands of files from ministerial and non-ministerial departments for consistency and quality controls. Central government spending accounts for 90.7% of the records in this dataset.

NHS Spending

Payment records from NHS Trusts and Clinical Commissioning Groups (CCGs) are collected by the [NHSSpend](https://github.com/crahal/NHSSpend) (<https://github.com/crahal/NHSSpend>) project. This covers payments exceeding £25,000 made by NHS institutions across England, spanning January 2010 to March 2020 when data collection concluded. NHS spending accounts for 7.2% of records.

Contracts Finder

[Contracts Finder](https://www.contractsfinder.service.gov.uk/) (<https://www.contractsfinder.service.gov.uk/>) is the UK government's online portal for public sector procurement opportunities and awarded contracts. Awarded contract notices are scraped by the TCSS project's own [Contracts Finder collector](https://github.com/uk-third-sector-database/contracts_finder) (https://github.com/uk-third-sector-database/contracts_finder). Unlike the payment-level data from the other two sources, Contracts Finder provides contract-level information (awarded amounts and dates), so each Contracts Finder row in this dataset is an awarded contract value rather than a cash payment. These records account for 2.1% of the dataset (18,184 rows). The full contract-level data, including contracts awarded to non-civil-society suppliers, is published separately — see the [Contracts Finder Records guide](https://uk-third-sector-database.github.io/guidance/tcss-contracts-finder-guidance.html) (<https://uk-third-sector-database.github.io/guidance/tcss-contracts-finder-guidance.html>).

3. Dataset Contents

The dataset is supplied as a single CSV file, `cso-procurement-dataset-refined.csv` , containing 883,816 records across 18 fields. Each row represents a single payment from a public sector funder to a civil society organisation. The fields are organised into four groups: identifiers, payment details, funder information, and summary statistics.

Identifiers & Organisation Info

Field	Description	Type	Coverage
uid	Unique organisation identifier from the TCSS Organisation Register (e.g., GB-COH-12345678 , GB-CHC-1234567 , GB-SC-SC012345)	Text	100%
organisation_name	Organisation name as recorded in the source data	Text	91.6%
supplier_id	Internal supplier identifier assigned during pipeline processing (e.g., S192664)	Text	100%
supplier_type	Organisation type: Charity , CIC , Co-operative/Mutual , or Other CSO	Text	100%
supplier_type_detail	A more granular supplier classification, refining supplier_type	Text	100%
is_multi_supplier	Indicates whether the source record named more than one supplier (True / False)	Boolean	100%

Payment Details

Field	Description	Type	Coverage
payment_year	Calendar year of the payment	Numeric	100%
payment_date	Date of payment (YYYY-MM-DD format)	Date	100%
amount	Payment amount in pounds sterling (£); negative values indicate reversals or corrections	Numeric	100%
data_source	Source dataset: Central Government , NHSSpend , Contracts Finder , or contractsfinder	Text	100%

Funder Information

Field	Description	Type	Coverage
funder_name	Canonical funder name (uppercase)	Text	{S['cov_funder_name']}
funder_name_alt	Alternative funder name, where applicable	Text	1.0%
funder_id	Unique funder identifier (e.g., F02683). Missing (together with funder_name) for 16 records whose funder is classified Junk/Invalid	Text	>99.9%
funder_type	Funder classification: UK Government , NHS , Local Government , Education Institution , CSO , Police, Fire and Rescue , Private Sector , Other , or Junk/Invalid (see Appendix A2)	Text	100%
funder_type_detail	A more granular funder classification, refining funder_type (for example, distinguishing types of body within UK Government)	Text	100%

Field	Description	Type	Coverage
funder_type_alt	Classification of the alternative funder name	Text	1.4%

Note: The `funder_name_alt` and `funder_type_alt` fields are populated only where an alternative funder name was assigned during the classification pipeline. This primarily applies to Contracts Finder funders that were manually refined to their parent Central Government department (e.g., DVSA mapped to DFTRANSPORT).

