

Procurement Records

A Guide to the Dataset

March 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 881,214 payment records covering 10,774 organisations and 1,421 public funders, spanning the period from 2010 to 2025. Together, these records offer a comprehensive view of the scale, distribution, and evolution of public procurement relationships with civil society across the UK.

881,214

Payment Records

10,774

Organisations

1,421

Public Funders

2010–2025

Years Covered

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 90.3% 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 approximately 2010 to April 2020 when data collection concluded. NHS spending accounts for 7.8% 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). These records account for 1.9% of the dataset.

3. Dataset Contents

The dataset contains 881,214 records across 15 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-C0H-12345678 , GB-CHC-1234567 , GB-SC-SC012345)	Text	100%
organisation_name	Organisation name as recorded in the source data	Text	80.3%
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	99.8%

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	100%
funder_name_alt	Alternative funder name, where applicable	Text	1.0%
funder_id	Unique funder identifier (e.g., F02683)	Text	100%
funder_type	Funder classification: UK Government , NHS , Local Government , Education Institution , CSO , Police, Fire and Rescue , Other ,OR Junk/Invalid	Text	100%
funder_type_alt	Classification of the alternative funder name	Text	1.3%

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 fields — `uid` , `payment_date` , `amount` , and `data_source` — are present in all records. The `organisation_name` field is missing for approximately one in five records; these payments are still linked to valid organisations via `uid` and `supplier_id` .

uid

payment_date

amount

100%

100%

100%

organisation_name

80.3%

Data Source Breakdown

The dataset draws from four data source tags. Central government spending dominates, accounting for over 90% of all records.

Source	Records	Unique Organisations	Share
Central Government	796,128	7,222	90.3%
NHSSpend	68,528	2,066	7.8%
Contracts Finder	16,558	4,397	1.9%

Note: The Contracts Finder row combines records tagged as `Contracts Finder` and `contractsfinder` in the `data_source` field, which reflect different collection batches from the same portal.

Funder Type Distribution

Funders are classified into eight types using a multi-layer cascade (see [Appendix A2](#)). UK Government departments account for the vast majority of funding.

Funder Type	Records	Share
UK Government	801,360	90.9%
NHS	70,107	8.0%
Local Government	7,610	0.9%
Education Institution	802	0.1%
Other	573	0.1%
CSO	423	<0.1%
Police, Fire and Rescue	319	<0.1%
Junk/Invalid	20	<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.

Supplier Type	Records	Share
Other CSO	610,993	69.3%
Charity	222,247	25.2%
Co-operative/Mutual	28,736	3.3%
CIC	17,576	2.0%
<i>(Missing)</i>	1,662	0.2%

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	5,140
2011	7,892
2012	52,304
2013	116,804
2014	166,657
2015	53,970
2016	29,817
2017	64,218
2018	63,025
2019	66,216
2020	59,325
2021	56,297
2022	61,212
2023	66,446
2024	10,911
2025	951

Note: Twenty-nine records fall outside the 2010–2025 range shown above. These include a small number with implausible payment years (e.g., 1900, 2027) due to source data errors, as well as records from 2005–2009 that predate the main collection period and 19 records dated 2026 from ongoing data collection. 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.

Prefix	Source Register	Records	Share
GB-COH	Companies House	629,751	71.5%
GB-CHC	Charity Commission for England & Wales	203,778	23.1%
GB-COO	Co-operatives UK	23,156	2.6%
GB-SC	Office of the Scottish Charity Regulator	18,623	2.1%
GB-MPR	Mutuals Public Register (FCA)	5,580	0.6%
GB-NIC	Charity Commission for Northern Ireland	183	<0.1%
GB-SHR	Scottish Housing Regulator	143	<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 full set of payment records to characterise the flow of public money to civil society organisations. Key findings include:

- **UK Government dominates** — UK Government departments are the largest source of public procurement spend to civil society across all organisation types, accounting for the majority of both payment volume and total value.
- **Charities receive the largest share** — Charities receive 42% of total procurement spend to civil society, reflecting their central role in public service delivery.

- **Scale of spending** — UK Government departments account for over £131 billion in cumulative payments, dwarfing the NHS (£15 billion), local government (£8 billion), and all other funder types combined.
- **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.

