

MOBIDATALAB

Labs for prototyping future mobility data sharing solutions in the cloud

Reference Data Catalogue Deliverable 4.4

Renee OBREGON, AKKODIS

mobidatalab.eu



MobiDataLab is funded by the EU under the H2020 Research and Innovation Programme (grant agreement No 101006879)

| CKAN

User demonstration



Discovering data in CKAN

- To discover and filter datasets on CKAN we search for keywords and tags.
- For example, we can search for a GTFS of Milan and explore its content details.

The screenshot shows a web browser window with the address bar displaying 'ckan.mobidatalab.eu'. The website has a dark blue header with the 'MOBIDATALAB' logo and navigation links for 'Datasets', 'Organizations', 'Groups', and 'About'. A search bar is located in the top right corner. Below the header, there is a 'Search data' section with a search box containing 'GTFS Milan' and a magnifying glass icon. Below the search box, there are 'Popular tags' including 'opendata', 'EU', and 'open data'. To the right of the search section, there is a promotional banner for 'CODAGON' by AETHON, featuring a person working on a laptop and text about a live co-creation event. Below the search section, there is a 'CKAN statistics' box showing '115.1k datasets', '275 organizations', and '25 groups'. The main content area displays a list of datasets, including 'OpenDataBerlin' and 'Mobility & Transport'. The footer contains links for 'About CKAN', 'CKAN API', and 'CKAN Association', along with a 'Powered by ckan' logo and a language selector set to 'English'.

Solr simple search

Modifiers for terms can be helpful:

- Include symbol +
- Exclude symbol -
- Phrase symbol “ ”

Datasets

Organizations

Open Data Vlaanderen 40

City of Leuven 7

De Lijn 2

Infrabel 2

TEC 1

Groups

Mobility & Transport 4

Traffic 4

Journey planning 2

+Leuven

52 datasets found for "+Leuven"

Order by: Re

Mobility Hubs Leuven

You will be able to find here the datasets provided by the city of Leuven for the x-thons of MobiDataLab.

gpkg

Parking data Leuven

Parking data for the city of Leuven.

JSON

Leuven -city

42 datasets found for "Leuven -city"

Order by: Relevance

Cycling network Leuven

This dataset has no description

GPKG

Traffic Accidents Leuven

This dataset contains all registered traffic accidents in Leuven between 2012 and 2021 Some further info on the included fields: -Field Comment -ID -Latitude -Longitude: WGS84...

GPKG

Origin-Destination data Leuven

OD Matrix Leuven 2017. The ids in the OD matrix refer to the zones defined in the GIS layer OD_zones_leuven.gpkg. This includes the whole of Belgium, though the zones get larger...

XL SX gpkg

"municipality of Eindhoven"

95 datasets found for ""municipality of Eindhoven""

Order by: Relevance

Tenders from the municipality of Eindhoven

Administration of the final documents associated with the various phases of a tender. Documents associated with a request include: assignment announcement, request,...

<https://www.iana.org/assignments/media-types/text/csv> <https://www.iana.org/assignments/media-types/application/json>
<https://www.iana.org/assignments/media-types/application/octet-stream> <https://www.iana.org/assignments/media-types/application/rdf+xml>
<https://www.iana.org/assignments/media-types/application/rd+json> <https://www.iana.org/assignments/media-types/text/turtle>
<https://www.iana.org/assignments/media-types/text/xml> <https://www.iana.org/assignments/media-types/application/xls>
<https://www.iana.org/assignments/media-types/application/vnd.openxmlformats-officedocument.spreadsheetml.sheet>

Tenders of the municipality of Eindhoven

Administration of the final documents associated with the various phases of a tender. Documents belonging to a request include: order announcement, request for tender,...

<https://www.iana.org/assignments/media-types/application/json> <https://www.iana.org/assignments/media-types/text/csv>
<https://www.iana.org/assignments/media-types/application/vnd.openxmlformats-officedocument.spreadsheetml.sheet>

Tenders of the municipality of Eindhoven

Administration of the final documents belonging to the different phases of a tender. Documents belonging to a request include: order announcement, request, specifications,...

<http://publications.europa.eu/resource/authority/file-type/JSON> <http://publications.europa.eu/resource/authority/file-type/CSV>

Trees Eindhoven

Trees in the municipality of Eindhoven

<http://publications.europa.eu/resource/authority/file-type/CSV>

Proclaimer Open data portal Eindhoven

Proclaimer Open data portal Eindhoven The open data portal of the municipality of Eindhoven (data.eindhoven.nl) is coming established under the responsibility of the

By default, all words or phrases specified by the user are treated as optional unless they are preceded by a “+” or a “-” symbol.

| Solr advanced search

Requires at least a colon (:)

Fielded search (field + colon (":") + term)

- **title: micro-mobility** this will look for all the datasets containing in its title the word “micro-mobility”
- **title:leuven || title:flemish** will look for datasets containing “leuven” or “flemish” in its title

Wildcards (“*”, proximity matching “~” terms)

- **title:share*** this will look for all the datasets containing in its title a word that starts with “share” like “shared” and “sharing”.
- **title: "pedestrian zone" ~ 4** A proximity search looks for terms that are within a specific distance from one another. This example will look for datasets which title contains the words “pedestrian” and “zone” within a distance of 4 words.

Fuzzy search (“~” symbol at the end of a single-word term)

- **text: yle~** this will find “cycle” and “CLE” terms in the text

title: micro-mobility 

1 dataset found for "title: micro-mobility"

Order by: Relevance 

Localization zones of the parking a
The dataset contains the localization zones

title: micro-mob~ 

10 datasets found for "title: micro-mob~"

Order by: Relevance 



Storage areas for micromobility offers

The data set contains the locations of the parking areas for micromobility offers.