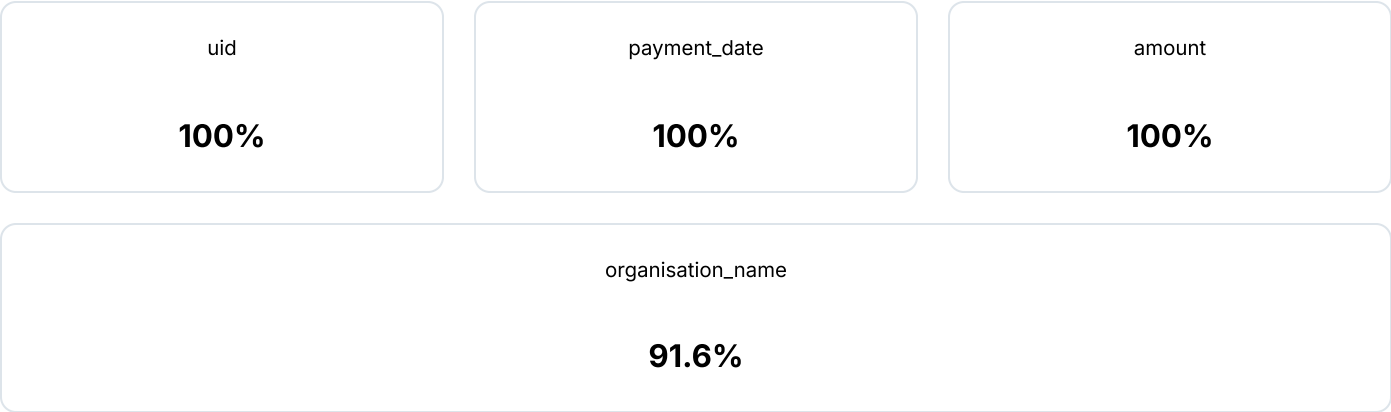
Summary Statistics

Field	Description	Type	Coverage
total_value_payments_to_org	Total value of all payments to this organisation across the full dataset (£)	Numeric	100%
total_number_payments_to_org	Total number of payments to this organisation across the full dataset	Numeric	100%

4. Coverage & Completeness

Core Field Coverage

The core payment fields — `payment_date` , `amount` , `supplier_id` and `data_source` — are present in every record; `funder_name` and `funder_id` are missing for just 16 records, all classified Junk/Invalid. Every record carries a `uid` — from this release a payment is included only if its supplier is a register-confirmed civil society organisation — and `organisation_name` is present for 91.6% of records (where missing, recover it from the Organisation Register on `uid`).



Data Source Breakdown

The dataset draws from four data source tags. Central government spending dominates, accounting for 90.7% of all records. Contracts Finder rows (2.1% combined) are awarded-contract values rather than payments.

Source (<code>data_source</code> value)	Records	Unique Suppliers	Share
Central Government	802,029	7,856	90.7%
NHSSpend	63,603	2,141	7.2%
Contracts Finder	14,658	4,293	1.7%
contractsfinder	3,526	1,679	0.4%

Note: The `data_source` field distinguishes two Contracts Finder collection batches — `Contracts Finder` and `contractsfinder` — both drawn from the same portal. The “Unique Suppliers” column counts distinct `supplier_id` values within each source; because a supplier may appear in more than one source, these figures do not sum to the dataset total.

Funder Type Distribution

Funders are classified into nine types (eight substantive types plus a Junk/Invalid catch-all) using a multi-layer cascade (see [Appendix A2](#)). UK Government departments account for the vast majority of records (91.4%).

Funder Type	Records	Distinct Funders	Share
UK Government	807,596	190	91.4%
NHS	65,281	517	7.4%
Local Government	8,612	331	1.0%
Education Institution	891	136	0.1%
CSO	424	138	<0.1%
Police, Fire and Rescue	356	72	<0.1%
Private Sector	303	55	<0.1%
Other	211	48	<0.1%
Junk/Invalid	142	16	<0.1%

Supplier Type Distribution

Civil society organisations in the dataset are classified by legal form. “Other CSO” includes companies limited by guarantee and other third sector bodies that do not fall into the three specific categories; the `supplier_type_detail` field shows that most of these records (95% of Other CSO rows, 65% of the whole dataset) are payments to education institutions — academy trusts, colleges and universities constituted as nonprofit companies — which should be borne in mind when interpreting totals.

Supplier Type	Records	Distinct Organisations	Share
Other CSO	606,603	3,955	68.6%
Charity	227,745	6,635	25.8%

Supplier Type	Records	Distinct Organisations	Share
Co-operative/Mutual	32,025	467	3.6%
CIC	17,443	843	2.0%

Year-by-Year Record Counts

Record counts vary substantially by year, reflecting the availability of source data over time. Coverage is strongest between 2012 and 2023.

Year	Records
2010	5,274
2011	8,085
2012	51,936
2013	115,708
2014	163,872
2015	53,517
2016	30,321
2017	64,876
2018	63,030
2019	66,280
2020	60,566
2021	57,887
2022	62,960
2023	66,690
2024	11,400
2025	1,380

Note: 34 records fall outside the 2010–2025 range shown above: 7 records dated 2005–2009 that predate the main collection period, 24 records dated 2026 from ongoing collection, and 3 records with implausible years (1900, 2027) caused by source data errors. All are retained in the dataset for transparency; users conducting temporal analysis may wish to filter to the core 2010–2025 range. The 2024 and 2025 figures are also incomplete as source data collection is ongoing.

