The Keys to Unlock Public Payments Data

Charles Rahal – charles.rahal at sociology.ox.ac.uk

Civil Society Data Partnership

Paper available: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2815845. Project website: http://tsrc-ncvo-csdp.com

Introduction

We outline the tools to operationalize some of the largest, richest, but as of yet unexamined sets of resources within developed, 'Open Data' economies. As an illuminating example, we construct a cleaned database of 24,581,192 local government payments subject to the Local Transparency Codes worth a total of £169.87bn. We develop a range of tools to map this database to various publicly available institutional registers, and we offer guidance on overcoming the substantial challenges of heterogenous provision and administrative recording errors in the absence of uniform resource identifiers. Our applications of the database provide guidance on what such an endeavor can be used for, and we apply various methodologies such as network visualization and spatial analysis as motivation for future work in this area.

Methodology

We develop a range of scripts in Python and Matlab which are able to locate, parse and clean tens of thousands of raw payments datasets mandated by law, and map these to registers of UK institutions which we expect are likely recipients of said payments. In particular, we develop a range of exact, normalization and approximate based tools which revolutionize analysis of this sort, adding a layer of a transparency to an otherwise opaque set of data.

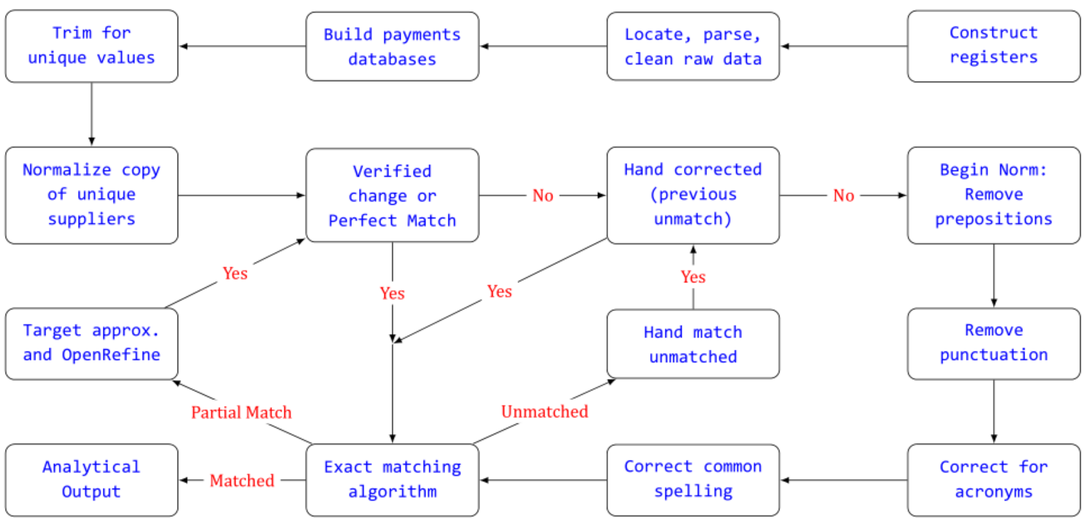

Figure 1: Flow Diagram for Database Construction

What can we match?

We map to six sets of institutional registers: Companies House, the Charity Comission (and CCNI and OSCR), a list of community amateur sports clubs (CASCs), a HSCIC-register of health care institutions, a list of all public entities and a list of all school and HEIs in the UK.