Table 4: Procurement by CSO (Civil Society Organisation)

Number of organisations, payments, and total value

	CSO (Civil Society Organisation)	Education Institution	Local Government	NHS (National Health Service)
CIC (Community Interest Company)	18 orgs 35 payments £7.1M	14 orgs 21 payments £1.1M	312 orgs 615 payments £795.4M	258 orgs 12,611 payments £4721.9M
Charity	154 orgs 247 payments £230.7M	167 orgs 443 payments £89.1M	1,784 orgs 5,353 payments £5386.0M	1,765 orgs 42,808 payments £8745.4M
Co-operative / Mutual	15 orgs 32 payments £9.4M	9 orgs 17 payments £3.3M	105 orgs 570 payments £1537.6M	96 orgs 3,607 payments £557.8M
Other	67 orgs 109 payments £64.2M	74 orgs 321 payments £39.6M	415 orgs 1,058 payments £386.7M	260 orgs 8,232 payments £1185.2M

Source: UK Civil Society Spine, Contracts Finder and procurement data



Table 9: Procurement by CSO (Civil Society Organisation) type and funder type. Each cell shows the number of organisations, payments, and total value. Source: UK Civil Society Spine, Contracts Finder and procurement data.

UK Government departments by CSO (Civil Society Organisation) type

Table 10 shows the proportion of each UK Government department's procurement spend that goes to each civil society organisation type. Departments vary considerably in where their procurement is directed. The Department for Culture, Media and 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 and Trade allocate over 75% to other nonprofit companies. The Department for Work and Pensions stands out for relatively high CIC (Community Interest Company) (6.1%) and co-operative/mutual (7.1%) shares compared with other departments.

Table 10: UK Government department procurement spend by CSO (Civil Society Organisation)

Proportion of each department's total spend to civil society by organisation type (%)

	Charity	CIC (Community Interest Company)	Co-operative / Mutual	Other CSO (Civil Society Organisation)
Bank of England	55.4%	0.0%	0.0%	

Source: UK Civil Society Spine, Contracts Finder and procurement data

Note: Includes 43 departments with total civil society spend of at least £10 million.

Table 10: UK Government department procurement spend by CSO (Civil Society Organisation)

Proportion of each department's total spend to civil society by organisation type (%)

	Charity	CIC (Community Interest Company)	Co-operative / Mutual	Other CSO (Civil Society)
Cabinet Office	72.4%	0.5%	0.0%	
Care Quality Commission	98.5%	0.0%	0.0%	
Crown Commercial Service	35.9%	0.3%	2.2%	
Defence Science & Technology Laboratory	76.3%	3.4%	0.0%	
Department for Business & Trade	17.0%	0.3%	0.5%	
Department for Culture, Media & Sport	98.0%	0.1%	0.1%	
Department for Education	23.3%	0.1%	0.3%	
Department for Energy Security & Net Zero	1.9%	2.0%	18.8%	
Department for Environment, Food & Rural Affairs	71.9%	0.7%	0.6%	
Department for International Development	92.9%	0.0%	0.0%	
Department for International Trade	7.8%	0.0%	0.0%	
Department for International Trade	6.5%	0.3%	0.0%	

Source: UK Civil Society Spine, Contracts Finder and procurement data

Note: Includes 43 departments with total civil society spend of at least £10 million.

Table 10: UK Government department procurement spend by CSO (Civil Society Organisation)

Proportion of each department's total spend to civil society by organisation type (%)

	Charity	CIC (Community Interest Company)	Co-operative / Mutual	Other CSO (Civil Society)
Department for Science, Innovation & Technology	38.7%		4.4%	0.0%
Department for Transport	20.9%		0.1%	1.7%
Department for Work & Pensions	57.6%		6.1%	7.1%
Department of Health & Social Care	82.1%		0.4%	0.0%
Driver & Vehicle Licensing Agency	21.4%		0.0%	0.0%
Food Standards Agency	12.6%		0.5%	0.0%
Foreign Office	93.9%		0.0%	0.0%
Foreign, Commonwealth & Development Office	84.0%		0.0%	0.0%
HM Land Registry	0.3%		5.8%	0.0%
HM Revenue & Customs	8.7%		0.1%	0.2%
HM Treasury	5.3%		0.1%	0.3%
Health Education England	35.8%		25.5%	0.3%
Highways England	94.6%		0.0%	0.0%
Home Office	38.2%		1.4%	0.2%

Source: UK Civil Society Spine, Contracts Finder and procurement data

Note: Includes 43 departments with total civil society spend of at least £10 million.

