

DEPARTMENT OF ECONOMICS

FOI as a Data Collection Tool for Economists

Joanna Clifton-Sprigg, Jonathan James & Sunčica Vujić

UNIVERSITY OF ANTWERP Faculty of Applied Economics

City Campus Prinsstraat 13, B.226 B-2000 Antwerp Tel. +32 (0)3 265 40 32 Fax +32 (0)3 265 47 99 www.uantwerpen.be



FACULTY OF APPLIED ECONOMICS

DEPARTMENT OF ECONOMICS

FOI as a Data Collection Tool for Economists

Joanna Clifton-Sprigg, Jonathan James & Sunčica Vujić

RESEARCH PAPER 2017-008 SEPTEMBER 2017

University of Antwerp, City Campus, Prinsstraat 13, B-2000 Antwerp, Belgium Research Administration – room B.226 phone: (32) 3 265 40 32 fax: (32) 3 265 47 99 e-mail: joeri.nys@uantwerpen.be

The research papers from the Faculty of Applied Economics are also available at <u>www.repec.org</u> (Research Papers in Economics - RePEc)

D/2017/1169/008

FOI as a Data Collection Tool for Economists¹

Joanna Clifton-Sprigg (University of Bath) Jonathan James (University of Bath) Sunčica Vujić (University of Antwerp)

Abstract

This paper sets out a method of generating a unique data set that has been underused by economists – a Freedom of Information (FOI) request. The FOI Act came into force in 2005 in the UK and allows the public to make requests of publicly held data. We explain how they can be made and provide suggestions on how to make effective data driven requests, those most frequently made by economists. Finally, we document the determinants of one particular FOI request. We applied for crime data from all police forces in the UK and examine the determinants of that request. In general, we find that observable characteristics of the local area or the police force neither determine whether the request was fulfilled, nor the speed at which it was responded to.

Keywords: Data Collection, Data Access

JEL Classification: C80, C81

¹ We thank Michèle Belot and John Hudson for useful comments and suggestions. This paper is accompanied by Supplementary Material in the Online Appendix available at: http://www.mwpweb.eu/JonathanJames/further_1.html

I. Introduction

This paper focuses on an important potential source of self-generated data that has not received much attention, nor has (yet) been fully exploited by economists - Freedom of Information (FOI) requests. FOI acts have been introduced in many countries over recent decades. The US passed a FOI act in 1966, followed by Australia, Canada and New Zealand in 1982, Ireland in 1997 and the UK in 2000 (coming into force on 1st January 2005).² FOI allows researchers access to bespoke datasets, which are typically administrative, given the FOI's focus on the public sector. Administrative data has a number of advantages over survey data. Specifically, administrative samples have universal coverage (and hence large sample sizes) and fewer issues with measurement error or attrition compared to traditional survey tools. Card et al. (2011) point out the wealth of administrative data that is collected across a wide range of domains from tax records to schooling and how lack of access to it is threatening the dominant position of the US in economics research.

While government data is being increasingly opened up and made available, it often lacks variables or characteristics that would be useful or essential to the researcher. Crime data in the UK is a particular example of this, even though detailed crime statistics are publicly available. Not only is the individual level crime data difficult to obtain but also to match to other relevant characteristics.³ Bell et al. (2014) who examine the impact of tougher sentencing on crime require detailed geographical information, more than is typically provided in published statistics. Francesconi and James (2015) exploit timing of alcohol consumption and require information on the time when crimes are committed. Hanes and Machin (2014) require monthly counts of the ethnicity of victims in their investigation of the

² Hazell and Worthy (2010) assess performance of the FOI act in the UK in comparison to other countries by considering the total number of requests made over time, percentage of requests granted and whether refusals to provide data are appealed against.

³ For a more detailed discussion see Machin, Marie and Vujić (2011).

impact of terrorist attacks on hate crime. Each of these papers used FOI requests to gather administrative data that is not routinely published.

In this paper we explain how researchers can make such a request and provide tips on how to make that request more successful. We then document the determinants of one particular FOI request we made - for data on hate crime from all police forces in the UK. We do this because, although public authorities are legally obliged to respond to FOI requests and provide data within 20 working days, they do not always do so. They may refuse to provide the data on the grounds of cost, or may not have the requested data. One concern with obtaining data from FOI requests could be that those authorities that respond might be systematically different from those which do not and therefore the researcher obtains a nonrepresentative sample. The aim here is to see whether there is any differential selection with respect to provision of data along observable characteristics.

We do not find that observable characteristics of the local area or the police force determine whether the request was not fulfilled at all or in part, nor the speed at which it was responded to. This suggests that the data from a FOI request is broadly representative of the institutions from which it was requested and that results obtained from analysis using them should be externally valid.