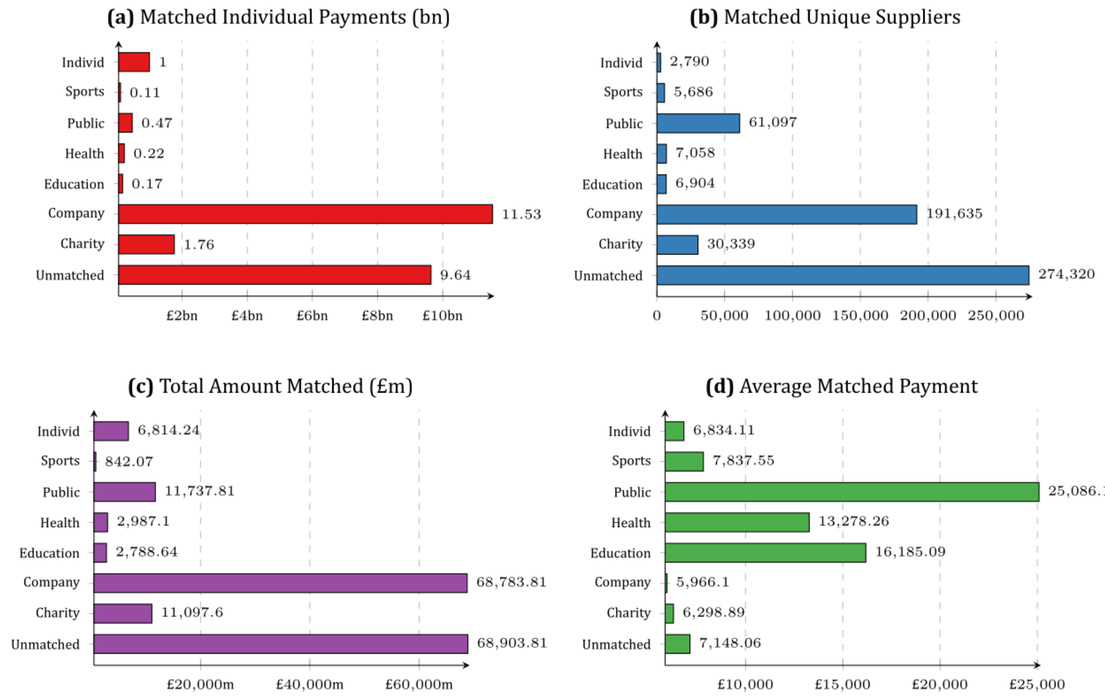

Figure 2: Matching Payments to Institutional Registers

Applications

We apply our methodological advancements to the granular data mandated by Local Authority Transparency Codes. Firstly, Figure 3 shows the relationship between each Local Authority (LA) with subsidiaries of G4S – a British multinational security services company. It visualization the 272 relationships between 180 unique LAs and 10 G4S subsidiaries unweighted across 7,695 payments totaling £19.88m.

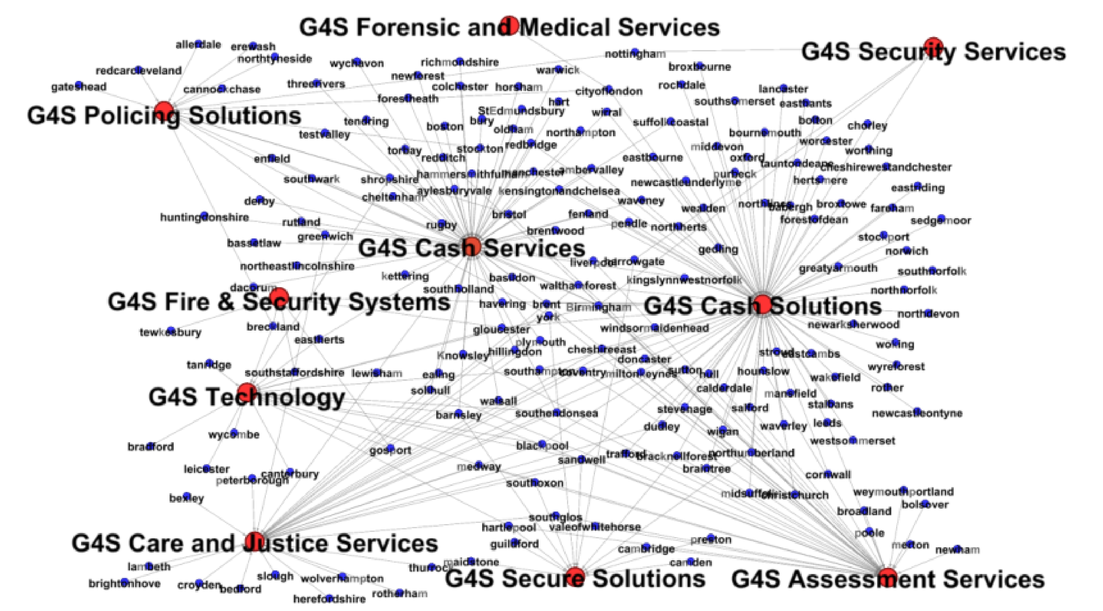

Figure 3: A Network Analysis of relationships with G4S Subsidiaries

Our second application analyzes third sector recipients. We decompose our matches to recipients only appearing on the Charity Commission register, and then aggregate to ICNPO categories. Figure 4a shows payments by value, and Figure 4b shows payments by frequency. As with companies, we also provide analysis of the most frequent and highest value individual suppliers.