Storage areas for micromobility offers

The database includes the previously set up parking areas in public and publicly accessible space for micromobility rental vehicles. Those using the rental vehicles should...

Storage areas for micromobility offers

The database includes the previously set up parking areas in public and publicly accessible space for micromobility rental vehicles. Those using the rental vehicles should...

Storage areas for micromobility offers

The map shows the previously set up parking areas in public and publicly accessible space for micromobility rental vehicles. Those using the rental vehicles should preferably...

Search results

- The datasets associated organization(s), group(s), tag(s), format(s) and license(s) will be displayed.
- The results can be filtered by restricting the search to datasets with particular tags, data formats, etc.

The screenshot shows the 'Datasets' page on the MobidataLab platform. On the left is a sidebar with filters: Organizations (Comune di Milano 1, Trenord 1), Groups (Mobility & Transport 1, Transport networks 1), Tags (GTFS 1, timetables 1, TPL 1), Formats (GTFS 2), and Licenses (Creative Commons At... 1). The main content area shows a search bar with 'GTFS Milan' and a magnifying glass icon. Below the search bar, it says '2 datasets found for "GTFS Milan"' and 'Order by: Relevance'. Two dataset cards are displayed: 'GTFS' and 'Trenord'. The 'GTFS' card describes it as generated by Amati from scheduled data for Milan's public transport, with a 'GTFS' tag. The 'Trenord' card states information comes from the MobilityDataBase, with a 'GTFS' tag. At the bottom, a link says 'You can also access this registry using the API (see API Docs)'.

Home / Datasets

Organizations

- Comune di Milano 1
- Trenord 1

Groups

- Mobility & Transport 1
- Transport networks 1

Tags

- GTFS 1
- timetables 1
- TPL 1

Formats

- GTFS 2

Licenses

- Creative Commons At... 1

GTFS Milan

2 datasets found for "GTFS Milan" Order by: Relevance

GTFS

Gtfs is generated by Amati starting from the scheduled data of the local public transport service transmitted by ATM to the Municipality of Milan, for the monitoring of the...

GTFS

Trenord


The information about these sources comes from the MobilityDataBase for more information please refer directly to their website.

GTFS

You can also access this registry using the API (see API Docs).

Explore and display dataset's metadata

[Home](#) / [Organizations](#) / [Comune di Milano](#)



Comune di Milano

The Municipality of Milan identifies the Open Government paradigm as a way to create an open Public Administration that empowers citizens and businesses to innovate: open data...

[read more](#)

Followers

0

Datasets

2

Organizations

Comune di Milano 2

Groups

Journey planning 1

Datasets

Activity Stream

About

2 datasets found

Order by: Relevance

Formats: GTFS

GTFS

GTFS is generated by Amat starting from the scheduled data of the local public transport service transmitted by ATM to the Municipality of Milan, for the monitoring of the...

GTFS

The Gtfs is generated by Amat starting from the scheduled data of the local public transport service transmitted by ATM to the Municipality of Milan, for monitoring the Service...

Dataset

Groups

Activity Stream

GTFS

Gtfs is generated by Amat starting from the scheduled data of the local public transport service transmitted by ATM to the Municipality of Milan, for the monitoring of the Service Contract which regulates the management of the service within the territory of Milan and the urban area. The dataset is updated periodically, varies from 2 weeks to a month and is you can download it at the following link: <https://dati.comune.milano.it/gtfs.zip> For information on the structure and format of gtfs refer to the following link: <https://developers.google.com/transit/gtfs/reference?hl=it> ## Additional information No information provided This dataset was released by the municipality of Milan.

Data and Resources

ds929_gtfs.zip

preview NOT available

Explore

GTFS

TPL

timetables

Additional Info

Field	Value
Maintainer	AMAT - Territory Environment Mobility Agency
Version	332
Last Updated	September 12, 2023, 7:42 PM (UTC+02:00)
Created	July 3, 2023, 11:52 AM (UTC+02:00)
Alternate identifier	
Conforms to	
Frequency	BIWEEKLY
Identifier	DS929
Issued	
Language	ITA
Modified	2023-09-12
Publisher name	Direzione Mobilità, Ambiente e Energia
Theme	[{"subthemes": [{"http://eurovoc.europa.eu/100238"}, {"theme": "TRAN"}]}
creator	[{"creator_identifier": "01199250158", "creator_name": {"en": "Unitiu00e0 Open Data", "it": "Unitiu00e0 Open Data"}]}
geographical_geonames_url	https://www.geonames.org/3173435
geographical_name	ITA_MIL
harvest_object_id	ffcfb2ac-a1a0-4acb-b8ee-6733edd4bf70
harvest_source_id	46dc7392-5467-4638-8206-63cda4ff88c8
harvest_source_title	Municipality of Milan - Open Data Portal
holder_identifier	c_f205
holder_name	Comune di Milano
is_version_of	https://www.amat-mi.it/it/servizi/pubblicazione-orari-trasporto-


Exploring and displaying resources

Home / Organizations / Hove / Daily Trajectory Logs datasets

Daily Trajectory Logs datasets

Followers
1

Organization



Hove

Hove puts data into action, working towards inclusive, shared, and sustainable mobility as part of the green transition.

read more

Social

Twitter

Facebook

License

Datasets

Organizations

Hove 2

mCLOUD 1

Groups

There are no Groups that match this search

Tags

mcloud_category_avi... 1

Dataset

Groups

Activity Stream

Daily Trajectory Logs datasets


This dataset collects the daily requests sent to the Navitia route planner during 2022 and the first quarter of 2023 starting or ending in the Île-de-France region. The responses included in the dataset are not real journeys but routes planned by the route planner. Every route includes the start and end points as well as some transit points.


Since this information could jeopardize the privacy of the users, an anonymization procedure was applied. The dataset was anonymized using the "Time partition Microaggregation" method, a version of the well-known microaggregation method for very large mobility datasets where the application of microaggregation would require a detailed description of the method can be found here.