In the next section, we explain the FOI law in the UK and illustrate how to make a FOI request (in the UK), while section III provides advice on how to make that request more successful. Section IV documents the determinants of a request we made to all police forces in the UK. Section V concludes.

II. How to make a FOI request

The Freedom of Information Act 2000 came into force on 1st January 2005 in the UK. It provides access to information public authorities hold by making them publish information about their actions and by allowing the public to request information from them. The public authorities covered by the Act include: central and local government, the National Health Service (NHS), maintained schools and other educational institutions, police, and other public bodies and offices. Requests must be made in writing. Authorities typically have a contact email address or an online application form through which the requests can be made. These can be found on dedicated FOI webpages as part of the authority's website.⁴ While it is not necessary to state that the request is made under the FOI act, doing so helps avoiding any confusion and sets the clock ticking. Requests must be acknowledged and answered within 20 working days. If clarification is required, the clock is re-set; the countdown begins again from the day clarification is received.

Before making a request, it is important to explore whether the required data can be found elsewhere. This is because the commitment by various governments to open up their data has led to a vast number of data sets being publicly available. Furthermore, the data of interest to researchers may have been previously requested. A good starting point is the website data.gov.uk which contains data sets from all central government departments, local authorities and various public sector bodies.⁵ Many organisations also publish the responses to previous FOI requests in their disclosure logs.⁶ Therefore it is useful to search through these prior to making a request.

There are several constraints that the researcher faces when making a FOI request. First, knowing whether the data required is available is sometimes difficult to ascertain. As described above, a prior search of the disclosure logs or the central government data collection website (data.gov.uk) can help narrow this down. However, the answer may still not be obvious. Therefore, in order to improve the success of a request, particularly when contacting multiple authorities with the same request (in our example we contacted every

⁴ See for example: <u>www.essex.police.uk/about-us/freedom-of-information/</u>

⁵ At the time of writing <u>data.gov.uk/data</u> contains 38,762 published datasets.

⁶ For example see: <u>www.bbc.co.uk/foi/publication-scheme/classes/disclosure-logs</u>

police force in the country), we recommend choosing one or two areas to pilot the request. This piloting will allow one to see whether the data is collected and held by the authority (department). In the case of requests to only one department an informal email asking about data availability might suffice.

The second major constraint is cost. A request can be refused if the cost of releasing the data exceeds £600 for central government, Parliament and the armed forces and £450 for other public authorities. This threshold is based on a standard rate of £25 per person per hour, meaning that a request can be refused if more than 18 hours are needed to complete it. The piloting will also allow the researcher to gauge whether the request is within the cost limits of the FOI Act.

The format of the data poses the third constraint. Obtaining data from a number of public authorities involves sending multiple requests and various institutions collect and catalogue data in different ways. This heterogeneity of the data collection process makes it unlikely that data requested is provided in the same format by all institutions. Therefore, we suggest providing the authority with an example of the format that is required, i.e., an empty excel sheet with an exemplary data format.

If a FOI request is denied, the authority will reply setting out the reasons for refusal. If the provided reasons are not satisfactory, one can in first instance ask for an internal review of the decision (by writing back to the authority refusing the request) and subsequently appeal to the Information Commissioner's Office (ICO).⁷ At each stage, the initial request and responses will be reviewed. However, the process is likely to be lengthy.⁸

⁷ The ICO is an independent regulatory office that deals with Data Protection Act 1998, Privacy and Electronic Communications (EC Directive) Regulations 2003, the Freedom of Information Act 2000, and the Environmental Information Regulations 2004.

⁸ A third of complaints took over three months to resolve in 2015/16: <u>https://ico.org.uk/about-the-ico/our-information/annual-operational-reports-201617/foi-complaints-and-appeals/</u>

III. Determinants of Response to a FOI request

In this section we examine the determinants of response to one FOI request that we have undertaken. When using data obtained by FOI requests, a researcher may worry that authorities that replied might be systematically different from these which did not; therefore selection bias may be a concern. This exercise is intended to establish whether any particular characteristic makes an authority more likely to respond. We recommend that researchers using FOI to gather data compare responders and non-responders to examine if replies are driven (at least) by observable characteristics.

On March 2nd 2017 we contacted 47⁹ police forces in the UK with an identical request¹⁰ to provide monthly statistics on the reported number of hate crimes by type of crime, by ethnicity and by nationality of the victim covering the period from January 2011 to February 2017. The requests were randomly allocated between the three authors with two exceptions.¹¹ The characteristics of the response across areas are provided in Table 1. Specifically, the response rate was 98%, with one police force not replying. The request was completed on average within 20 days, the exact time within which institutions are obliged to respond. In 28% of cases the request was completed late, usually due to the need to follow up with clarification. The police forces responded in three ways – by refusing to share any information (11%) due to cost or data unsuitability, by providing some data and by providing all data. When only partial information was sent back, it was typically justified by lack of data or the cost of providing additional data being greater than the cost limits set out in Section III. Among the successful responses, some data provided was not as requested.

