

Procurement Data and the UK Third Sector Database

Charlie Rahal

Gradel Institute of Charity, November 2025



The greatest things David Cameron achieved in office:

- ~~The Brexit Referendum~~
- Mandating procurement data being made public.

The greatest things David Cameron achieved in office:

- ~~The Brexit Referendum~~
- Mandating procurement data being made public.

The greatest things David Cameron achieved in office:

- ~~The Brexit Referendum~~
- **Mandating procurement data being made public.**



Original Article

The Keys to Unlocking Public Payments Data

Charles Rahal

First published: 20 April 2018 | <https://doi.org/10.1111/kykl.12171> | Citations: 7[Read the full text >](#)[PDF](#) [TOOLS](#) [SHARE](#)

Summary

We mechanize some of the richest yet significantly under-utilized data resources within developed, 'Open Data' economies. We show how it is possible to scrape, parse, clean and merge tens of thousands of disaggregated public payments datasets in an attempt to bridge the methodological gap between newly available data from the administrative sphere and applications in empirical social science research. We outline techniques to unambiguously link records to various freely available institutional registers. In particular, we offer guidance on overcoming the substantial challenges of heterogeneous provision and administrative recording errors in the absence of Uniform Resource Identifiers, namely in the form of an approximate, domain-specific 'record-linkage' type matching algorithm. As an illuminating example, we construct a cleaned database of 24,581,192 local government payments subject to the Local Transparency Codes which total £169.87bn in value. We overcome various challenges in a detailed examination of the procurement of services by local government from the voluntary sector: an important contemporary issue due to the rise of the 'Big Society' political ideology of the early 21st century. Finally, we motivate future work in this area and discuss potential international applications and practical advancements.

References

Volume 71, Issue 2
May 2018
Pages 310-337

Advertisement

[References](#) [Related](#) [Information](#)

Tools for Transparency in Central Government Spending

Charles Rahal

Department of Sociology, University of
Oxford, 42-43 Park End Street, Oxford OX1
1JD, UK

Abstract

We first introduce centgovspend, an open source software library which provides functionality to automatically scrape and parse central government spending within the United Kingdom at the micro and meso levels. The library then optionally reconciles suppliers and subsequently analyzes payments made to private entities. We briefly discuss the policy environment surrounding the library before explaining the modular structure, implementation and execution which results in scraping over 4.9m payments worth over £3.5tn in value. We then provide two prototype applications in the fields of public administration and sociology; one of which analyzes government procurement across Standard Industry Classifier (SIC) and one which analyzes the social stratification of company officers and persons of significant control who supply the government. The project acts as a prototype in an international context, aiming to highlight an unrealised possibility of 'Big, Open Data' for public policy making and government efficiency.

"I do not think that Ministers understand how little trust there is left."

Lisa Nandy MP (Wigan) (Lab), House of Commons, 25th March, 2019 [1]

Introduction

David Cameron's introduction of new requirements in May 2010 enabled the United Kingdom to lead the world in terms of transparency and 'Open Data' [2]: an important ambition to realize given that roughly one in every three pounds spent by the public sector is spent on procurement [3]. The new regime applies first and foremost to central government departments which procure from thousands of suppliers varying in size, ranging from large 'Strategic Suppliers' to small businesses. It theoretically allows us to transparently analyze and track longitudinal changes in fiscal policy at a granular level during a period of aggressively targeted deficit reduction.

Procurement data on either individual purchases or payments related to contractual obligations is able to promote government efficiency and effectiveness, as well as empowering citizens with an understanding of the inner workings of the public sector. The provision of such data is mandated across a range of administrative levels (at various financial thresholds), such as central government departments, local authorities [4], smaller local councils [5], and a range of other public institutions such as NHS bodies, emergency services and public transport network providers [6]. At a national level, the publication of information on expenditure over £25,000 is one of the few mandated data-sets that ministerial and non-ministerial departments must provide, with information required on the supplier, the date of transaction, the



The cover of the IJPSD Standard Issue, Issue 1, Vol 4, 2019. It features a collage of images related to data science and public administration, including a group of people in a meeting, a cityscape, and various data visualizations. The IJPSD logo is at the top, and the text 'The Science of Data About People' is below it. The title 'STANDARD ISSUE' is prominently displayed, followed by 'Issue 1, Vol 4' and '2019'. Logos for Open Access, Crossref, and DOAJ are at the bottom.

