



OpenDataMonitor: insights and findings from the platform
and stakeholder research. [@opendatamonitor](https://opendatamonitor.org)

Amanda Smith [@ayymanduh](https://twitter.com/ayymanduh)

OpenDataMonitor is a monitoring, analytics and visualisation tool.

It addresses the need to give insight into available open data resources, and gain understanding as to how the open data ecosystem is developing.

It harvests metadata from existing local, national and pan-European data catalogues.

Purpose #1

support entrepreneurs looking for reusable data with
which to create new services

Purpose #2

show open data users what high quality open data looks like

Purpose #3

enable policy makers to better understand how other cities, regions and countries are adopting open data and the impact its having

Purpose #4

enable researchers to gain insights into open data
publishing patterns

We've designed metrics which focus on **quality** and **quantity** of open data, to **measure** and **monitor** the open data ecosystem.

- total number of catalogues harvested
- total number of countries harvested
- percentage of datasets with open licenses
- percentage of datasets that are machine readable
- percentage of datasets with complete core metadata
- software platform used
- and number of unique publishers.



European Data Catalogues Visualisations

Dashboard

- Open licenses
- Total vs open
- Unique publishers

COMPLEXITY LEVEL

- Basic
- Advanced
- Expert



Pending Catalogues

HARVESTED DATA CATALOGUES
70
From 226 Harvested Catalogues

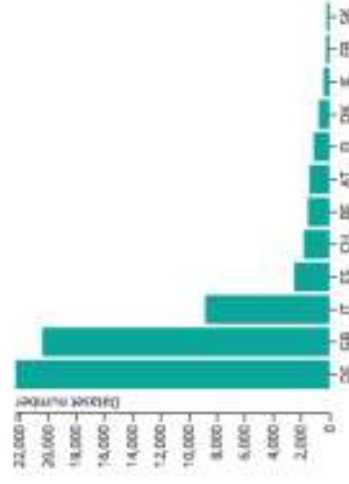
COUNTRIES HARVESTED
18
From 28 European Member Countries

OPEN LICENSES
23%
In 202,474 distributions with a license attached

MACHINE-READABLE
17%
Found in 381,735 distributions harvested

DATASETS HARVESTED
163,823

TOTAL SIZE
0.52 TB



Distributions

Total distribution

381,735

Distribution maximum size

37.8 MB

Distributions with a license attached

202,474

Distributions with an open license

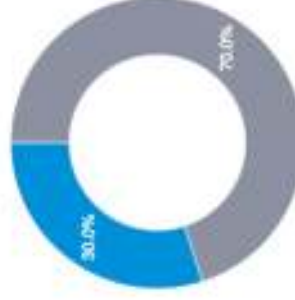
88,603

Distributions machine-readable

66,747

Software platform in the harvested catalogues

The OpenDataMonitor platform provides information and statistics on European open data catalogues, harvested through two methods. The first, OAH, is currently the most widely used open source data management system. In addition, for those catalogues that are not built on an open source management system, HTML scraping procedures are being used to extract information from websites. We are also exploring the feasibility of working with other publishing platforms to improve the coverage of data.



OAH HTML



European Data Catalogues Overview

Harvested catalogues ●
Pending catalogues ●

Harvested data catalogues

* Click to the eye icon of each row to view a specific catalogue.

Showing 1-10 of 70 items.