Organisation Register Composition

The `uid` prefix indicates which source register each organisation originates from in the TCSS Organisation Register. Companies House registrations dominate, reflecting the large number of companies limited by guarantee and other nonprofit company forms. Every record in this release carries a `uid`.

Change from the June 2026 release. Earlier releases also contained 62,175 records (£32.9 billion, 18.2% of value) whose supplier carried a civil society `supplier_type` from the project’s earlier name-based classification but no register match; on inspection some of the largest (for example a Scottish health board typed as a charity) were not civil society organisations. From this release the rule is *Spine-only*: a supplier is a civil society organisation only if it matches the TCSS Organisation Register (or carries a manually confirmed register identity), so those rows are no longer included and every row has a `uid`. Suppliers demoted by this rule that are in fact civil society organisations will be restored in future releases as they are matched to the Register.

Prefix	Source Register	Records	Share
GB-COH	Companies House	623,380	70.5%
GB-CHC	Charity Commission for England & Wales	202,287	22.9%
GB-SC	Office of the Scottish Charity Regulator	25,477	2.9%
GB-COOP	Co-operatives UK	23,253	2.6%
GB-MPR	Mutuals Public Register (FCA)	8,772	1.0%
GB-SHR	Scottish Housing Regulator	352	<0.1%
GB-NIC	Charity Commission for Northern Ireland	294	<0.1%
GB-SHPE	Social Housing England	1	<0.1%

5. What Can You Learn?

The procurement dataset enables a wide range of research questions about the relationship between the public sector and civil society in the United Kingdom.

Research Questions

- **Scale and trends** — How much does the UK government spend with civil society organisations, and how has this changed over time?
- **Funder analysis** — Which government departments and NHS bodies are the largest funders of civil society? How does spending vary across funder types?
- **Sectoral composition** — What types of civil society organisations receive the most public procurement funding — charities, community interest companies, or co-operatives?

- **Concentration** — How concentrated is procurement spending? Do a small number of organisations receive the majority of payments?
- **Cross-sector linkage** — By linking to the TCSS Organisation Register, researchers can explore how procurement recipients differ from the broader civil society population in terms of size, age, location, and industrial classification.

Example: Mapping UK Civil Society Procurement

The procurement dataset was used in the research report *Mapping and Understanding the UK Civil Society Sector* (McDonnell et al., 2026), which analysed the payment records to characterise the flow of public money to civil society organisations. The report worked from an earlier build of the dataset; the figures and tables below have been recomputed from the current release so that they match the file you download. Key findings:

- **UK Government dominates** — UK Government departments are the largest source of public procurement spend to civil society across all organisation types, accounting for 91.4% of records and £123.1 billion of the £148.4 billion recorded — far ahead of the NHS (£12.2 billion) and local government (£9.8 billion). Note that the Department for Education alone accounts for £80.4 billion, most of it paid to academy trusts and other education bodies classified as “Other CSO”.
- **Charities and other nonprofit companies receive most of the money** — charities receive 38% of total value (26% of records) and “Other CSO” suppliers 55% of value (69% of records); community interest companies and co-operatives/mutuals together receive the remainder.
- **Department-level variation** — analysis at the individual department level reveals substantial variation in how much each department spends with different types of civil society organisation, from health-focused charities receiving NHS payments to social enterprises delivering local government contracts.

The table below shows, for each combination of funder type (rows) and civil society organisation type (columns), the number of distinct organisations paid, the number of payment records, and their total value. Records whose funder was classified Junk/Invalid are excluded.

Funder type	Charity	CIC	Co-operative/Mutual	Other CSO
UK Government	4,326 orgs 180,932 payments £41,786.2M	324 orgs 4,060 payments £943.1M	268 orgs 25,723 payments £1,083.4M	3,579 orgs 596,881 payments £79,237.7M
NHS	1,829 orgs 39,353 payments £5,686.2M	268 orgs 12,503 payments £4,530.6M	152 orgs 5,460 payments £920.2M	281 orgs 7,965 payments £1,079.3M
Local Government	2,041 orgs 6,042 payments £6,372.7M	374 orgs 715 payments £865.1M	174 orgs 721 payments £1,812.6M	446 orgs 1,134 payments £723.7M
Education Institution	205 orgs 489 payments £230.6M	19 orgs 27 payments £2.6M	13 orgs 55 payments £7.2M	78 orgs 320 payments £41.1M
CSO	170 orgs 269 payments £466.6M	18 orgs 22 payments £4.0M	14 orgs 25 payments £29.7M	69 orgs 108 payments £59.5M
Police, Fire and Rescue	125 orgs 231 payments £420.7M	22 orgs 33 payments £11.3M	6 orgs 7 payments £2.8M	27 orgs 85 payments £25.8M

Funder type	Charity	CIC	Co-operative/Mutual	Other CSO
Private Sector	111 orgs 185 payments £686.0M	24 orgs 48 payments £713.8M	14 orgs 19 payments £66.1M	37 orgs 51 payments £51.8M
Other	101 orgs 133 payments £357.1M	18 orgs 21 payments £21.6M	9 orgs 13 payments £3.2M	28 orgs 44 payments £17.0M

Table 1: Procurement by funder type and CSO type. Each cell shows the number of organisations, payments, and total value. Source: TCSS Procurement Records (payments), current release.

