



**EU Grant Agreement number: 645852**

**Project acronym: DIGIWHIST**

**Project title: The Digital Whistleblower: Fiscal Transparency, Risk Assessment and the Impact of Good Governance Policies Assessed**

Work Package: 2 – Data Collection and Cleaning

**Title of deliverable:**  
**D2.4 – Raw data**

Due date of deliverable: 29/2/2016  
Actual submission date: 31/05/2016

Author: Jan Hrubý, Mihály Fazekas, Tomáš Pošepný, Jakub Krafka, Michal Říha, Marek Mikeš, Tomáš Mrázek

Organization name of lead beneficiary for this deliverable:  
University of Cambridge-Sociology

Dissemination Level		
<b>PU</b>	Public	X
<b>PP</b>	Restricted to other programme participants (including the Commission Services)	
<b>RE</b>	Restricted to a group specified by the consortium (including the Commission Services)	
<b>Co</b>	Confidential, only for members of the consortium (including the Commission Services)	

All rights reserved. This document has been published thanks to the support of the European Union's Horizon 2020 research and innovation Programme under grant agreement No 645852.

The information and views set out in this publication are those of the author(s) only and do not reflect any collective opinion of the DIGIWHIST consortium, nor do they reflect the official opinion of the European Commission. Neither the European Commission nor any person acting on behalf of the European Commission is responsible for the use which might be made of the following information.

## Objective

The main goal of this deliverable is to obtain unstructured and semi-structured data from a pool of 100+ potentially relevant datasources where further data processing is assessed to be plausible. In order to determine the right strategy for data scraping, this also requires analysis of the architecture and structure of online data portals reporting data on public procurement tenders, private and public entities such as companies, public sector organisations' budgets, asset declarations, and political office holders.

The main outputs are therefore the technical evaluation of the quality of individual datasources and the collection of raw data from those chosen for further processing. As stressed in the project proposal, this task required the prioritisation of the better-structured and more policy-relevant sources. The selected sources and thus also raw data outputs represent a solid basis for achieving the project's research goals.

## Technical solution

Regardless of the dataset type, there are common issues in extracting data from external sources. A unified general approach has therefore been used across dataset types. Based on the mapping of individual sources for each jurisdiction, provided by project partners, the UCAM team has explored several approaches to web crawling and raw data storage. The main issues we had to deal with were:

1. a variable internet environment which allows many different technical solutions for source applications from pure HTML to advanced javascript clients;
2. the performance of a whole data scraping system that requires sending millions of requests to tens of sources simultaneously; and
3. changing the data model for each source in order to store not only raw data itself but also much additional information obtained during a web crawling.

Initially the University of Cambridge Computer Laboratory (UCAM-CL) attempted to create an innovative new method to automatically gather the data from all jurisdictions using a single data collection tool based on machine learning. However, after nine months, it was clear that this methodology could not deliver data with the required quality and scope within a timeframe that allowed the rest of the project to progress. Consequently, in the interests of the wider project, it was agreed to revert to a more traditional approach whereby various methodologies are used according to the different ways that countries store their public procurement data - a method project partners have worked with successfully in the past. This means implementing a source-specific crawler for each source application which benefits from our detailed knowledge of such source. Our crawlers use an HTMLUnit technology which is an implementation of a web browser without a graphical user interface developed for Java programs. This technology enables us to simulate the user's behaviour and helps us to overcome many difficulties associated with HTML being obscured with Javascript or AJAX elements in a web page or solving cookies issues.

To solve a performance issue when we have to download data from millions of URLs we have chosen a microservice architecture pattern which allows us to benefit from the possibilities that a cloud environment offers. Each crawler is a small stand-alone application (service) that is designed specifically for one purpose. We can run multiple instances of one service in parallel, which enables us to boost the performance of any task. Communication between services is ensured by the messaging system RabbitMQ which assigns work to each running service and this way balances a workload of individual instances. This architecture has proved to be very efficient when we run it in an Amazon web service environment.

Raw data does not contain only HTML pages itself but in many cases it contains additional information that we also need to store. Such information can be search form values, HTML snippets, cookies value, dates etc. The structure of such data differs from source to source and we want to keep it so, without any transformation. MongoDB is a database which is based on storing documents instead of relational data and such a document can be in any structure. This way we can create a mirror of a source without losing any information and without the need to adjust the relational database model whenever we need to store some new information.