Specific parameters: • K: 10 • Interval: 3600 seconds • Clustering_method: MDAV • Aggregation method: Mean

Along with the daily datasets, origin-destination matrixes are also included. These were computed using a tool corresponding to the French zip codes. A geojson file with the layout of the postal codes of the region is also available.

Data and Resources

 **Description Navitia route planner data anonymization**
Anonymization description

 **fr-idf_202201.zip**

trajectory logs

Q


3 datasets found for "trajectory logs"

Order by: Relevance


Daily Trajectory Logs datasets

This dataset collects the daily requests sent to the Navitia route planner during 2022 and the first quarter of 2023 starting or ending in the Île-de-France region. The...

DOCX ZIP XLSX

 **fr-idf_202303.zip**

Explore

 **Navitia measures**

Explore

navitia

route planner

trajectory

Additional Info

Field	Value
Source	https://storageaccount11111111.blob.core.windows.net/container1/MobiDataLab/anonymization_description_2.docx
Last Updated	October 18, 2023, 10:57 AM (UTC+02:00)
Created	September 13, 2023, 6:03 PM (UTC+02:00)
Anonymizer module (github)	https://github.com/MobiDataLab/mdl-anonymizer
Country code	FR
Municipality	Paris
Region	Ile-de-France
Time Partitioned Microaggregation description (github)	https://github.com/MobiDataLab/mdl-anonymizer/blob/master/docs/anonymization/TimepartMicroaggregation.md

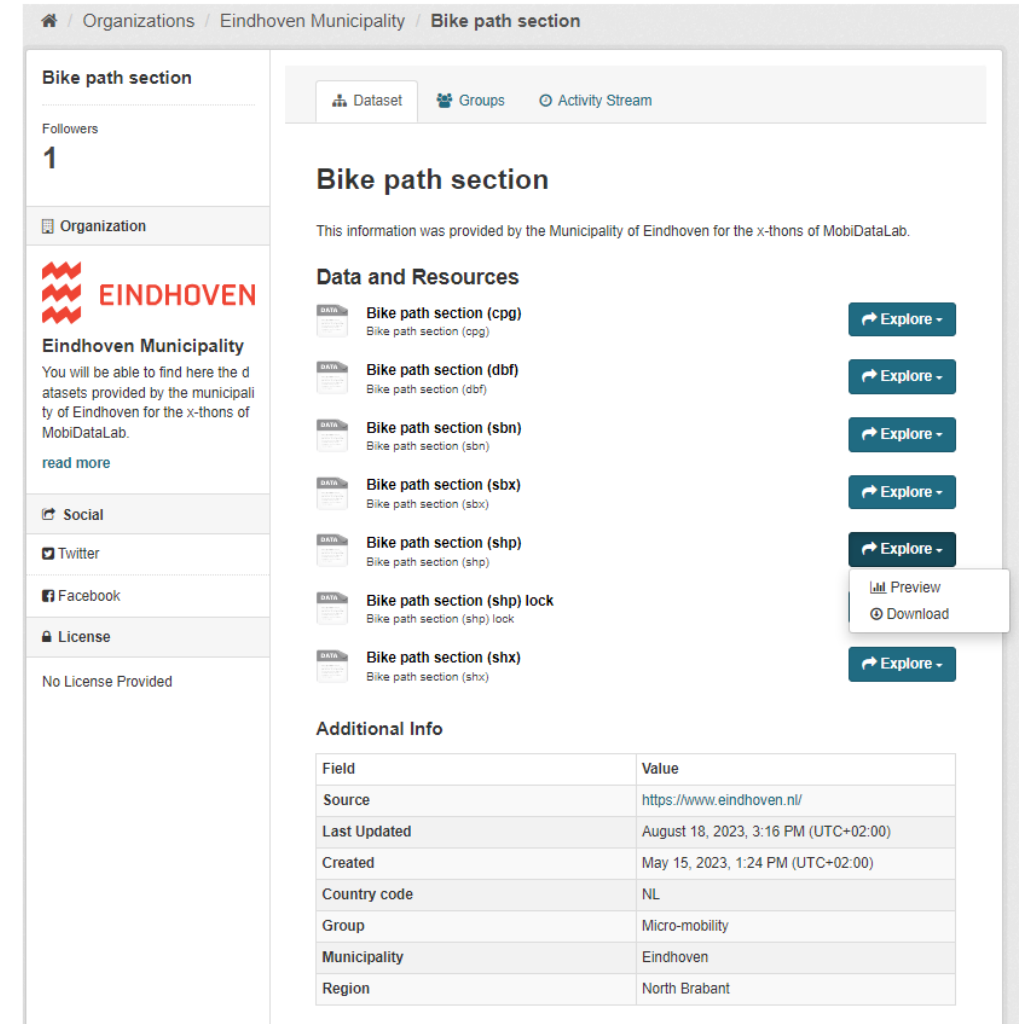
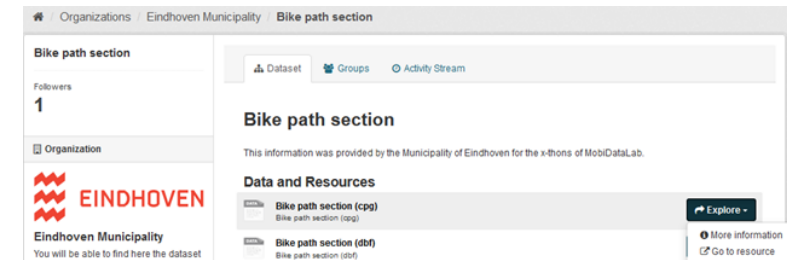
MOBIDATALAB

Labs for prototyping future mobility data sharing solutions in the cloud

Exploring and displaying resources

A dataset can contain more than one resource.