UK Government departments by CSO type

Table 2 shows the proportion of each UK Government department's procurement spend with civil society that goes to each organisation type, for the 44 departments and bodies with at least £10 million of such spend. Departments vary considerably in where their procurement is directed. The Department for Culture, Media & Sport and the Department for International Development direct over 90% of their civil society spend to charities, while the Department for Education and the Department for Business & Trade allocate 81% and 73% respectively to other nonprofit companies. The Department for Work & Pensions stands out for relatively high CIC (5.8%) and co-operative/mutual (7.9%) shares compared with most other departments.

Department / body	Charity	CIC	Co-operative/Mutual	Other CSO	Total to CSOs
Bank of England	55.4%	0.0%	0.0%	44.6%	£11.5M
Cabinet Office	42.7%	41.4%	0.1%	15.8%	£901.7M
Care Quality Commission	98.8%	0.0%	0.0%	1.2%	£24.4M
Crown Commercial Service	44.4%	0.2%	1.9%	53.5%	£123.6M
Department for Business & Trade	26.1%	0.3%	0.5%	73.1%	£3,222.8M
Department for Culture, Media & Sport	96.0%	2.1%	0.1%	1.8%	£6,164.5M
Department for Education	18.4%	0.1%	0.4%	81.1%	£80,401.5M
Department for Energy Security & Net Zero	11.7%	1.5%	26.6%	60.2%	£61.9M
Department for Environment, Food & Rural Affairs	71.4%	0.9%	1.9%	25.7%	£428.7M
Department for International Development	93.1%	0.1%	<0.1%	6.7%	£5,155.0M
Department for International Trade	9.0%	0.2%	0.0%	90.9%	£72.0M
Department for Science, Innovation & Technology	64.3%	2.5%	0.0%	33.2%	£42.8M
Department for Transport	9.9%	0.2%	1.9%	88.0%	£4,155.2M
Department for Work & Pensions	55.2%	5.8%	7.9%	31.1%	£1,619.3M
Department of Health & Social Care	63.2%	1.7%	0.1%	35.1%	£4,793.6M
Driver & Vehicle Licensing Agency	20.4%	0.0%	0.0%	79.6%	£15.5M
Financial Conduct Authority	98.2%	0.4%	0.0%	1.4%	£12.5M

Department / body	Charity	CIC	Co-operative/Mutual	Other CSO	Total to CSOs
Food Standards Agency	15.0%	0.5%	0.0%	84.5%	£20.3M
Foreign Office	94.0%	<0.1%	0.2%	5.7%	£1,744.1M
Foreign, Commonwealth & Development Office	76.7%	<0.1%	0.0%	23.3%	£522.5M
HM Land Registry	0.3%	5.9%	0.0%	93.8%	£25.3M
HM Revenue & Customs	9.8%	<0.1%	0.2%	89.9%	£201.4M
HM Treasury	6.2%	0.1%	0.3%	93.5%	£167.4M
Health And Social Care Information Centre	97.7%	0.0%	0.0%	2.3%	£18.2M
Health Education England	32.3%	26.7%	<0.1%	41.0%	£169.2M
Highways England	96.1%	0.0%	0.0%	3.9%	£52.0M
Home Office	68.5%	0.6%	0.1%	30.8%	£590.3M
Homes England	95.2%	0.0%	3.3%	1.5%	£41.1M
Ministry of Defence	74.6%	0.7%	5.8%	18.9%	£1,094.2M
Ministry of Housing, Communities & Local Government	54.4%	0.8%	4.8%	40.0%	£784.7M
Ministry of Justice	72.3%	3.3%	3.5%	20.9%	£1,176.4M
Money & Pensions Service	100.0%	0.0%	0.0%	0.0%	£10.8M
NHS Blood & Transplant	2.8%	0.0%	93.8%	3.4%	£23.7M
National Employment Savings Trust (NEST)	98.9%	0.0%	0.0%	1.1%	£20.9M
National Highways	60.5%	1.9%	0.0%	37.6%	£15.3M
Office For National Statistics	94.6%	0.0%	0.0%	5.4%	£15.2M
Public Health England	82.8%	5.3%	0.3%	11.6%	£28.8M
Scotland Office	61.2%	0.2%	3.4%	35.2%	£8,244.2M
Skills Funding Agency	50.0%	0.0%	0.0%	50.0%	£35.6M
Transport for London	50.9%	23.7%	23.8%	1.6%	£110.5M
UK Health Security Agency	1.8%	<0.1%	0.0%	98.2%	£192.0M
UK Research & Innovation	45.8%	19.2%	3.2%	31.7%	£25.5M
UK Shared Business Services	36.9%	6.4%	2.4%	54.3%	£65.6M
Ukstatauth	96.0%	0.0%	0.0%	4.0%	£27.9M

Table 2: UK Government department procurement spend by civil society organisation type — proportion of each department's total spend to civil society (%), departments and bodies with at least £10 million of such spend. Department labels follow the research report; the Department for International Trade appears under a single label after merging two source names. Source: TCSS Procurement Records (payments), current release.