Besides using the web scraping technologies described above, we obtained some additional datasets from our partners following the project plan, such as procurement data from Romania. Such raw data are stored within the system as well.

## **Data description**

This section describes data that we downloaded and stored for further processing. Where possible we are continuing with incremental scraping (e.g. on daily basis) and so our database continues to grow.

- *Public procurement data*

The main goal for this deliverable is to implement a set of web crawlers that will be able to effectively search for procurement data and store it in its original format. We have implemented 21 scrapers that cover TED + 20 national sources. One scraper is still under development. We have also incorporated several CSV files that contain a limited amount of data for some countries.

Table 1: Public procurement data coverage for individual countries

Country	Source	Number of downloaded tender records*
Armenia	Not covered	
Austria	Covered by TED	
Belgium	National source	77,340
Bulgaria	In progress	
Croatia	National source	216,601
Cyprus	Covered by TED	
Czech Republic	National source	300,138
Denmark	Covered by TED	
Estonia	National source	155,221
European Commission	Covered by TED	
Finland	Covered by TED	
France	National source	215,548
Georgia	National source	162,126
Germany	Covered by TED	
Greece	Covered by TED	
Hungary	National source	273,294
Iceland	Not covered	
Ireland	National source	70,970
Italy	National source	29,081**
Latvia	National source	118,611
Lithuania	National source	126,554
Luxembourg	Covered by TED	
Malta	Covered by TED	
Netherlands	National source	41,519
Norway	National source	180,023
Poland	National source	2,682,855
Portugal	National source	697,368
Romania	National source	871,537
Serbia	National source	41,508
Slovakia	National source	129,712
Slovenia	National source	150,080
Spain	National source	318,271
Sweden	Covered by TED	
Switzerland	National source	94,813
United Kingdom	National source	431***

\* Our database of TED forms currently contains 2,341,141 records

\*\* XML files containing multiple tender records

\*\*\* Large XML files containing hundreds of records

Table 2 contains links to downloadable reports. Each report contains links to individual tenders or announcements that we downloaded into our database. These links direct to original data on the internet, not to our database, because we store data in human unreadable format and we also have our database highly secured since we also store confidential data. We do not have this archive for Romanian data because we obtained these data from our partners therefore we do not have links to web pages.

Each archive that is linked from Table 2 is encrypted by standard encryption software GnuPG. To decrypt archives on Windows, the installation of software like gpg4win (<https://www.gpg4win.org/download.html>) is required. Linux distributions and Mac operation systems usually have this software built-in. After installation of gpg4win, the decrypt function is available in a context menu. Decryption also requires a password which we provide on demand by emailing [digiwhist.aws@gmail.com](mailto:digiwhist.aws@gmail.com).