All



#	ACTIONS	TITLE	DESCRIPTION	PLATFORM	COUNTRY
		Search for a catalog	Search...		Search...

1		daten_hamburg.de	Open Data HamburgDie bisher unter dem Open Data Portal veröffentlichten Daten, finden Sie nun hier im Transparenzportal.	ckan	Germany
2		data.gov.uk	The government is opening up its data for other people to re-use. This is only about non-personal, non-sensitive data – information like the list of schools, crime rates or the performance of your council.To make sure that we do this properly and work with the web, this work is being...	ckan	United Kingdom
3		dati.gov.it	Il portale di Dati.gov.it è stato progettato e realizzato nel rispetto dei criteri di accessibilità indicati nel Decreto Ministeriale di attuazione della Legge Stanca (Legge 4/2004 'Disposizioni per favorire l'accesso dei soggetti disabili agli strumenti informatici') e alle linee guida WCAG 1.0 previste dallo standard internazionale WAI 'Web Accessibility Initiative....	ckan	Italy
4		govdata.de	Auf GovData, dem Datenportal für Deutschland sind Daten aller Verwaltungsebenen zentral zugänglich. Aktuelle Infos und Ähnliches finden Sie im Bereich 'Neues', dem Blog von GovData. Auf dem Weg zu „Open Data ‘Unser Ziel ist klar: Wir wollen Daten der öffentlichen Hand möglichst umfassend und offen zur Weiterverwendung zur Verfügung stellen....	ckan	Germany
5		open_data_europa.eu	The European Union Open Data Portal is the single point of access to a growing range of data from the institutions and other bodies of the European Union (EU). Data are free for you to use and reuse for commercial or non-commercial purposes.By providing easy and free access to data....	html	International



Alert Monitor Dataset List

An incomplete dataset is defined on the following missing metadata:

- License
- Author
- Organisation
- Date released
- Date updated

Incomplete datasets - 3.9%

* If you are an open data publisher, please go to your original dataset source to add the required metadata.

Showing 1-10 of 6,435 items.

 All

 2 -

#	ACTIONS	TITLE	CATALOGUE URL	PLATFORM	DESCRIPTION
1		Portal de Datos abiertos - Conjuntos de datos	 http://datos.madrid.es	html	Listado con el catálogo completo de conjuntos de datos que el Ayuntamiento de Madrid dispone de los ciudadanos.
2		Portal de Datos abiertos - Conjuntos de datos	 http://datos.madrid.es	html	Listado con el catálogo completo de conjuntos de datos que el Ayuntamiento de Madrid dispone de los ciudadanos.
3		Portal de Datos abiertos - Conjuntos de datos	 http://datos.madrid.es	html	Listado con el catálogo completo de conjuntos de datos que el Ayuntamiento de Madrid dispone de los ciudadanos.
4		Portal de Datos abiertos - Conjuntos de datos	 http://datos.madrid.es	html	Listado con el catálogo completo de conjuntos de datos que el Ayuntamiento de Madrid dispone de los ciudadanos.