Tip: This dataset can be linked to the [TCSS Organisation Register](https://uk-third-sector-database.github.io/guidance/tcss-organisation-register-guidance.html) (<https://uk-third-sector-database.github.io/guidance/tcss-organisation-register-guidance.html>) using the `uid` field, enabling enrichment with organisation characteristics such as location, registration dates, and industrial classification codes. See [Appendix A5](#) for worked examples.

6. Limitations & Caveats

Threshold Bias

The source data includes only transactions above £25,000, as mandated by UK government transparency requirements. Smaller payments and grants below this threshold are not captured. This means the dataset over-represents larger contracts and under-represents routine smaller purchases, and total spending figures will understate the true volume of public procurement from civil society.

Missing Organisation Names

Approximately 8.4% of records lack an `organisation_name` value. These records are still linked to valid organisations through the `uid` and `supplier_id` fields, but the name was not always present in the source spending data. Users can recover organisation names by joining with the [TCSS Organisation Register](https://uk-third-sector-database.github.io/guidance/tcss-organisation-register-guidance.html) (<https://uk-third-sector-database.github.io/guidance/tcss-organisation-register-guidance.html>), on `uid`.

Year Outliers

A handful of records (3 in total) carry implausible payment years (1900, 2027), resulting from errors in the source data, and a further 7 pre-date 2010. These records are retained for transparency. Users conducting temporal analysis should filter to the core range of 2010–2025, and may also wish to note that the 2024 and 2025 counts are incomplete as source data collection is ongoing.

Duplicate and Reversal Entries

Some records represent payment corrections or reversals, indicated by negative values in the `amount` field (2,563 records), and 2,115 records carry a zero amount. These are retained to preserve the source data faithfully. Users conducting aggregate analysis should be aware that naïve summation of amounts may overstate or understate totals; consider filtering or handling negative amounts depending on the research question.

NHS Coverage Gap

The NHSSpend data collection concluded around April 2020. NHS procurement records after this date are not included in the dataset, creating a gap in NHS-specific coverage for 2020 onwards. Central government and Contracts Finder data continue beyond this date.

Register-Confirmed Membership (Spine-only rule)

From the August 2026 release a payment is included only when its supplier matches the TCSS Organisation Register (every row has a `uid`). Earlier releases also included 62,175 records from 1,131 suppliers (£32.9 billion) typed as civil society by the earlier name-based masterlist without a register match; several of the largest were misclassified public bodies or companies, so totals in earlier releases and in the June 2026 guidance are not comparable with this release. The other side of the rule is that genuine civil society organisations whose names have not yet been matched to the Register are excluded; they will be added as matching improves.

Supplier Matching

Organisations in the raw payment data were matched to the TCSS Organisation Register using a combination of exact and fuzzy name matching (Jaro-Winkler similarity, threshold ≥ 0.90). Some false positives (incorrect matches) and false negatives (missed matches) are possible, particularly for organisations with common or ambiguous names. The alias resolution process (see [Appendix A3](#)) mitigates but does not eliminate this issue.

Funder Classification

Funders are classified using a multi-layer cascade of metadata signals, external lookups, and keyword rules (see [Appendix A2](#)). While the cascade achieves high accuracy for well-known funder types (NHS, UK Government), edge cases — particularly funders with ambiguous names or those not present in external reference databases — may be misclassified. The `Junk/Invalid` category captures clearly erroneous funder names (16 distinct names, 142 records).

What's NOT in the Data

The dataset does not include:

- Contract descriptions or service categories (available for Contracts Finder awards in the companion [Contracts Finder Records](#) (<https://uk-third-sector-database.github.io/guidance/tcss-contracts-finder-guidance.html>), dataset)
- Geographic detail of the contract delivery location
- Payments below £25,000
- Payments to organisations not identified in the TCSS Organisation Register
- Procurement from non-civil-society suppliers (private companies, individuals, etc.) — the Contracts Finder Records dataset does include these, for awarded contracts only

7. Citation & Licence

Licence: This dataset is licensed under the [Creative Commons Attribution 4.0 International License](#) (<https://creativecommons.org/licenses/by/4.0/>) (CC BY 4.0). You are free to share, adapt, and build upon this data for any purpose, provided you give appropriate credit.