*Table 2: Public procurement data overview*

Country	Link
Belgium	<a href="http://digiwhist-data.s3-website.eu-central-1.amazonaws.com/D2_4/tender/be.csv.gz.gpg">http://digiwhist-data.s3-website.eu-central-1.amazonaws.com/D2_4/tender/be.csv.gz.gpg</a>
Croatia	<a href="http://digiwhist-data.s3-website.eu-central-1.amazonaws.com/D2_4/tender/hr.csv.gz.gpg">http://digiwhist-data.s3-website.eu-central-1.amazonaws.com/D2_4/tender/hr.csv.gz.gpg</a>
Czech Republic	<a href="http://digiwhist-data.s3-website.eu-central-1.amazonaws.com/D2_4/tender/cz.csv.gz.gpg">http://digiwhist-data.s3-website.eu-central-1.amazonaws.com/D2_4/tender/cz.csv.gz.gpg</a>
Estonia	<a href="http://digiwhist-data.s3-website.eu-central-1.amazonaws.com/D2_4/tender/ee.csv.gz.gpg">http://digiwhist-data.s3-website.eu-central-1.amazonaws.com/D2_4/tender/ee.csv.gz.gpg</a>
France	<a href="http://digiwhist-data.s3-website.eu-central-1.amazonaws.com/D2_4/tender/fr.csv.gz.gpg">http://digiwhist-data.s3-website.eu-central-1.amazonaws.com/D2_4/tender/fr.csv.gz.gpg</a>
Georgia	<a href="http://digiwhist-data.s3-website.eu-central-1.amazonaws.com/D2_4/tender/ge.csv.gz.gpg">http://digiwhist-data.s3-website.eu-central-1.amazonaws.com/D2_4/tender/ge.csv.gz.gpg</a>
Hungary	<a href="http://digiwhist-data.s3-website.eu-central-1.amazonaws.com/D2_4/tender/hu.csv.gz.gpg">http://digiwhist-data.s3-website.eu-central-1.amazonaws.com/D2_4/tender/hu.csv.gz.gpg</a>
Ireland	<a href="http://digiwhist-data.s3-website.eu-central-1.amazonaws.com/D2_4/tender/ie.csv.gz.gpg">http://digiwhist-data.s3-website.eu-central-1.amazonaws.com/D2_4/tender/ie.csv.gz.gpg</a>
Italy	<a href="http://digiwhist-data.s3-website.eu-central-1.amazonaws.com/D2_4/tender/it.csv.gz.gpg">http://digiwhist-data.s3-website.eu-central-1.amazonaws.com/D2_4/tender/it.csv.gz.gpg</a>
Latvia	<a href="http://digiwhist-data.s3-website.eu-central-1.amazonaws.com/D2_4/tender/lv.csv.gz.gpg">http://digiwhist-data.s3-website.eu-central-1.amazonaws.com/D2_4/tender/lv.csv.gz.gpg</a>
Lithuania	<a href="http://digiwhist-data.s3-website.eu-central-1.amazonaws.com/D2_4/tender/lt.csv.gz.gpg">http://digiwhist-data.s3-website.eu-central-1.amazonaws.com/D2_4/tender/lt.csv.gz.gpg</a>
Netherlands	<a href="http://digiwhist-data.s3-website.eu-central-1.amazonaws.com/D2_4/tender/nl.csv.gz.gpg">http://digiwhist-data.s3-website.eu-central-1.amazonaws.com/D2_4/tender/nl.csv.gz.gpg</a>
Norway	<a href="http://digiwhist-data.s3-website.eu-central-1.amazonaws.com/D2_4/tender/no.csv.gz.gpg">http://digiwhist-data.s3-website.eu-central-1.amazonaws.com/D2_4/tender/no.csv.gz.gpg</a>
Poland	<a href="http://digiwhist-data.s3-website.eu-central-1.amazonaws.com/D2_4/tender/pl.csv.gz.gpg">http://digiwhist-data.s3-website.eu-central-1.amazonaws.com/D2_4/tender/pl.csv.gz.gpg</a>
Portugal	<a href="http://digiwhist-data.s3-website.eu-central-1.amazonaws.com/D2_4/tender/pt.csv.gz.gpg">http://digiwhist-data.s3-website.eu-central-1.amazonaws.com/D2_4/tender/pt.csv.gz.gpg</a>
Slovakia	<a href="http://digiwhist-data.s3-website.eu-central-1.amazonaws.com/D2_4/tender/sk.csv.gz.gpg">http://digiwhist-data.s3-website.eu-central-1.amazonaws.com/D2_4/tender/sk.csv.gz.gpg</a>
Slovenia	<a href="http://digiwhist-data.s3-website.eu-central-1.amazonaws.com/D2_4/tender/si.csv.gz.gpg">http://digiwhist-data.s3-website.eu-central-1.amazonaws.com/D2_4/tender/si.csv.gz.gpg</a>
Spain	<a href="http://digiwhist-data.s3-website.eu-central-1.amazonaws.com/D2_4/tender/es.csv.gz.gpg">http://digiwhist-data.s3-website.eu-central-1.amazonaws.com/D2_4/tender/es.csv.gz.gpg</a>
Switzerland	<a href="http://digiwhist-data.s3-website.eu-central-1.amazonaws.com/D2_4/tender/ch.csv.gz.gpg">http://digiwhist-data.s3-website.eu-central-1.amazonaws.com/D2_4/tender/ch.csv.gz.gpg</a>
TED	<a href="http://digiwhist-data.s3-website.eu-central-1.amazonaws.com/D2_4/tender/eu.csv.gz.gpg">http://digiwhist-data.s3-website.eu-central-1.amazonaws.com/D2_4/tender/eu.csv.gz.gpg</a>
United Kingdom	<a href="http://digiwhist-data.s3-website.eu-central-1.amazonaws.com/D2_4/tender/uk.csv.gz.gpg">http://digiwhist-data.s3-website.eu-central-1.amazonaws.com/D2_4/tender/uk.csv.gz.gpg</a>