- Some might be the same resource in a different format or a different date.
- Some might contain additional information such as the metadata of each column.
- Some might just be a dataset composed of several files, such as the well-known shape file.



Preview of a resource in CVS

Organizations / Ministry of the... / Unfinished works / Unfinished Works 2016 CSV format

Unfinished Works 2016 CSV format

Manage

Download

URL: <https://dati.mit.gov.it/catalog/dataset/35772d25-5b63-42c0-b7fd-0cf874ba4ac3/resource/7749fd7-1df8-4a4f-a24c-eed5b96726e/download/opereincompiute2...>

Data for 2016 published on June 30, 2017.

Data Explorer

Fullscreen

Embed

Grid

Graph

Map

752 records

1 - 100

Search data ...

Go »

Filters

public...	anno_rif	cup	cup_ma...	ambito...	denomi...	codice...	codice...	natura...	tipolog...	codice...	tipolog...	localizz...	localizz...
Regione...	2016	B59D10...			COMU...	489010...	3	REALIZ...	b) i lavo...	1	NUOVA...	3018060	
Regione...	2016	G44E89...			COMU...	820005...	3	REALIZ...	a) i lavo...	1	NUOVA...	17078019	ITF5
Regione...	2016	D31B91...			Comun...	800019...	3	REALIZ...	b) i lavo...	1	NUOVA...	17078036	ITF5
Regione...	2016	D69D11...			comune...	8000411...	3	REALIZ...	a) i lavo...	1	NUOVA...	17078084	ITF5
Regione...	2016	I15E080...			Comun...	850004...	3	REALIZ...	b) i lavo...	51	COMPL...	17078043	ITF5
Regione...	2016	E83B06...			COMU...	800044...	3	REALIZ...	b) i lavo...	4	RISTR...	17078014	ITF5
Regione...	2016	D38D14...			COMU...	800163...	3	REALIZ...	b) i lavo...	51	COMPL...	3018089	
Regione...	2016	E45I870...			AMM.N...	800005...	3	REALIZ...	a) i lavo...	3	RECUP...	17077008	
Regione...	2016	H28B80...			COMU...	800049...	3	REALIZ...	a) i lavo...	51	COMPL...	17078017	ITF5
Regione...	2016	B34B82...			comune...	8100117...	3	REALIZ...	a) i lavo...	1	NUOVA...	17077007	ITF5
Regione...	2016	I73G01...			Comun...	820009...	3	REALIZ...	a) i lavo...	1	NUOVA...	17077029	ITF5
Regione...	2016	H11B03...			PROVI...	800027...	3	REALIZ...	b) i lavo...	51	COMPL...	17078007	ITF5
Regione...	2016	C76J15...			COMU...	800025...	3	REALIZ...	a) i lavo...	54	COMPL...	17077011	ITF5
Regione...	2016	E53D06...			COMU...	800042...	3	REALIZ...	a) i lavo...	51	COMPL...	17078002	ITF5
Regione...	2016	E33F09...			Consorz...	91600767	3	REALIZ...	b) i lavo...	51	COMPL...	17078063	
Regione...	2016	J72J10...		Regionale	COMU...	238090...	3	REALIZ...	c) i lavo...	7	MANUT...	18079033	ITF5
REGIO...	2016			Regionale	Fondazi...				b) i lavo...			10054039	
Regione...	2016	F71B07...		Regionale	Provinci...	800037...	3	REALIZ...	a) i lavo...	1	NUOVA...	18078003	ITF5
REGIO...	2016	I33F050...		Regionale	Regione...				b) i lavo...			10054024	
Regione...	2016	I69B110...		Regionale	Comun...	800202...	3	REALIZ...	a) i lavo...	1	NUOVA...	15085135	ITF3
Regione...	2016	G47B90...		Regionale	Comun...	581980...	3	REALIZ...	b) i lavo...	1	NUOVA...	15083083	ITF3
Regione...	2016	H49G09...		Regionale	COMU...	800215...	3	REALIZ...	b) i lavo...	55	COMPL...	15085014	ITF3

Resources

Unfinished Works 2016...

metadata

metadata (PDF)

Unfinished Works 2017...

Unfinished works 2016...

Unfinished works 2017...

Social

Additional Information

Field	Value
Data last updated	March 4, 2022
Metadata last updated	August 22, 2023
Created	unknown
Format	CSV
License	CC-BY-4.0

Show more

MOBIDATALAB
Labs for prototyping future mobility data sharing solutions in the cloud