Suggested Citation

McDonnell, D. et al. (2026). *TCSS Procurement Records (Payments)*. UK Third and Civil Society Sector Database. Available at: <https://uk-third-sector-database.github.io/data/> (<https://uk-third-sector-database.github.io/data/>). Licensed under CC BY 4.0.

If you would like to learn more about this dataset and how it can be applied to your project or research programme, please contact research@brawdata.com (<mailto:research@brawdata.com>).

8. Changelog

Version	Date	Changes
1.3	August 2026	August 2026 data release (883,816 records, 11,886 organisations). Supplier classification now follows the Spine-only rule: a supplier is a civil society organisation only if matched to the TCSS Organisation Register, so the 62,175 previously included records without a <code>uid</code> (£32.9 billion) are removed and every row now carries a <code>uid</code> ; 13 large charities whose register identities were recorded as compound identifiers (e.g. British Red Cross, RNIB, Sense, NatCen) are now correctly classified and included. All figures and tables re-derived. Caution notes about unmatched rows replaced by a description of the rule; the line break inside the <code>Central Government / Contracts Finder</code> source labels is fixed in the data. Supplier identifiers (<code>supplier_id</code>) are re-minted each release; use <code>uid</code> to link across releases.
1.2	August 2026	Guidance split into two standalone documents: this guide (payments dataset) and a new Contracts Finder Records guide (https://uk-third-sector-database.github.io/guidance/tcss-contracts-finder-guidance.html). Dataset unchanged from the June 2026 release at that point. All figures re-derived from the published file; corrected the source-share percentages in Section 2, the share of records without an organisation name and the counts of implausible years and Junk/Invalid records in Section 6, and replaced the two tables in Section 5 with versions computed from the current data. Documented the line break in the <code>data_source</code> values, the Private Sector funder type, and the unmatched CSO-typed suppliers (no <code>uid</code>) that inflate value totals. Added a “related dataset” note and a distinct suppliers column to the source table.
1.1	June 2026	Refreshed the payments dataset (944,449 records) and refreshed all coverage and distribution figures. Added documentation for the companion contract-level Contracts Finder dataset (now in its own guide).
1.0	March 2026	Initial release of the Procurement Records guidance document and dataset.

A1. Pipeline Overview

The procurement dataset is produced by a six-step R preprocessing pipeline that transforms raw spending data into the final linked dataset. The pipeline is orchestrated by a single script (`run-all-preprocessing-pipeline.R`) that runs each step in sequence, skipping steps whose inputs have not changed since the last successful run.

Pipeline Steps

Step	Script	Purpose
1	<code>02-build-funder-lookup.R</code>	Classifies the 1,503 unique funders into nine types (eight substantive types plus Junk/Invalid) using a multi-layer cascade of metadata signals, external lookups, and keyword rules. See Appendix A2 .
2	<code>03-build-supplier-lookup.R</code>	Matches supplier names from the raw payment data to the TCSS Organisation Register using exact and fuzzy name matching (Jaro-Winkler, threshold ≥ 0.90). Assigns a <code>uid</code> and <code>supplier_type</code> to each matched supplier.
3	<code>03a-generate-alias-batches.R</code>	Identifies potential supplier name aliases — cases where the same organisation appears under different names — using deterministic rules. Generates review batches for manual or LLM-assisted validation.
4	<code>03b-assemble-alias-decisions.R</code>	Assembles alias decisions from both rule-based determinations and LLM-reviewed batch results into a single validated alias file.
5	<code>03c-apply-alias-merges.R</code>	Applies validated alias merges to the supplier lookup, consolidating duplicate supplier entries under a single canonical <code>uid</code> .
6	<code>04-assemble-final-datasets.R</code>	Joins funder classifications, supplier lookups, and raw payment data into the final output file. Computes per-organisation summary statistics (<code>total_value_payments_to_org</code> , <code>total_number_payments_to_org</code>).

Note: Steps 3–5 handle the alias review pipeline. In a fully automated run, Step 3 generates candidate batches, but the LLM review must occur externally before Step 4 can assemble decisions. When reviewed batches already exist, all steps run in sequence without manual intervention.

A2. Funder Classification

Funders are classified into nine types (eight substantive types plus a Junk/Invalid catch-all) using a multi-layer cascade. Classification is applied independently to `funder_name` (producing `funder_type`) and `funder_name_alt` (producing `funder_type_alt`). The cascade proceeds in order; the first match wins.

Classification Cascade

Pre-filter: Junk/Invalid Detection

Entries with purely numeric names, hash-like strings, or very short names (fewer than 3 meaningful characters) are flagged as `Junk/Invalid` and excluded from subsequent cascade layers. This affects 16 funder names (142 records).