[Download/View Article](#)

Published: Jul 10, 2019

Updated: 2019-07-10

DOI: <https://doi.org/10.23889/ijpsd.v4i1.1092>

Keywords: Civic Technology, Social Data Science, Procurement, Public Administration

Statistics from Altmetric

Journal of the Royal Statistical Society

Series A: Statistics in Society

[Issues](#) [More content ▼](#) [Submit ▼](#) [Purchase](#) [Alerts](#) [About ▼](#)[Journal of the Royal](#) ▼[Volume 188, Issue 4](#)
[October 2025](#)

Article Contents

Abstract

1 Introduction

2 Nonstate providers of healthcare in the British NHS: rationale, extent of involvement, and difficulties in quantification

3 Data

4 Procurement from third sector organizations

5 Discussion and conclusion

Acknowledgments

Funding

Data availability

Supplementary material

References

Footnotes

Author notes

JOURNAL ARTICLE

The role of the third sector in public health service provision: evidence from 25,338 heterogeneous procurement datasets

[Charles Rahal](#) ✉, [John Mohan](#) [Author Notes](#)

Journal of the Royal Statistical Society Series A: Statistics in Society, Volume 188, Issue 4, October 2025, Pages 1085–1106, <https://doi.org/10.1093/jrssa/qnae092>

Published: 27 September 2024 [Article history ▼](#)

[PDF](#) [Split View](#) [Cite](#) [Permissions](#) [Share ▼](#)

Abstract

We examine the role of nonprofits within publicly funded healthcare which runs parallel to private provision in a 'two-tier' system through a unique Big Data pipeline. Scraping tens of thousands of heterogeneous transaction datasets across a commissioning hierarchy, the processed dataset contains over £445Bn worth of transactions in 1.9m+ rows of clean data, spanning 2012–2020. We utilize this dataset to test a range of hypotheses related to the introduction of the Health and Social Care Act of 2012. The proportion of contracts placed with nonprofit organizations is relatively low, with limited evidence as to whether the Act increased the involvement of third sector organizations in line with NHS 'marketization'. We analyse the pattern of procurement by corporate category, the field of service delivery (International Classification of Nonprofit Organizations and Standard Industrial Classification codes), and 'expense area' to show the unique array of services which voluntary organizations supply. We also analyse the pattern of commissioning across entity class and size distributions of registered charities. We conclude with a consideration of high-value Community Interest Companies, and discuss potential further areas of research within a healthcare context which such government transaction data makes possible.

Current Project: ES/X000524/1

- Creating a national database integrating diverse UK third sector organisations.
- Linking administrative datasets to reveal income, employment, and partnerships.
- Enhancing research, policy, and recognition of the voluntary sector's impact.

UK Data Service
ReShare

[Home](#) [Legal](#) [Review procedures](#) [Metrics](#)

Third Sector and Civil Society Organisations in the UK, 1869-2025

Rutherford, Alasdair and Mohan, John and McDonnell, Diarmuid and Rahal, Charles and Clifford, David and Couper Kenney, Fiona and Duggan, Alan and McNicol, Stacey (2025). *Third Sector and Civil Society Organisations in the UK, 1869-2025*. [Data Collection]. Colchester, Essex: UK Data Service. [10.5255/UKDA-SN-85800](https://beta.ukdataservice.ac.uk/datacatalog/studies/study?id=10.5255/UKDA-SN-85800)

The voluntary sector is widely acknowledged as containing very large numbers of organisations that make an enormous contribution to well-being and social cohesion in the UK. It encompasses charities, social enterprises, mutuals, cooperatives, and many less formal voluntary and community organisations. We know a great deal from survey data about patterns of individual giving to charities, and about patterns of volunteering. But there is a substantial gap in the availability of high-quality data about voluntary organisations. And it is argued that better-quality information and evidence would lead to the contribution of those organisations being properly recognised, leading in turn to higher levels of public and voluntary support for them.