- *Company data*

In order to gain powerful insights into the workings of public procurement markets, information on the participating companies also has to be available. Consequently, in addition to the collection, scraping, parsing and organization of public procurement data by DIGIWHIST, company data also had to be obtained. There are four relevant variable groups to be obtained: company registry information (company name, ID, incorporation date, address, company size etc.), financial data (annual turnover, profit rate, liabilities etc.), ownership and manager information. However, in most European countries, there is no readily available and detailed company data. Although, national company registries exist, they are not always free to use and often only contain a limited set of information (e.g. no ownership or financial data is available). Furthermore, open data repositories on company characteristics do not contain enough data either (e.g. opencorporates.com), hence they can be only used for cross-checking data quality. Therefore, the full company data set from all 34 jurisdictions was purchased from a private data provider.

After an initial screening of possible data providers (e.g. Bisnode, Bureau van Dijk), a public procurement tender (in line with both EC and UCAM procurement rules) was launched after agreeing on a comprehensive list of variables and the related details by all DIGIWHIST partners and the public procurement experts at UCAM. As a result of the open tender a three-phased data delivery plan was agreed with Bureau van Dijk, whereby the first data dump was provided in January 2016. After verifying diligent delivery, the work on data cleaning, structuring and its preparation for linking and analysis has been already started by UCAM.

Table 3: Company data overview

Country	Number of unique companies included in the registry dataset	Number of companies having ownership information from 2014	Number of companies having ownership information between 2009-2014	Number of companies with turnover data	Number of companies with P/L information	Number of companies having manager information
Armenia	531	472	791	31	31	30
Austria	886,040	648,741	812,591	68,401	9,770	447,343
Belgium	2,692,093	669,465	699,446	31,939	473,442	563,407
Bulgaria	1,472,697	1,537,690	1,687,847	523,021	272,707	1,207,911
Croatia	142,024	106,600	128,261	118,735	118,733	113,315
Cyprus	137,044	50,223	122,071	1,019	1,102	64,459
Czech Republic	2,326,746	864,195	1,070,649	1,587,391	201,885	363,823
Denmark	1,067,331	562,396	692,808	41,177	279,497	271,520
Estonia	228,909	80,052	176,401	106,620	123,936	184,139
Finland	1,298,559	137,341	210,430	452,489	208,283	597,238
France	2,588,761	1,381,614	2,462,110	1,511,319	1,318,490	1,439,498
Georgia	65,143	969	1,293	58,713	19	136
Germany	3,036,657	3,295,539	1,534,004	536,124	119,703	1,555,043
Greece	139,333	41,140	128,488	41,375	41,410	46,450
Hungary	1,519,752	722,053	929,411	446,938	507,947	522,481
Iceland	45,159	40,146	53,302	13,420	32,233	30,142
Ireland	578,113	172,425	239,784	26,908	29,978	191,697
Italy	4,378,031	1,605,773	2,163,064	3,115,346	1,275,359	3,398,294
Latvia	352,394	315,247	361,990	165,060	123,115	178,254
Lithuania	160,997	17,727	39,865	78,341	11,167	119,288
Luxembourg	147,379	71,070	86,575	12,808	12,148	28,326
Malta	522,39	13,034	16,589	18,815	18,817	36,260
Netherlands	4,373,015	602,143	813,505	22,415	45,465	2,142,623
Norway	1,849,666	1,550,235	1,750,726	335,704	335,703	975,731
Poland	1,276,340	153,826	1,387,222	147,976	153,435	815,026
Portugal	647,470	545,777	766,791	375,229	425,910	355,923
Romania	2,554,976	2,219,806	2,576,759	780,580	780,575	853,712
Serbia	218,031	326,051	382,086	97,490	97,482	170,846
Slovakia	749,295	419,974	498,057	424,486	176,882	212,256
Slovenia	370,084	41,815	51,748	155,417	167,280	35,238
Spain	4,049,368	1,358,249	2,343,140	909,131	1,000,304	1,733,111
Sweden	1,787,506	225,148	289,067	1,233,528	466,094	475,871
Switzerland	788,286	479,400	636,075	502,418	869	560,023
United Kingdom	9,308,875	3,695,417	5,092,390	1,203,518	372,788	4,358,362

- *Public sector data*

Public sector data comprises three main datasets: contracting authorities; public officials; and budgets.