| CKAN

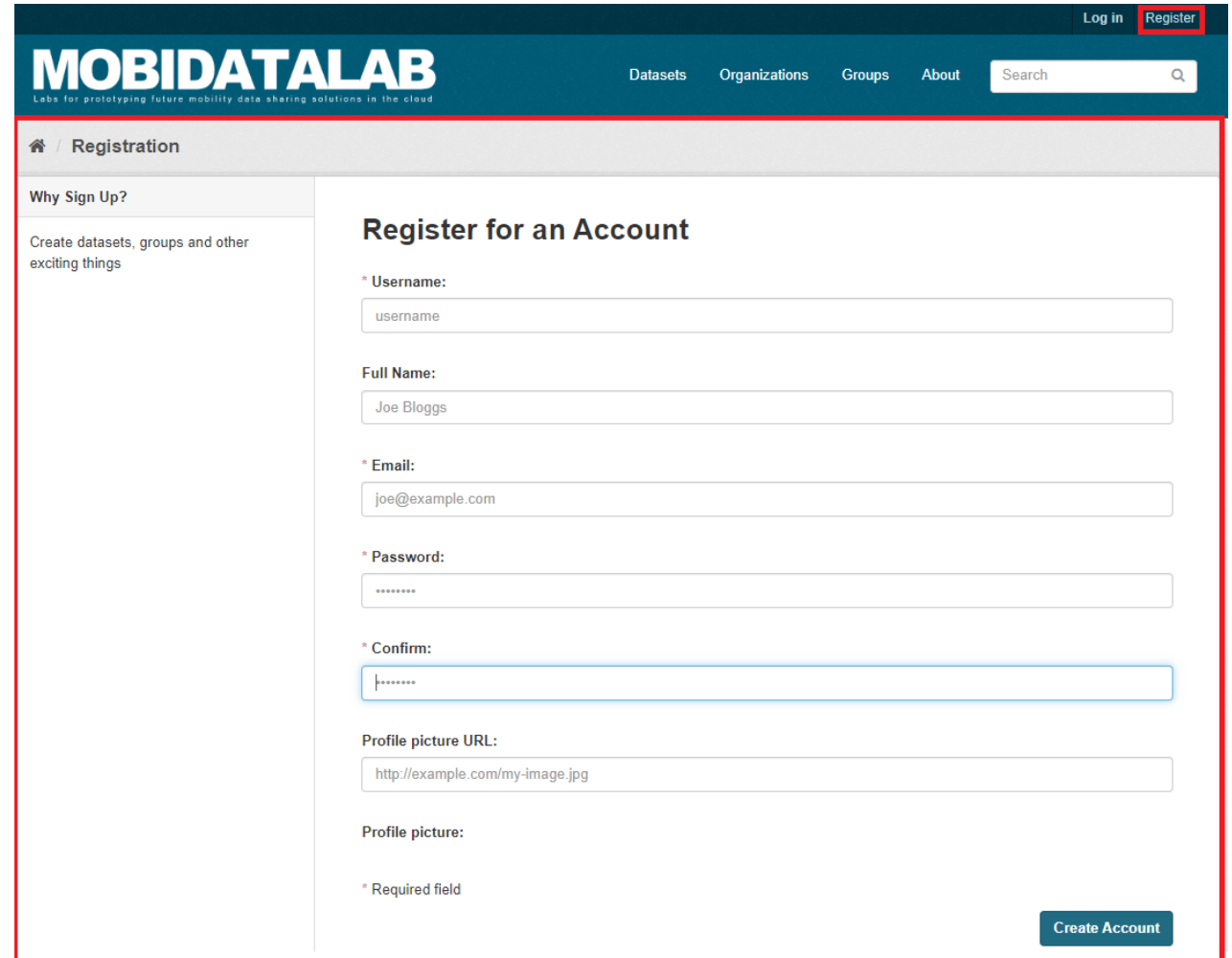
Demo management



| Register and log in

Registration is useful for “following” datasets, but it is mostly for administrators and organizations publishing and personalizing features.

- To create a user ID, click on “Register” at the top of any page.



The screenshot shows the MOBIDATALAB website's registration page. The header is dark blue with the MOBIDATALAB logo and tagline 'Labs for prototyping future mobility data sharing solutions in the cloud'. Navigation links include 'Datasets', 'Organizations', 'Groups', 'About', and a search bar. In the top right corner, there are 'Log in' and 'Register' buttons, with 'Register' highlighted in red. The main content area is titled 'Registration' and is divided into two columns. The left column, titled 'Why Sign Up?', contains the text 'Create datasets, groups and other exciting things'. The right column, titled 'Register for an Account', contains a registration form with the following fields: 'Username' (placeholder: username), 'Full Name' (placeholder: Joe Bloggs), 'Email' (placeholder: joe@example.com), 'Password' (placeholder: *****), 'Confirm' (placeholder: *****), 'Profile picture URL' (placeholder: http://example.com/my-image.jpg), and 'Profile picture'. A legend at the bottom left of the form indicates that an asterisk (*) denotes a required field. A 'Create Account' button is located at the bottom right of the form.

MOBIDATALAB
Labs for prototyping future mobility data sharing solutions in the cloud

Log in **Register**

Datasets Organizations Groups About Search

Registration

Why Sign Up?

Create datasets, groups and other exciting things

Register for an Account

* Username:

Full Name:

* Email:

* Password:

* Confirm:

Profile picture URL:

Profile picture:

* Required field

Create Account

Manage Organizations

- Datasets can be also discovered on the organizations and groups pages with the help of filters.
- The organizations page allows the discovery of data providers covering one or more municipalities. Their description is often complemented with links to the website and API's endpoints of the organizations.

The screenshot displays the CKAN Organizations interface. On the left, a sidebar contains a search bar and a list of organizations. The main content area shows a grid of organization cards. The top card is for 'Open Data Vlaanderen', which is highlighted. Below the grid, there is a detailed view for 'Open Data Vlaanderen'.

Open Data Vlaanderen
Open data portal of the Flemish region. It gives you access to an extensive range of information, so that it can be used in a reliable and secure way. All the information from Flemish authorities on one central platform. <https://metadata.vlaanderen.be/>

Additional Info

Field	Value
Country code	BE
Municipality	Leuven
Region	Flemish region
URL API	https://metadata.vlaanderen.be/srv/dut/catalog.search#/home
URL website	https://www.vlaanderen.be/datavindplaats/catalogus

| Add Organizations

- Log into CKAN
- Go to the **Organizations** page and select **Add organization**.
- Enter the publisher's name (e.g., myCity) and description
- Add the logo of the organization

[Home](#) / [Organizations](#) / **Create an Organization**

What are Organizations?

CKAN Organizations are used to create, manage and publish collections of datasets. Users can have different roles within an Organization, depending on their level of authorisation to create, edit and publish.

Create an Organization

Name:

* URL: processor1.francecentral.cloudapp.azure.com/organization/<organization> [Edit](#)

Description:

A little information about my organization...

You can use [Markdown formatting](#) here

Image:

[Upload](#) [Link](#)

Parent:

None - top level

* Required field

[Create Organization](#)

Manage datasets

What's a resource?