⁹ We contacted but dropped from the main analysis Doverport and British Transport Police as they are nonterritorial police forces.

¹⁰ The full text of the request can be found in the supplementary material.

¹¹ One exception was the Essex police force, which was contacted as a pilot area to determine the feasibility of the request. The second exception was Avon and Somerset police force; the police force was contacted by Jonathan James to determine whether being contacted by an academic based locally might increase the likelihood of a positive response.

We want to investigate whether these differential responses were determined by either characteristics and circumstances of the police force or of the area it was covering. We examine two outcomes – i) whether the police force provided any data (*sensu lato*) and ii) whether the provided data was what we requested (e.g., monthly and not quarterly or annual) (*sensu stricto*). Among the determinants of the FOI request we consider police force characteristics including total size, overall funding and staff available to process requests. Forces with more funding might be better placed to deal with the FOI requests and may be less inclined to divert funds from administrative staff towards more front-line policing. The likelihood of the request being completed and on time should increase with the size of the administrative team. We also examine local area characteristics such as the demographic and ethnic make-up of the area and the proportion of hate crime offences in the period prior to when the request was made. It could be the case that more diverse areas have experienced greater exposure to hate crime and as such have already implemented better data collection practices making it easier to access. Finally, we examine the local labour market conditions of the area.

Table 2 presents the results of this analysis with two different dependent variables – whether any data was provided and whether the right data was provided. We present the marginal effects from a set of probit regressions.¹² In columns 1 and 6 we only include the police force characteristics – total size, number of admin staff and funding of the police force.¹³ We find no relationship between provision of right data and the police force characteristics. Police force size is positively correlated with any data provision. Next, we examine whether previous hate crimes are correlated with the response to the FOI request

¹² All regressions include the variables with coefficients reported in the table as well as two dummy variables indicating the contact person on the research team responsible for the request.

¹³ Note that the regressions are run on a sample of 45 police forces, as two police forces (British Transport Police and Doverport) are non-territorial. Hence, there are no local area characteristics available for them.

(columns 2 and 7).¹⁴ We find no relationship. Then we consider the demographic characteristics of the area (columns 3 and 8) and find correlations between provision of data and the share of population who are white (+), the share of population who are of non-Christian religion (+) and the share of population of working age (–). There is no relationship between these characteristics and provision of right data. Lastly, we consider the role played by the local labour market characteristics (columns 4 and 9) and find no relationship with either of the dependent variables. When we include these variables together in a regression (columns 5 and 10), all coefficients become statistically insignificant. Overall, we conclude that the above characteristics do not determine the responsiveness of police forces to our FOI request.

In the supplementary material we also provide results of the regressions using two alternative measures: i) whether the police force provided all requested data (i.e., number of hate crimes by type, by ethnicity of victim and by nationality of victim), and ii) whether the request was completed late (i.e., after the 20 working day target). No characteristics we study seem to determine late provision of data. We find positive and marginally significant correlations between the local population characteristics and provision of all requested data. As an additional check, we have also dropped two London-based police forces (Metropolitan and City) from the main regressions. This is because due to the London allowance wages paid to administrative staff, processing the requests in London increases the cost of complying. The regression results considering provision of any data remain unaltered. In regressions considering provision of right data, when all controls are included together, the coefficients on some of the local population characteristics become significant but only at 10% level.

¹⁴ We have also used a more general measure of crime (i.e., total crimes committed per 100 population) instead of the hate crime variable. The regression results remain unchanged.

IV. Conclusion

We have put forward what we believe to be an underused (by economists) method of collecting a potentially unique data set – a Freedom of Information request. We briefly describe how to make a request with a focus for economists who in the main will be requesting the data. Our two main tips to improve the success of requests are i) to pilot the request and ii) provide the authority with an example of the format that is required. Finally, we examine the drivers of one request that was made to all police forces in the UK. We find that the observable characteristics are not significantly correlated with the probability that a request was satisfied, suggesting that the data from this FOI request are broadly representative of the institutions from which it was requested.

References

Bell, B., Jaitman, L. and Machin, S., 2014. Crime deterrence: Evidence from the London 2011 riots. *The Economic Journal*, 124(576): 480-506.

Card, D., Chetty, R., Feldstein, M., and Saez, E., 2010. Expanding Access to Administrative Data for Research in the United States. NSF SBE 2020 White Paper, National Science Foundation Directorate of Social, Behavioral, and Economic Sciences, Arlington, VA.