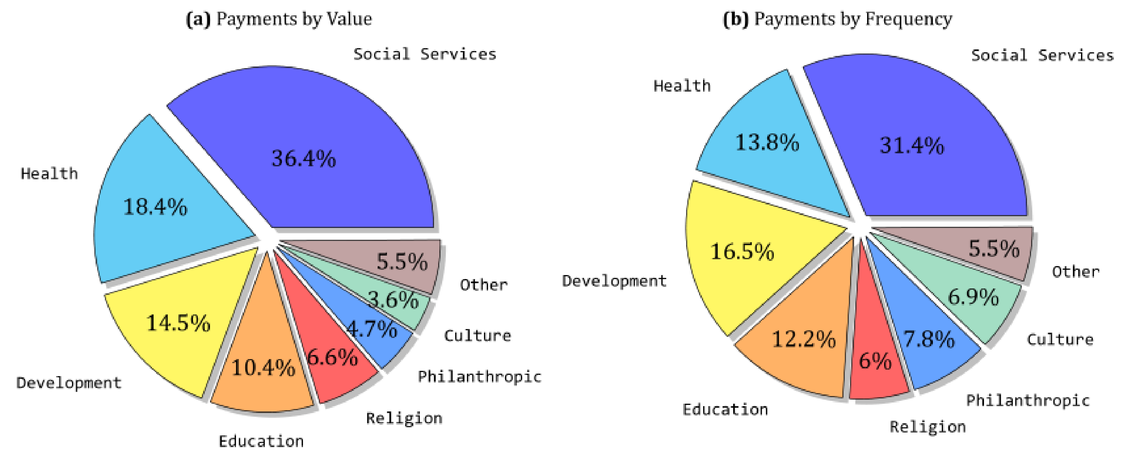

Figure 4: A Breakdown of Third Sector Recipients by ICNPO Group

Data of this sort has rich potential for spatial analysis. Choosing Bristol as an arbitrary example, we can show the distribution of payments from Bristol City Council to general-practice health-care institutions on the HSCIC register. Having linked the string matched payments to this dataset which contains postcodes, we are able to geocode the location of the recipients as in Figure 5.