Table 10: UK Government department procurement spend by CSO (Civil Society Organisation)

Proportion of each department's total spend to civil society by organisation type (%)

	Charity	CIC (Community Interest Company)	Co-operative / Mutual	Other CSO (Civil Society)
Homes England	95.0%	0.0%	3.5%	
Institute for Apprenticeships & Technical Education	100.0%	0.0%	0.0%	
Medicines & Healthcare Products Regulatory Agency	99.6%	0.0%	0.0%	
Ministry of Defence	71.2%	0.8%	5.0%	
Ministry of Housing, Communities & Local Government	79.1%	0.3%	2.1%	
Ministry of Justice	63.4%	3.1%	3.3%	
Money & Pensions Service	100.0%	0.0%	0.0%	
NHS (National Health Service) Blood & Transplant	2.7%	0.0%	93.9%	
Public Health England	59.9%	10.8%	0.6%	
Scotland Office	53.8%	0.2%	4.3%	
Sellafield Ltd	1.1%	0.0%	0.0%	
Skills Funding Agency	50.0%	0.0%	0.0%	
Transport for London	66.7%	25.0%	6.5%	

Source: UK Civil Society Spine, Contracts Finder and procurement data

Note: Includes 43 departments with total civil society spend of at least £10 million.

Table 10: UK Government department procurement spend by CSO (Civil Society Organisation)

Proportion of each department's total spend to civil society by organisation type (%)

	Charity	CIC (Community Interest Company)	Co-operative / Mutual	Other CSO (Civil Society)
UK Health Security Agency	1.8%		0.0%	0.0%
UK Research & Innovation	42.0%		20.6%	3.3%
UK Shared Business Services	34.3%		7.1%	3.5%

Source: UK Civil Society Spine, Contracts Finder and procurement data

Note: Includes 43 departments with total civil society spend of at least £10 million.

Table 10: UK Government department procurement spend by civil society organisation type. Proportion of each department's total spend to civil society (%). Departments with at least £10 million total spend. Source: UK Civil Society Spine, Contracts Finder and procurement data.

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 19.7% 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 small number of records (22 in total) carry implausible payment years such as 1900 or 2027, resulting from errors in the source data. 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. 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.

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 entries (20 records).

What's NOT in the Data

The dataset does not include:

- Contract descriptions or service categories
- 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.)

7. Citation & Licence

Licence: This dataset is licensed under the [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/) (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*. 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
1.0	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 approximately 1,400 unique funders into eight types 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.

Step	Script	Purpose
4	03b-assemble-alias-decisions.R	Assembles alias decisions from both rule-based determinations and LLM-reviewed batch results into a single validated alias file.
5	03c-apply-alias-merges.R	Applies validated alias merges to the supplier lookup, consolidating duplicate supplier entries under a single canonical <code>uid</code> .
6	04-assemble-final-datasets.R	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 eight types 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 20 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.
- **TCCS 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. **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
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 Enrichment

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 (1.9%) but contribute a disproportionate number of unique organisations (approximately 4,400), as they capture a broader range of contract awards that may not appear in the £25k+ payment transparency data.

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/) (<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
Charity Financial Records (https://uk-third-sector-database.github.io/guidance/tcss-charity-financial-records-guidance.html)	<code>uid</code>	Annual income, expenditure, and detailed financial breakdowns for registered charities (GB-CHC, GB-SC, GB-NIC prefixed UIDs)
Nonprofit Financial Records (<i>guidance forthcoming</i>)	<code>uid</code>	Companies House accounts data — balance sheets, profit and loss, employee numbers — for nonprofit companies (GB-COH prefixed UIDs)
CIC 36 Forms (https://uk-third-sector-database.github.io/guidance/tcss-cic36-forms-guidance.html)	<code>uid</code>	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("tcss-procurement-records.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("tcss-procurement-records.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-NI
C"))]

# 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.