- i. Contracting authorities

This dataset is indispensable for cleaning and linking public procurement data. We mapped sources and decided to extract data from those that contain at least some of the following data fields: name, contracting authority ID, street, town, zip code, contact person, phone, email, fax, web, activity type or contracting authority type.

We used 3 different methods to obtain data for a specific country:

- Data extraction from PDF: Some sources provide data only in PDF format. This has obvious disadvantages since such data are not updated on a regular basis but we can still obtain an almost comprehensive list of contracting authorities.
- Scraping data from HTML source: We wrote a set of web crawlers that scrape raw data in HTML format and we further process such data to get structured information about contracting authorities. The most common tools for such processing are XPath expressions and Regular expressions. Sources that provide data in HTML format can be regularly checked for updates so that we do not miss any new records.
- Importing data from CSV files: For some countries we were able to get data in a structured form in CSV files. Therefore we implemented a tool that enables us to import such data directly into the database.

This has resulted in 16 datasets covering Austria, Bulgaria, Czech Republic, Germany, Greece, Hungary, Ireland, Lithuania, Malta, Netherlands, Norway, Portugal, Romania, Slovakia, Spain, Switzerland and United Kingdom.

From the outset we knew that we do not have a comprehensive list of all contracting authorities for each country, making the identification of procuring organisations challenging. Hence, we obtained datasets which will be an important input for linking and matching algorithms and may significantly improve its results because the more records we have, the more accurate our clustering algorithms can be.

Further plans for the design of cleaning mechanisms also involve the use of expertise within DG GROW, which has offered to collaborate with us on clustering algorithms currently developed for TED data. Their system uses two stages of clustering plus some machine learning based on user corrections.

Table 4: Contracting authorities scraping results

Country	Number of downloaded records	Country	Number of downloaded records
Armenia		Italy	20,339
Austria	5,739	Latvia	
Belgium		Lithuania	8,781
Bulgaria	6,032	Luxembourg	
Croatia		Malta	417
Cyprus		Netherlands	1528
Czech Rep.	5877	Norway	1,175
Denmark		Poland	
Estonia		Portugal	125,687*
European Union Institutions		Romania	586
Finland		Serbia	
France		Slovakia	1921
Georgia		Slovenia	
Germany	952	Spain	3,713
Greece	852	Sweden	
Hungary	14,166	Switzerland	2,785
Iceland		United Kingdom	5,362
Ireland	4904		

\* mix of buyers and suppliers

## ii. Public officials

To get public officials data we have implemented 4 web scrapers for 4 generic sources that cover multiple countries:

- CIA world leaders – a PDF file, published monthly, that contains information about current presidents, prime ministers and other ministers;
- Everypolitician.org - one of the world's biggest open databases of politicians;
- Political data yearbook – a database of presidents and ministers containing historical data;
- Rulers.org – a database containing also data for regional governors or big cities.

Beyond these 4 generic sources we have been able to map sources for public officials' information stemming from each individual country. Although the heterogeneity of these websites makes it difficult to extract information, we will keep these sources in case further complementarity is needed or we feel particular countries warrant more in-depth exploration.

Names and profiles of public officials will not be released following the advice of our ethics consultant. However, public officials data becomes relevant when producing aggregate and systemic measures of links between the public and the private sectors.

Asset declarations: regarding the collection of public asset declarations for public officials, we have worked at three levels: firstly, we have made use of the data collected for the EuroPAM database (WP1 and WP4) in order to understand which regulations different types of public officials and politicians are subject to regarding both asset and conflict of interest declarations. There is great heterogeneity and for many countries there is no obligation to present asset declarations, although requirements to declare conflicts of interest are more widespread. For each jurisdiction we have therefore identified the relevant link where either asset declarations or conflicts of interest are reported for the relevant public officers. Secondly, we screened for readily available country-level asset declaration statistics which were not available in any of the screened countries. Thirdly, as an alternative route to data collection, we assessed the possibility of automated web-crawling to obtain individual level data on which we would then report aggregate statistics. However, the screened sources turned out to be highly irregular in their 13 formats even within the same country, often using scanned PDFs which are impenetrable using computerized data collection. As a result, we estimate that for most countries DIGIWHIST will not be able to provide country-level statistics.