This project responds to this need by creating the first national database on the population of organisations forming the third sector through bringing together information about charities with information about different kinds of noncharitable civil society organisations (including Community Interest Companies, Co-operatives and Mutuals, and non-profit Companies Limited by Guarantee). A variety of relevant registers are maintained by different regulators and public bodies across the four nations of the UK. Each register holds partial and sometimes overlapping information about civil society organisations. This project combines them into a single, deduplicated dataset, linking records that refer to the same organisation.

The project will open up new avenues for research on, for example, the survival and growth of third sector organisations; the contribution of the third sector to public service delivery; the level of voluntary sector activity in different parts of the country. In this way, we will make the contribution of the sector much more visible to stakeholders.

In strategic terms this work provides an important demonstration of the uses of administrative data for research on organisations - a field in which most work to date has focussed on administrative data on individuals. The dataset can be linked to publicly available information on government and NHS spending, allowing users to identify which organisations receive public funding. It can be used to support further analysis of the size, scope, and distribution of third sector organisations across the UK. This will be of value to funders (such as charitable foundations), central and local government, local and regional councils for voluntary service, which sustain these organisations locally, and commissioners of public services such as the NHS, which rely substantially on the voluntary sector to deliver services.

DATA DESCRIPTION (ABSTRACT)

This dataset brings together information on all organisations registered with at least one charitable or non-profit regulator in the United Kingdom. It was developed by combining and deduplicating records from ten separate registers, covering organisations across all four nations of the UK. The Third Sector and Civil Society (TSCS) in the UK are overseen by a range of regulatory bodies, including three national charity regulators and several additional registers for non-profit entities. Companies House, the UK's official register of companies, also includes non-profit types such as companies limited by guarantee and community interest companies.

Because these registers operate independently and often overlap—many organisations appear in more than one—the available data is fragmented. This fragmentation makes it difficult to form a clear picture of civil society and its role within the UK. By integrating these diverse sources and linking records where evidence suggests they refer to the same organisation, we produce a unified “spine” of TSCS organisations. This consolidated list enables consistent mapping, analysis, and monitoring of the sector as a whole. The spine can also be linked to other datasets—such as government procurement records—to examine patterns in public spending, or to demographic data to explore the geographic and social distribution of civil society activity across the UK.

Project Accreditation

Principal Investigator(s): John Mohan
Principal Investigator(s): Alasdair Rutherford
Co-Investigator: David Clifford
Co-Investigator: Diarmuid McDonnell
Co-Investigator: Charles Rahal

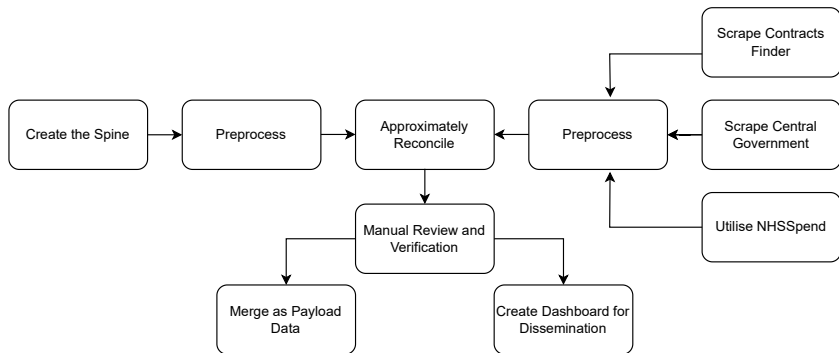
Initially funded by the ESRC, with generous support from the Gradel Institute for the purpose of building a procurement dashboard.

Research Associates

- Fiona Cooper Kenney
- Stacey McNicol
- Alan Duggan
- Steve Barnard
- Annabel Young

These research associates played a vital role in data curation, linkage, and analysis, underpinning the project's creation of a comprehensive national database of third sector organisations.

Research Pipeline



Scraping Libraries

The raw procurement data comes in three forms:

1. Contracts Finder: [10.5281/zenodo.14951539](#)
2. Central Government: [10.5281/zenodo.14951531](#)
3. NHSSpend: [10.5281/zenodo.5054679](#)

uk-third-sector-database

Type to search



+ ▾

+ ▾



Overview

Repositories 11

Projects 1

Packages

Teams

People 6

Insights

Settings



UK Third Sector Database

Welcome to the home of the UK Third Sector Database!