Layer 1: Metadata Signals

Information already present in the source data is used as the first classification signal:

- Funders from the `NHSSpend` data source are classified as **NHS**
- Funders from the `Central Government` data source are classified as **UK Government**
- The `note` field in the funder masterlist may contain signals such as "ministerial" or "non-ministerial", indicating **UK Government**

Layer 2: External Lookups

Unclassified funders are matched against two external reference databases:

- **findthatcharity** (<https://findthatcharity.uk/>) — a comprehensive lookup of UK organisations. The `organisationType` field is mapped to funder types (e.g., `nhs-trust` → NHS, `local-authority` → Local Government). Both exact and fuzzy matching (Jaro-Winkler, threshold ≥ 0.90) are used.
- **TCSS Organisation Register** — funders that match the Organisation Register are classified as **CSO** (a civil society organisation acting as a funder).

Layer 3: Keyword Rules

Remaining unclassified funders are matched using pattern-matching rules applied to the funder name. Rules are applied in priority order; the first match wins:

1. **NHS** — names containing "NHS" combined with "TRUST", "CCG", "ICB", etc.
2. **UK Government** — known department abbreviations (FCDO, DFID, etc.) and patterns like "BRITISH EMBASSY", "SCOTTISH GOVERNMENT"
3. **Local Government** — names containing "CITY COUNCIL", "COUNTY COUNCIL", "BOROUGH COUNCIL", etc.
4. **Police, Fire and Rescue** — names containing "CONSTABULARY", "POLICE", "FIRE" + "RESCUE"
5. **Education Institution** — names containing "UNIVERSITY", "COLLEGE", "ACADEMY TRUST", school patterns
6. **CSO** — names ending in "CIC" or containing "TRUST" or "CHARITY" (after NHS and Academy Trusts have been captured)
7. **Private Sector** — commercial companies identified by the external lookups (registered companies that are not civil society organisations)
8. **Other** — all remaining unclassified funders

Funder Type Taxonomy

Funder Type	Description	Examples
UK Government	Central government departments, agencies, and arm's-length bodies	DEPARTMENT FOR EDUCATION, MINISTRY OF DEFENCE, DVLA
NHS	NHS trusts, clinical commissioning groups, integrated care boards	NHS ENGLAND, BARTS HEALTH NHS TRUST, NHS HRAW CCG
Local Government	County, district, borough, and unitary councils; combined authorities	MANCHESTER CITY COUNCIL, KENT COUNTY COUNCIL
Education Institution	Universities, colleges, academy trusts, schools	UNIVERSITY OF OXFORD, HARRIS FEDERATION
CSO	Civil society organisations acting as funders (grant-makers, intermediaries)	THE NATIONAL LOTTERY COMMUNITY FUND
Police, Fire and Rescue	Police forces, fire and rescue services	METROPOLITAN POLICE, LONDON FIRE BRIGADE

Funder Type	Description	Examples
Private Sector	Commercial companies appearing as the paying body — typically procurement intermediaries, managed-service providers or housing companies acting on behalf of public bodies (identified mainly through the external lookups)	ATTAIN HEALTH MANAGEMENT SERVICES LTD, ATAMIS LTD
Other	Funders not classifiable into the above categories	Various unclassified public bodies
Junk/Invalid	Clearly erroneous entries (numeric strings, hash tokens)	3149053, 37300

A3. Supplier Matching & Alias Resolution

The raw spending data contains supplier names as entered by government departments — often inconsistent in spelling, abbreviation, and formatting. The pipeline matches these names to organisations in the TCSS Organisation Register to assign a standardised `uid` to each supplier.

Matching Process

Matching proceeds in two stages:

1. **Exact matching** — supplier names are normalised (uppercased, punctuation removed, whitespace collapsed) and matched exactly against the Organisation Register.
2. **Fuzzy matching** — unmatched suppliers are compared to the Register using Jaro-Winkler string similarity. Matches with a similarity score ≥ 0.90 are accepted. This captures variations in spelling, abbreviation (e.g., “LTD” vs “LIMITED”), and minor data entry errors.

Alias Resolution

After initial matching, the pipeline identifies potential aliases — cases where the same organisation appears under different supplier names. This is common when departments record the same supplier differently (e.g., “ST LUKE’S HOSPICE” vs “SAINT LUKES HOSPICE”).

Alias resolution proceeds in three steps:

1. **Candidate generation** (Script 03a) — deterministic rules identify supplier name pairs that may refer to the same organisation, based on shared UIDs, similar names, or overlapping funder relationships.
2. **Batch review** — candidate pairs are grouped into batches and reviewed using a combination of LLM-assisted classification and manual checks. Each pair is labelled as a confirmed alias or a false positive.
3. **Merge** (Scripts 03b–03c) — confirmed aliases are assembled into a validated alias file, and the supplier lookup is updated to consolidate duplicate entries under a single canonical record.