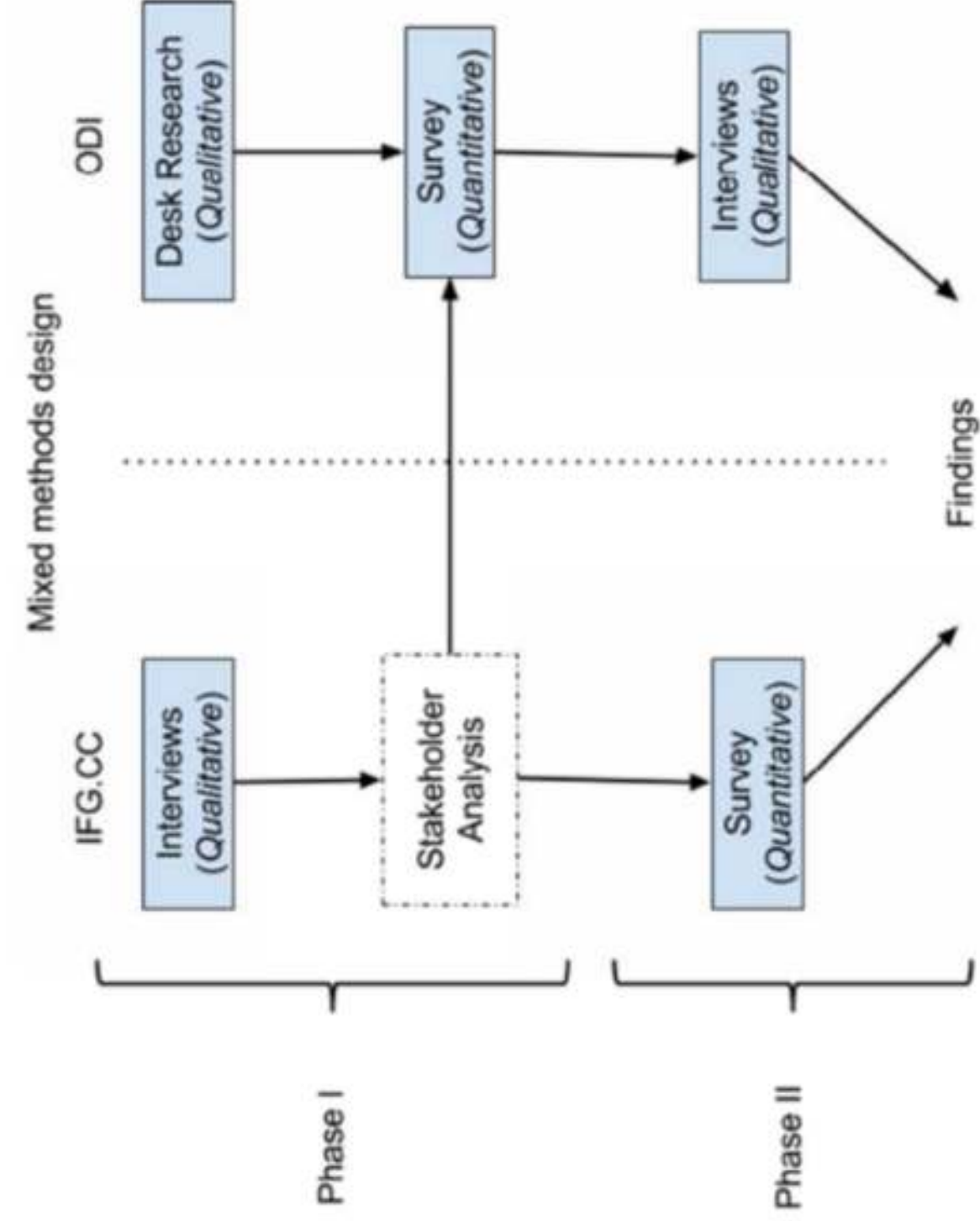
So far, we've discovered the following insights:

- **220 catalogues** identified across Europe to harvest from
- Currently collected data from **18 out of 28 countries**
- **Over 70 catalogues** harvested to date (a mix of national and local)
- **53%** of datasets have a license attached
 - **23%** of datasets have open licenses
- **17%** of datasets are machine readable
- **30%** of platforms use CKAN software
- **70%** use custom platforms that require HTML scraping



Our stakeholder research explored and analysed the roles, interests and understandings of the various stakeholders in the open data ecosystem.

As well as identifying the main stakeholders, we wanted to understand the core issues related to the publication and use of open data.



“Overall, 81 per cent of respondents have reportedly used an open data set in the past, despite the fact that only 26 per cent classified themselves as users.”

The support and push for open data comes from:

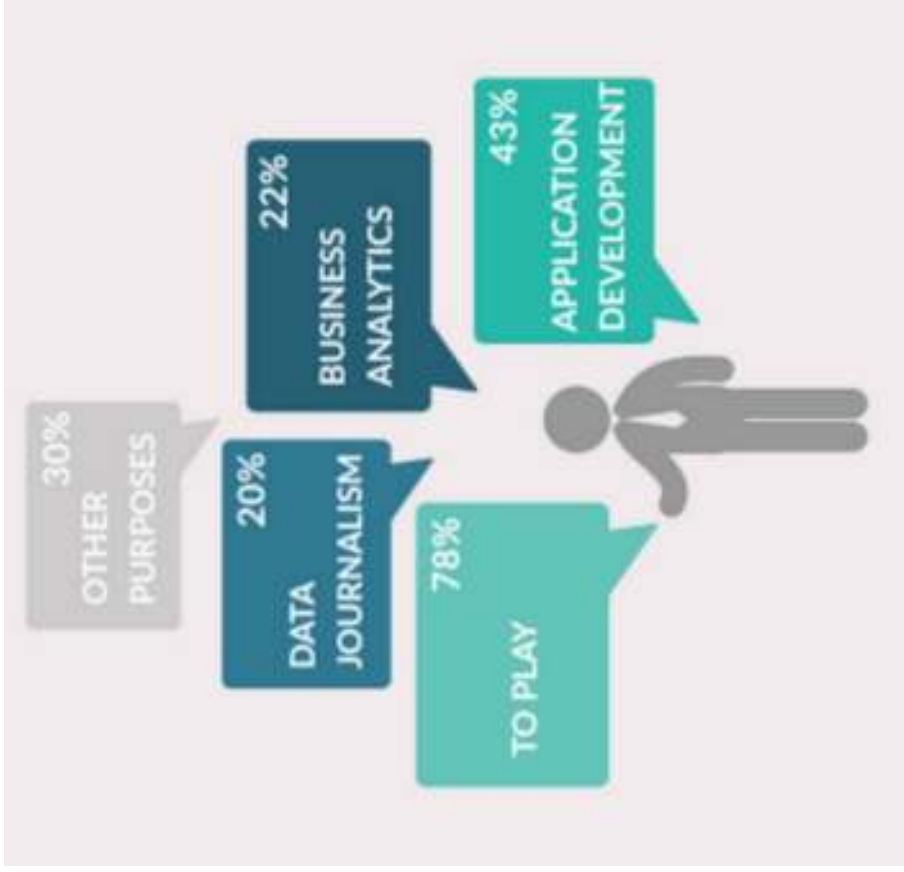
- civic and political activists / activist groups (79%)
- international bodies (53%)
- journalists and media industry (39%)
- senior and mid-level public managers (32%)
- politicians (29%)
- businesses and business associations (15%)

“In more than half of the surveyed cases, the software provides some sort of support to the public administrators who generate the data.

About 30% of the respondents need to rely solely on their manual ability to somehow generate open data. 12% were unaware of the software’s capability.”

“In about one fifth of the cases the software the software can support the manual export of data and by itself generate predefined metadata.”

Why use open data?



The most common “other purpose” specified in a free text field was “research” .