A resource can be any file or link to a file containing useful data.

1 Create dataset

2 Add data

Data:

Upload

Link

Name:

eg. January 2011 Gold Prices

Description:

Some useful notes about the data

You can use Markdown formatting here

Format:

eg. CSV, XML or JSON

This will be guessed automatically. Leave blank if you wish

Previous

Save & add another

Finish

Datasets / Create Dataset

What are datasets?

A CKAN Dataset is a collection of data resources (such as files), together with a description and other information, at a fixed URL. Datasets are what users see when searching for data.

1 Create dataset

2 Add data

Title:

demo

URL:

/dataset/ demo

Description:

demo description

You can use Markdown formatting here

Tags:

transport

License:

Creative Commons Non-Commercial (Any)

License definitions and additional information can be found at creativecommons.org/licenses/by-nc/4.0/

Organization:

ADME

Visibility:

Private

Source:

<https://mobidatalab.eu/>

Version:

1.0

Author:

MobiDataLab

Author Email:

joe@example.com

Maintainer:

Joe Bloggs

Maintainer Email:

joe@example.com

Custom Field:

Key: Country code

Value: BE, FR, ES, AU, JP, LT, LU, NO, SE, SI

Custom Field:

Key: Municipality

Value: Leuven

The data license you select above only applies to the contents of any resource files that you add to this dataset. By submitting this form, you agree to release the metadata values that you enter into the form under the Open Database License.

* Required field


Next: Add Data


Manage groups


- Groups can help to find more easily datasets related to a certain theme or category.
- In the MobiDataLab reference data catalogue, groups were defined in relation to the use case challenges of the living and virtual lab challenges.


25 groups found


Order by: Name Ascending ▾


**Accessibility**
Accessibility to facilities and public spaces for all kind of users....
59 Datasets


**Accidents**
Historical data on fatalities or accidents (covering different modes).
22 Datasets


**Addresses**
Location of properties based on address identifiers, usually by road name,...
11 Datasets


**Administrative units**
Units of administration, dividing areas where Member States have and/or...
23 Datasets


**Amenities**
Features that provide comfort, convenience or pleasure.
19 Datasets


**Analytics & Learning**
Generic datasets for the "analytics and learning" use case in which tools are...
13 Datasets


**Atmospheric Conditions and meteorological geographical features**
Physical conditions in the atmosphere. Includes spatial data based on...
36 Datasets


**Culture**
Datasets related to cultural activities or sites.
17 Datasets


**Data services**
Heterogeneous-stored data web services, APIs and tools to find and extract data.
24 Datasets


**Freight**
Goods of cargo, cargo control data, cargo network, cargo places,...
6 Datasets


**Geography & Environment**
Various datasets in relation to geographic features or environmental...
40 Datasets


**Journey planning**
Datasets in relation to travel planning. Including also information about...
53 Datasets


**Environmental monitoring facilities**
Location and operation of environmental monitoring facilities include...
34 Datasets


**Mobility & Transport**
Information about public transport, transport networks, roads, transport...
346 Datasets


**Parking**
General parking information. Information about parking fees for a right or...
65 Datasets


**Points of Interest**
Point of interest (POI) datasets. POI is a term used in cartography and...
35 Datasets

**Land use**
Land Use theme is defined as the use and functions of a territory and it is...
45 Datasets

**Shared mobility**
Datasets of shared transport including areas, costs and offers. Share...
36 Datasets

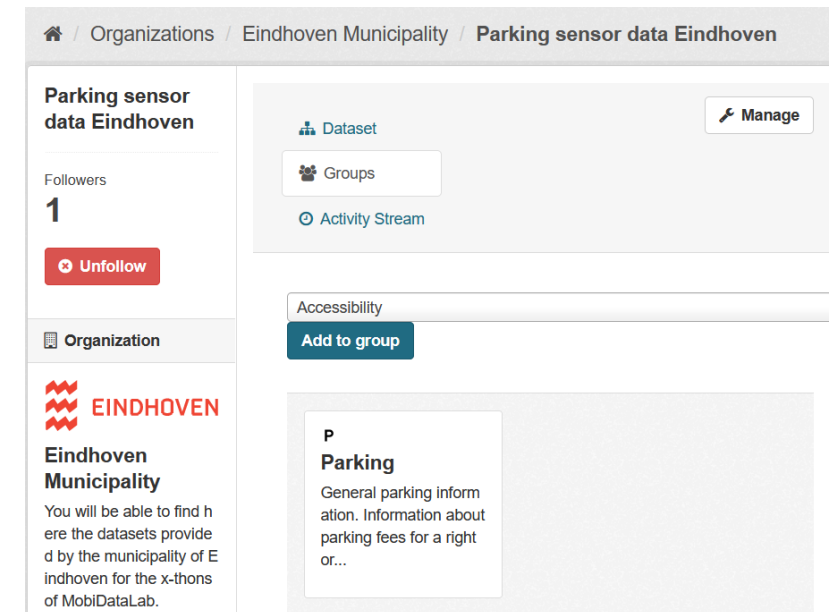
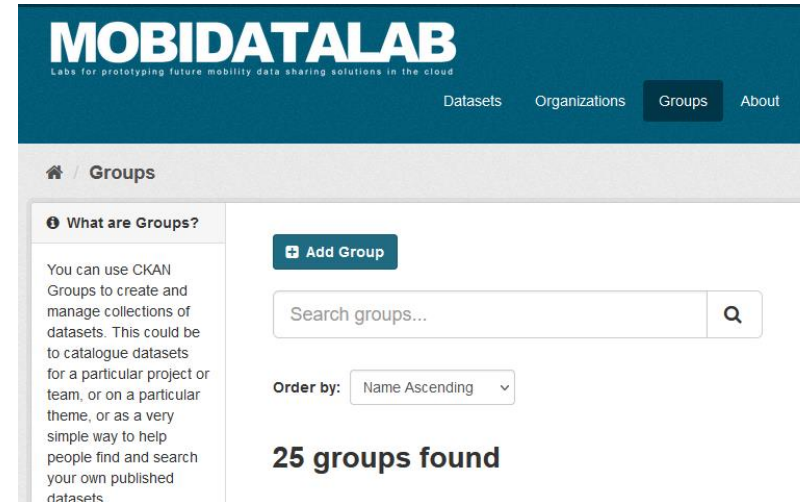
**Transport networks**
Road, rail, air and water transport networks and...
36 Datasets