6 followers United Kingdom <http://uk-third-sector-database.github...>

Unfollow

README .md



Welcome to the home of the UK Third Sector Database! We are an [ESRC funded group of collaborators](#) seeking to improve data infrastructure surrounding the UK's third sector and civil society more generally. The grant is entitled "Improving Access to and Use of Organisation-Level Data on the Third Sector and Civil Society" (Project Reference: ES/X000524/1). In this repository you can find all of the code that powers our data, dashboards, public APIs and tutorials. We also have a [website](#).

Datasets

You can access some of our public data products as follows:

- UK third sector database of organisations | [Data](#) | [Project background](#) |

View as: Public ▾

You are viewing the README and pinned repositories as a public user.

You can [pin repositories](#) visible to anyone.

You can [hide the tasks we've suggested](#) on this page and bring them back later.

Discussions

Set up discussions to engage with your community!

[Turn on discussions](#)

We think you're gonna like it here.

We've suggested some tasks here in your organization's overview to help you get started.

 Invite your people



Invite your first member



Customize members'

Repositories

[procurement_dashboard](#)

 Jupyter Notebook

☆0

Updated yesterday

[uk-third-sector-database.github.io](#)

 JavaScript

☆0

Updated on Sep 15

[tso-database-builder](#)

 Jupyter Notebook

☆0

Updated on Sep 15

Even after scraping the data, parsing it represents a considerable challenge.

- Data is heterogenously provided.
- It comes in an unbelievable multitude of formats
- Organisations do not follow provided schemas.
- We do our best...

Even after scraping the data, parsing it represents a considerable challenge.

- Data is heterogenously provided.
- It comes in an unbelievable multitude of formats
- Organisations do not follow provided schemas.
- We do our best...

Even after scraping the data, parsing it represents a considerable challenge.

- Data is heterogenously provided.
- It comes in an unbelievable multitude of formats
- Organisations do not follow provided schemas.
- We do our best...

Even after scraping the data, parsing it represents a considerable challenge.

- Data is heterogenously provided.
- It comes in an unbelievable multitude of formats
- Organisations do not follow provided schemas.
- We do our best...

Even after scraping the data, parsing it represents a considerable challenge.

- Data is heterogenously provided.
- It comes in an unbelievable multitude of formats
- Organisations do not follow provided schemas.
- We do our best...

Expenditure Over Threshold

[illegible]

13 / 17

Expenditure Over Threshold

[illegible]

Expenditure Over Threshold

[illegible]

Even after parsing the data, linkage represents a considerable challenge:

- There's no unique identifiers.
- We begin with approximate matching informed manual linkage
- This gets supplemented with 'unverified' computer matches.
- We do our best...

Even after parsing the data, linkage represents a considerable challenge:

- There's no unique identifiers.
- We begin with approximate matching informed manual linkage
- This gets supplemented with 'unverified' computer matches.
- We do our best...

Even after parsing the data, linkage represents a considerable challenge:

- There's no unique identifiers.
- We begin with approximate matching informed manual linkage
- This gets supplemented with 'unverified' computer matches.
- We do our best...

Even after parsing the data, linkage represents a considerable challenge:

- There's no unique identifiers.
- We begin with approximate matching informed manual linkage
- This gets supplemented with 'unverified' computer matches.
- We do our best...

Even after parsing the data, linkage represents a considerable challenge:

- There's no unique identifiers.
- We begin with approximate matching informed manual linkage
- This gets supplemented with 'unverified' computer matches.
- We do our best...

Some Aggregate Numbers

Some useful aggregate numbers to have in mind over the course of the following slides...:

- Number of unique suppliers: 213,599
- Number of payments: 9,244,448
- Total value of payments: £5,043,474,546,039.619

Some Aggregate Numbers

Some useful aggregate numbers to have in mind over the course of the following slides...:

- Number of unique TSOs: 13,057
- Number of payments to TSOs: 708,818
- Total value of payments to TSOs: £121,077,073,087.71

<https://procurementdashboard.streamlit.app/>