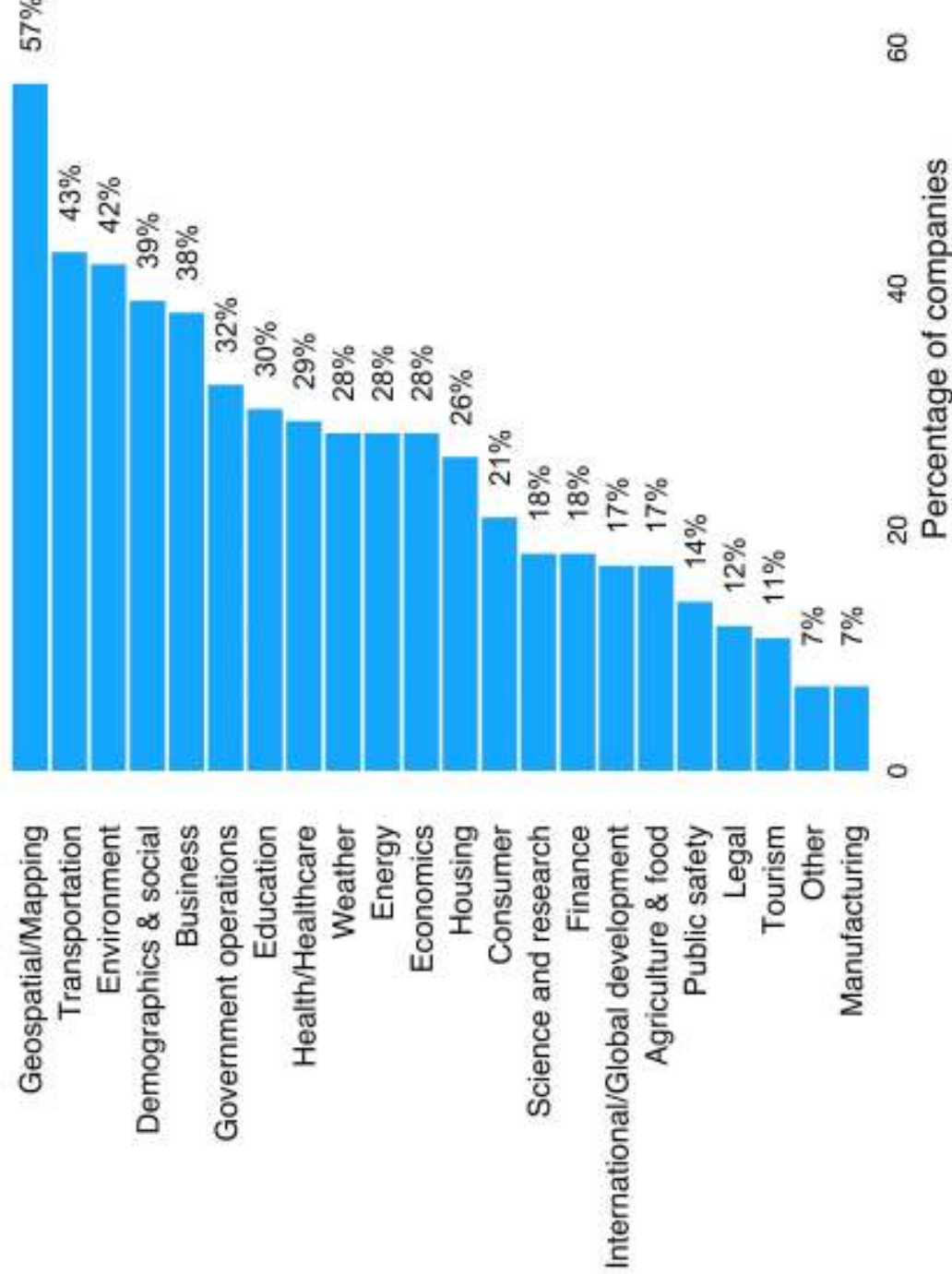
43% of public administrators and **67%** of consultants also used the data to develop an application.

Business analytics and data journalism were far less common purposes of open data use, but were used by **44% of consultants**.

Geospatial/mapping open data is used by more than half (57%) of all respondent companies.

Several other types are used by more than a third of respondents. **transportation (43%) environment (42%) demographic & social (40%) business (38%)**

79% of companies use data related to **more than one sector** **14%** of them use data related to **more than 10**.



“Insight generally comes from combining more than one dataset together. As each dataset becomes available as open data, the opportunities for taking that and coming up with completely new business models, and new ideas about what’s possible, are tremendous.”

~ Chris Taggart, OpenCorporates

Which **issues**, or **characteristics** influence the reuse of open data by companies?

80% found **accuracy of data** important, with **39% of companies** saying it had **great influence** on their decision to use open data.

Over half (55%) of the companies said that the way that datasets are **licensed** had a **great influence** on their **decision to use data**. **Only 8%** suggested that licensing had **little or no influence**.

65% found timeliness of data important, with 31% of companies saying it had great influence on their decision to use open data.

43% of the companies agreed the format of data important, however only 12% marked it as having a great influence, and 11% said it had little or no influence at all.

62% said that help and support from publishers wasn't important - only 3% agree that assistance had great influence on them.

“Data needs to be in a computer readable format. The format type does not matter.”

“There is little or no benefit to the user if they do not trust that the data is accurate and up to date.”

“It is notable that public data stores are ordinarily presented as 'separate' from corporate web pages, such that open data of relevance to specific policy objectives and/or services do not appear in the same place to promote awareness of them and/or facilitate ease of access for non-technical users.”

*“The art of it is to make it (the data) simple...
The general public out there are not interested in all the
rubbish out there, they just want to get to the
information they need to get to.”*



Noun icon by: [Luboš Volkov](#)



opendatamonitor.eu
project.opendatamonitor.eu
[@opendatamonitor](https://twitter.com/ayymanduh)
[@ayymanduh](https://twitter.com/ayymanduh)