**Micro-mobility**
Datasets related to micro-mobility services or features.
36 Datasets

**Topography**

Add groups

- To add a group is simple, the data publisher needs to log into the CKAN portal and go to **Groups**. Then click on **Add Group**, add the name and the description of the group and save it.
- To add a dataset manually into a group, select a dataset, click on the **Groups** tab, select a group from the menu and click on **Add to Group**.



| Data and metadata storage

FileStore API

- Files can be uploaded to the FileStore by using the `resource_create()` and `resource_update()` action API functions.

DataStore

- The DataStore extension provides a database for the storage of structured data from CKAN resources.
- Data can be extracted from resource files and stored in the DataStore. This one then enables data previews and a data API for the resources. The DataStore is distinct but complementary to the FileStore.

Metadata file to add datasets into CKAN

Legend									
M - Mandatory R - Recommended O - Optional MDL - To be added by the team of MobiDataLab									
1 = required, but not repeatable 1..N = required and repeatable 0..1 = optional, but not repeatable 0..N = optional and repeatable									

```

CKAN_CTW = {
    "token": "a2",
    "api_key": "a2"
}

CKAN_P1 = {
    "url": "http://",
    "token": "a2",
    "api_key": "a2"
}

CKAN_CTW = {
    "url": "http://",
    "token": "a2",
    "api_key": "a2"
}

def create_multiple_new_packages(
    filepath: str,
    package_data: pd.DataFrame,
    str_to_dict: Callable[[str], Dict[str, str]]):
    for item in package_data:
        if _DEBUG:
            print('Line handling : ' + str(item))

        result = create_package(
            item['name_package'], item['title_package'], item['license_package_id'],
            item['url_package'], item['tags_package'],
            item['country'], item['region'],
            item['municipality'], item['metadata_modified'], item['group_id'], item['group_name'],
            item['owner_org_name'], item['notes_package'])

        if result.json()['success'] == True:
            PACKAGE_ID = result.json()['result']['id']
            print('Package ' + item['name_package'] + ' created with the ID: ' + PACKAGE_ID)
        else:
            print(result.json())
            print('Error, package ' + item['name_package'] + ' not created ')
            print('Reason : ' + result.json()['error']['name'][0])

def add_resource(
    id_package: str,
    name_package: str,
    name: str,
    description: str,
    url: str,
    owner_org_name: str,
    format: str,
    groups: str):
    result = requests.post(
        CKAN_ADDRESS[ 'url' ] + '/api/action/resource_create',
        json={
            "package_id": id_package,
            "package_name": name_package,
            "name": name,
            "description": description,

```

Example of use of the metadata file with a CKAN API toolbox

Metadata file elements explanation

Metadata file elements example

Visualising data in CKAN

Home / Organizations / Paris open Data / Accessibility diagnoses of... / Accessibility diagnoses Paris (shp) / **Edit**

Accessibility diagnoses Paris (shp)


Format
SHP


[Edit resource](#)[All resources](#)[View resource](#)

[Views](#)

[New view](#)

[Reorder resource view](#)

 Image

 Shapefile

Home / Organizations / Paris open Data / Pedestrian frequentation... / **Pedestrian frequentation...**

Pedestrian frequentation of Paris streets (shp)


[Download](#)

URL: <https://storageaccount11111111.blob.core.windows.net/container1/Paris-Mobility/Pedestrian-frequentation-Paris.zip>

Estimated hourly pedestrian frequentation and densities of Paris streets, in June 2022, at peak hours, on weekdays. These estimations were produced by Hove for the City of Paris. They are based on GNSS data collected by smartphone applications, redressed with socio-demographic data and processed to quantify foot trips along Paris streets (using OpenStreetMap routes). Pedestrian densities are calculated for each street segment, by taking into account the median width of sidewalks. The data also contains indexes, from 1 to 10, meant to classify streets in terms of frequentation and density.

[Shapefile](#)

[Fullscreen](#) [Embed](#)



Leaflet | Map tiles by Stamen Design (CC BY 3.0), Data by OpenStreetMap (CC BY SA)

[Resources](#)

[Pedestrian...](#)

[Social](#)

[Twitter](#)

[Facebook](#)

Additional Information

Field	Value
Data last updated	unknown
Metadata last updated	September 14, 2023
Created	unknown
Format	SHP
License	Open Data Commons Open Database License (ODbL)

Multi-lingual management: translation

- It is possible to change metadata fields of language by selecting on the menu or on the search bar your preferred language.

ckan.mobidatalab.eu/es/dataset/shapefile-of-the...

Identificador	https://ressources.data.sncf.com/explore/dataset/formes-des-lignes-du-rtn/
Idioma	["http://id.loc.gov/vocabulary/iso639-1/fr"]
Modificado	2022-11-16T17:23:51+00:00
Nombre del publicador	SNCF Réseau
Publicado	2022-11-16T17:23:51+00:00
Tema	["https://ressources.data.sncf.com/explore/?refine.theme=Description+du+r%C3%A9seau"]
harvest_source_title	SNCF

Acerca de CKAN
API CKAN
CKAN Association
[OPEN DATA](#)
Gestionado con ckan
Idioma:

ckan.mobidatalab.eu/en/dataset/shapefile-of-the-lines...