Francesconi, M. and James, J., 2015. The Cost of Binge Drinking. IZA Discussion paper 8849.

Hamermesh, D.S., 2013. Six decades of top economics publishing: Who and how?. *Journal of Economic Literature*, 51(1): 162-172.

Hanes, E. and Machin, S., 2014. Hate crime in the wake of terror attacks: Evidence from 7/7 and 9/11. *Journal of Contemporary Criminal Justice*, 30(3): 247-267.

Hazell, R. and Worthy, B., 2010. Assessing the performance of freedom of information. *Government Information Quarterly*, 27:352-359.

Machin, S., Marie, O. and Vujić, S., 2011. The crime reducing effect of education. *The Economic Journal*, 121(552), 463-484.

Table 1: Summary statistics

Panel A. Descriptives			Panel B. Justification provided			
				Number of police forces		
Number of police forces	47		Reasons for refusal to provide data overall			
Request response rate	98%		cost	5	5	
Follow up / clarification required	28%		not appropriate for this research	1		
Late request completion	28%					
average completion time (working days)	19.8		Reasons not all data provided			
Refused data provision	11%		information not held	5		
Provided data of some kind	89%		cost	2		
all requested data	44.7%					
most (or all) requested data	78.7%		Reasons data provided not suitable			
unsuitable data	10.3%		wrong information (e.g. annual or quarterly)	3		
			police force not territorial	1		
Panel C. Police force characteristics			Panel D. Local area characteristics			
	mean	st.dev.		mean	st.dev.	
Total police force	6111.96	12600.76	Disposable income per head in 2016	18640.83	2573.63	
Total administrative staff	83.54	93.77	Unemployment rate (16-64 year olds) in 2016	4.49	1.07	
Administrative staff (per 100 police force)	2.02	0.81				
Administrative staff (per 1 inhabitant)	1.69	6.23	% population working age	63.29	2.62	
Number of non-white employees (per 100 police force)	1.50	1.63	% population UK born	88.37	14.97	
			% population white	91.63	9.08	
Funding (per 10000 population)	2771296	8446972				
			% population Christian	62.78	6.88	
Total offences (per 100 police force)	2210.78	643.65	% population no religion	25.19	5.19	
Reported hate crimes, 2014-2015 (per 100 population)	0.101	0.11				
Reported racist incidents, 2014-2015 (per 100 population)	0.112	0.102	Migration flow (per 100 population)	0.32	0.26	

Note: Data come from the following sources: police force characteristics and crime statistics - Home Office; local area demographic characteristics - 2011 Census; labour market characteristics - ONS. Information about the police forces refers to year 2015/2016, the most recent data available. The crime statistics capture year 2014/2015. The local area demographics reflects the situation at the time of the 2011 Census. The reference year for labour market characteristics is 2015.

Dependent variable			Any Data Right Data							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Police force characteristics										
size of the police force	0.00004**				0.000002	1.209				2.520
	(0.00002)				(0.00006)	(0.928)				(2.515)
size of admin team (per 100										
population)	-3.812				3.979	-36.396				-74.989
	(7.355)				(18.547)	(22.095)				(63.694)
total funding (per 10000	0.000002				-0.00004	- 0.000007				-0 00004
population)	0.000002				0.00001	(0.00002				0.00001
	(0.000005)				(0.0001))				(0.00003)
Incidence of hate crime										
number of hate crimes (per										
100 population)		-0.486			7.622		-0.566			0.804
T I I		(0.416)			(16.240)		(0.508)			(2.979)
characteristics										
% population working age			-0.079**		-0.105			-0.045		0.034
			(0.031)		(0.208)			(0.043)		(0.063)
% population white			0.085**		0.285			0.049		0.215
			(0.035)		(0.521)			(0.043)		(0.202)
% population non-UK born			-0.007		0.009			0.014		0.050
			(0.016)		(0.026)			(0.026)		(0.063)
% population of non-Christian										
religion			0.174***		0.543			0.064		0.249
			(0.062)		(0.987)			(0.066)		(0.234)
Local labour market										
disposable income per head				-0.00002	-0.0001				-0.00002	-0.000001
				(0.00002)	(0.0001)				(0.00003)	(0.00004)
unemployment level				-0.00000006	0.000002				-0.0000003	0.000004
				(0.000001)	(0.00003)				(0.000001)	(0.000003)
n	45	45	45	45	45	45	45	45	45	45

Table 2. The determinants of the response to a FOI request of all police forces in the UK

Note: robust standard errors in parentheses. *, ** and *** respectively denote statistical significance at the 10, 5 and 1% level. The dependent variables is a dummy equal to 1 if any of the requested data were provided. The reported coefficients are marginal effects from probit regressions. Data sources: FOI requests, 2011 Census, Home Office and ONS.