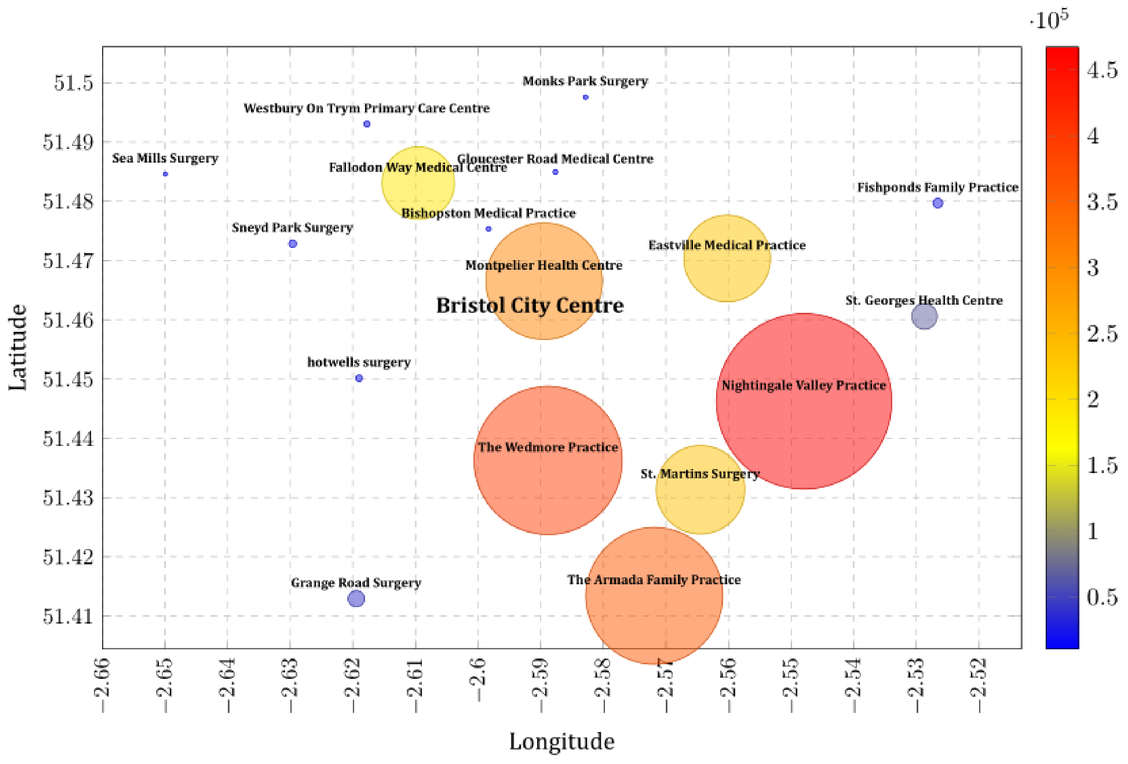

Figure 5: A Spatial Analysis of General Practice Health-care Payments in Bristol

We are also able to link payments from local authorities to individual schools found on the Key Stage 4 attainment register. Firstly, we use Free School Meal (FSM) Eligibility as a measure of family socio-economic status, and secondly, we consider payments to schools scattered against the percent of non-native English speaking pupils at the school. While the paper associated with the poster also considers payments with respect to educational attainment (GCSE results), obvious (multivariate) extensions would be to estimate a model which also included data such as family income, parental employment and parental achievements.

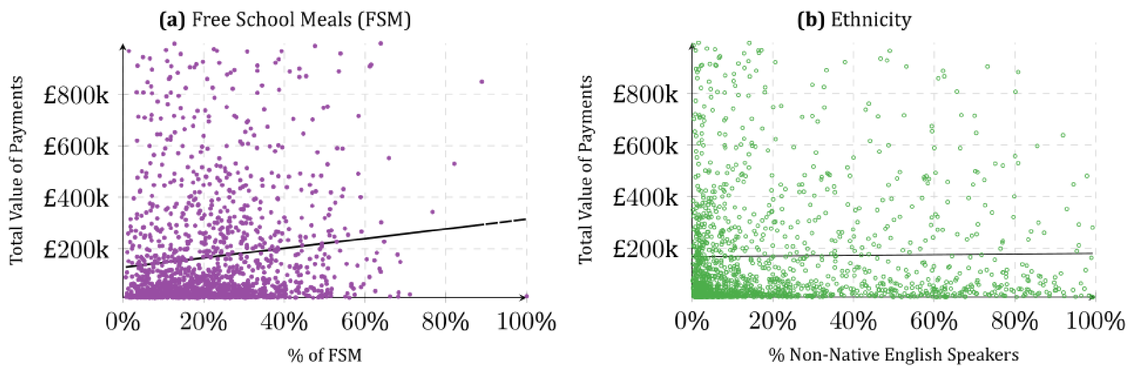

Figure 6: Payments to Schools

Conclusions

- We present a range of tools which are able to mechanize a wealth of hitherto unexamined data within the realm of public economics/public administration/political economy.
- Other examples of potential applications using LA data are shown in the paper. This includes a breakdown by SIC code, which sports are funded through community sports clubs, and which types of public bodies are receiving LA payments.
- This is only made possible by open, transparent governance structures, despite the the non-standard provision of the data at source.
- Our response to the LTA consultation can also be found at the project website, and provides the opportunity to further revolutionize the flourishing provision of Open Data for Policy

Forthcoming Research

The number of potential applications within this domain is endless. One specific example of the deployment of this methodology explicitly examines payments from clinical commissioning groups (CCGs) since the introduction of the Health and Social Care Act, and a working paper version of this can be found on the project website.

Acknowledgments

Thanks are due to John Mohan, David Kane, Marc Lawson and Cin Man Winnie Yeung. Financial support gratefully received for this project by the Economic and Social Research Council (project number ES/M010392/1).