Note: The LLM review step occurs externally between Scripts 03a and 03b. In a fully automated run where reviewed batches already exist, all scripts execute in sequence without manual intervention.

A4. Contracts Finder Records in the Payments Dataset

Contracts Finder (<https://www.contractsfinder.service.gov.uk/>) is the UK government's portal for public procurement opportunities. The TCSS project operates its own Contracts Finder collector (https://github.com/uk-third-sector-database/contracts_finder) that scrapes awarded contract notices from the portal's API.

Integration Process

Contracts Finder records are integrated into the procurement dataset through a crosswalk file (`contracts-finder-crosswalk.csv`) that links awarded contract notices to the main payment data. The crosswalk maps Contracts Finder notice identifiers and awarded amounts to the standardised format used by the rest of the dataset.

Unlike the payment-level data from Central Government and NHS sources, Contracts Finder provides contract-level information — each record represents an awarded contract rather than an individual payment. These records are assigned a `data_source` value of `Contracts Finder` or `contractsfinder` (reflecting different collection batches).

Note: Contracts Finder records account for a relatively small share of the dataset (2.1%, 18,184 rows) but contribute a disproportionate number of unique organisations (4,977 register-matched organisations), as they capture a broader range of contract awards that may not appear in the £25k+ payment transparency data. The complete contract-level data — all awarded notices, including those won by non-civil-society suppliers, with full contract metadata — is published as the companion Contracts Finder Records (<https://uk-third-sector-database.github.io/guidance/tcss-contracts-finder-guidance.html>) dataset.

A5. Linking to Other TCSS Datasets

The procurement dataset can be linked to other datasets in the UK Third and Civil Society Sector Database (<https://uk-third-sector-database.github.io/data/>) using the `uid` field. This unique organisation identifier is consistent across all TCSS datasets, enabling researchers to enrich procurement records with organisation characteristics, financial data, and other information.

Available Linkages

Dataset	Join Key	What It Adds
Organisation Register (https://uk-third-sector-database.github.io/guidance/tcss-organisation-register-guidance.html)	<code>uid</code>	Organisation name, postcode, registration and removal dates, source registers, SIC codes

Dataset	Join Key	What It Adds
Charity Financial Records (https://uk-third-sector-database.github.io/guidance/tcss-charity-financial-records-guidance.html)	uid	Annual income, expenditure, and detailed financial breakdowns for registered charities (GB-CHC, GB-SC, GB-NIC prefixed UIDs)
Nonprofit Financial Records (https://uk-third-sector-database.github.io/guidance/tcss-nonprofit-financial-records-guidance.html)	uid	Companies House accounts data — balance sheets, profit and loss, employee numbers — for nonprofit companies (GB-COH prefixed UIDs)
Contracts Finder Records (https://uk-third-sector-database.github.io/guidance/tcss-contracts-finder-guidance.html)	uid (matched civil society suppliers) or supplier_id	Contract-level detail for Contracts Finder awards — title, description, CPV codes, estimated and awarded values, dates, buyer flags — and the wider contract market including non-civil-society suppliers
CIC Founding Purposes (CIC 36/37 forms) (https://uk-third-sector-database.github.io/guidance/tcss-cic-statements-guidance.html)	uid	Community interest statements, beneficiary descriptions, and activity summaries for Community Interest Companies

Example: Linking to the Organisation Register

```
import pandas as pd

# Load datasets
procurement = pd.read_csv("cso-procurement-dataset-refined.csv")
spine = pd.read_csv("TSCS_spine.spine.csv")

# Link procurement records to organisation characteristics
merged = procurement.merge(
    spine[["uid", "organisationname", "postcode", "dateregistered"]],
    on="uid",
    how="left"
)

# Example: count procurement recipients by region (requires postcode lookup)
print(merged.groupby("postcode").size().sort_values(ascending=False).head(10))
```

Example: Linking to Charity Financial Records

```
import pandas as pd

# Load datasets
procurement = pd.read_csv("cso-procurement-dataset-refined.csv")
charity_finance = pd.read_csv("cso-spine-charity-financial-history.csv")

# Filter procurement to charities only
charity_procurement = procurement[procurement["uid"].str.startswith(("GB-CHC", "GB-SC", "GB-NIC"))]

# Link to latest financial year
latest_finance = charity_finance.sort_values("fy").groupby("uid").last().reset_index()
merged = charity_procurement.merge(
    latest_finance[["uid", "inc", "exp"]],
    on="uid",
    how="left"
)

# Example: compare procurement spend to charity income
print(merged[["uid", "amount", "inc"]].head(10))
```

Tip: When linking datasets, use a `left` join from the procurement data to preserve all payment records, even if some organisations are not found in the target dataset. Check for missing values after the join to assess linkage coverage.