We used several criteria for selecting the countries where we deem asset declaration processing most feasible and desirable:

1. Quality of procurement data - we prefer countries where we have detailed procurement data;
2. Downloadability: we prefer sources from which asset declarations can be downloaded;
3. Machine readability of asset declaration data - HTML or structured PDF is preferred.

Table 5: Asset declaration decision matrix

Country	Procurement data from national source	Downloadable asset declarations	Machine readable asset declarations	Chosen for processing
Latvia	YES	YES	HTML	YES
Lithuania	YES	YES	HTML	YES
Slovakia	YES	YES	HTML	YES
United Kingdom	YES	YES	HTML	YES
EU Institutions	YES	YES	PDF	YES
Georgia	YES	YES	PDF	YES
Ireland	YES	YES	PDF	YES
Netherlands	YES	YES	PDF	YES
Norway	YES	YES	PDF	YES
Spain	YES	YES	PDF	YES
Switzerland	YES	YES	PDF	YES
Denmark	NO	YES	HTML	NO
Germany	NO	YES	HTML	NO
Finland	NO	YES	HTML	NO
Iceland	NO	YES	HTML	NO
Serbia	NO	YES	HTML	NO
Croatia	NO	YES	HTML+PDF	NO
Austria	NO	YES	PDF	NO
Bulgaria	YES	YES	NO	NO
France	YES	YES	NO	NO
Hungary	YES	YES	NO	NO
Italy	YES	YES	NO	NO
Poland	YES	YES	NO	NO
Belgium	YES	NO	NO	NO
Czech Rep.	YES	NO	NO	NO
Estonia	YES	NO	NO	NO
Portugal	YES	NO	NO	NO
Slovenia	YES	NO	NO	NO
Greece	NO	YES	NO	NO
Luxembourg	NO	YES	NO	NO
Romania	NO	YES	NO	NO
Cyprus	NO	NO	NO	NO
Sweden	NO	NO	NO	NO
Armenia	NO	NO	NO	NO
Malta	NO	NO	NO	NO

Based on this matrix we selected countries that have both downloadable declarations and procurement data of high quality. From those we had to exclude Latvia and Lithuania because we need to search for specific names on these sources. We will get these names later after we process data on public officials and we will implement scrapers at that point. This means we implemented web crawlers for four sources where we will be able to extract data from structured PDFs or HTML pages: the EC, United Kingdom, Slovakia and Georgia. For the remaining countries we were able to get large structured PDF files containing declarations for multiple public officials (e.g for all members of parliament at once). These countries are Ireland, Netherlands, Norway, Spain and Switzerland.

### iii. Budgets

Our mapping efforts revealed that in almost all countries there are no services or public portals that aggregate budget reports. We were only able to identify single documents with such information scattered over the internet. Such data are totally inappropriate for automatic downloading and we had to download all files manually. Furthermore these are mainly PDF documents that do not follow any template and contain information with different levels of detail hidden in few tables in large documents that also contain a lot of irrelevant information. It will therefore be very complicated, if not impossible, to find and implement a data extraction approach that can be applied to multiple sources. As a result we are forced to do this work manually or semi-manually with the help of tools like Tabula that can help us extract data from tables contained in PDF files.

We have sought a collaboration with a related Horizon 2020 funded research project working on this issue: [openbudgets.eu](http://openbudgets.eu). While this is as yet uncertain, we might be able to improve our results by collaborating with [openbudgets.eu](http://openbudgets.eu) on specific datasets in countries of common interest.

*Table 6: Budget data availability*

Country	Budget data	Country	Budget data
Armenia	NO*	Italy	YES
Austria	YES	Latvia	NO*
Belgium	YES	Lithuania	NO*
Bulgaria	NO*	Luxembourg	YES
Croatia	NO*	Malta	YES
Cyprus	NO*	Netherlands	YES
Czech Rep.	YES	Norway	NO*
Denmark	NO*	Poland	NO*
Estonia	NO*	Portugal	YES
European Union Institutions	YES	Romania	NO*
Finland	NO*	Serbia	NO*
France	YES	Slovakia	NO*
Georgia	NO*	Slovenia	NO*
Germany	YES	Spain	YES
Greece	NO*	Sweden	NO*
Hungary	YES	Switzerland	YES
Iceland	NO	United Kingdom	YES
Ireland	YES		

\* Sources have been mapped, but we foresee difficulties in data extraction