Identifier	https://ressources.data.sncf.com/explore/dataset/formes-des-lignes-du-rtn/
Issued	2022-11-16T17:23:51+00:00
Language	["http://id.loc.gov/vocabulary/iso639-1/fr"]
Modified	2022-11-16T17:23:51+00:00
Publisher URI	https://ressources.data.sncf.com/explore/?refine.publisher=SNCF+R%C3%A9seau
Publisher name	SNCF Réseau
Theme	["https://ressources.data.sncf.com/explore/?refine.theme=Description+du+r%C3%A9seau"]
URI	https://ressources.data.sncf.com/explore/dataset/formes-des-lignes-du-rtn/
harvest_source_title	SNCF

About CKAN
CKAN API
CKAN Association
[OPEN DATA](#)
Powered by ckan
Language:

Harvesting

ckan.mobidatalab.eu/harvest

MOBIDATALAB
Labs for prototyping future mobility data sharing solutions in the cloud

Datasets Organizations Groups About

Harvest Sources

Frequency

- MANUAL 26
- WEEKLY 11
- BIWEEKLY 10
- MONTHLY 6

Type

- dcat_rdf 23
- ckan 9
- csv 9
- dcat_json 9
- inspire Harvester 3

Add Harvest Source

Search harvest sources...

Order by: Name Ascending

53 harvests found

Update frequency: Biweekly

Configuration:

```
{
  "default_tags": [{"name": "transport"}],
  "default_groups": ["Mobility & Transport"],
  "default_extras": {"encoding": "utf8"},
  "remote_groups": "only_local",
  "remote_orgs": "create"
}
```

Organization: infrabel

Delete Save

- This extension allows to harvest metadata from other catalogues and integrate them in the CKAN portal. Most of the data in the reference data catalogue is harvested from different data sources, but mainly from governmental organizations catalogues.

| Harvested metadata

115k+ datasets harvested on CKAN

- From 34 data portals were harvested for the living and virtual labs of MobiDataLab
 - Including the following countries:
 - Greece
 - Germany
 - Italy
 - France
 - Netherlands
 - Belgium

A screenshot of the CKAN Datasets page. The left sidebar shows a list of organizations and groups with their respective dataset counts. The main content area displays a search bar, the total number of datasets found (115,102), and a list of dataset descriptions with download links for various formats like DOCX, ZIP, and XLSX.

Datasets

Search datasets...

115,102 datasets found Order by: Relevance

Daily Trajectory Logs datasets
This dataset collects the daily requests sent to the Navitia route planner during 2022 and the first quarter of 2023 starting or ending in the Île-de-France region. The...
DOCX ZIP XLSX

Municipality of Rozzano - Population Statistics - Natural Population Balance
Natural Balance of the Population (last 15 years)

Population forecast 2016 (BP2016): population by age
Raw data population forecast 2016: population by age. The South Holland Population Forecast 2016 (BP2016) is intended as a quantitative basis for the provincial housing policy....
PDF WFS

Population Development Indicators (Population)
Description Utrecht in Figures is the online database of the municipality of Utrecht. Here the Research & Advice department publishes figures about the city of Utrecht and...
<http://publications.europa.eu/resource/authority/file-type/XLSX> <https://data.overheid.nl/format/unknown>

Indicators Population Forecast (Population)
Description Utrecht in Figures is the online database of the municipality of Utrecht. Here the Research & Advice department publishes figures about the city of Utrecht and...
<http://publications.europa.eu/resource/authority/file-type/XLSX> <https://data.overheid.nl/format/unknown>

Organizations

- GeoDatiGovIt RNDT 23660
- OverheidNI 16295
- Land Brandenburg 10527
- Open Data Vlaanderen 9427
- NationaalGeoregisterNL 8343
- Transparenzportal H... 8295
- The Federal Ministr... 8015
- GDI-DE 4002
- Open-Government-Dat... 3549
- Open Data of Lombardia 3361

Show More Organizations

Groups

- Mobility & Transport 346
- Traffic 83
- Micro-mobility 75
- Parking 65
- Accessibility 59
- Transport networks 57
- Journey planning 53
- Population distribu... 49

Customized harvesting for the Datathon & Hackathon use-cases challenges

country	municipal	title	name	source_ty	frequency	url	harvested	datasets	errors 19/	/2023
France	National	France - Open Data	datagouvfr	dcat_rdf	BIWEEKLY	https://www.data.gouv.fr/catalog.xml	done	78	25	
Netherlands	Eindhoven (N)	Eindhoven - Open Data	eindhoven-op	dcat_rdf	BIWEEKLY	https://data.eindhoven.nl/api/v2/catalog/exports/dcat	done	53	64	
Netherlands	Eindhoven (N)	Georegister provincie	georegister-p	dcat_rdf	BIWEEKLY	https://georegister.brabant.nl/geonetwork/srv/eng/r	done	64	36	
France	National	IGN FR	ign-fr	csw	BIWEEKLY	https://wxs.ign.fr/inspire/inspire/r/wms?SERVICE=W	done	52	0	
Belgium	National	Infrabel Open Data	infrabel	dcat_rdf	BIWEEKLY	https://infrabel.opendatasoft.com/api/v2/catalog/ex	done	89	2	
Italy	National	Ministry of the Interior	dati-mit-gov	ckan	BIWEEKLY	http://dati.mit.gov.it/catalog/	done	68	1	
Belgium	National	NAP ITS Belgium	transportdat	ckan	BIWEEKLY	https://www.transportdata.be/	done	154	0	
Netherlands	National	Nationaal Georegister	ngr	dcat_rdf	BIWEEKLY	https://www.nationaalgeoregister.nl/geonetwork/srv	done	130	3	
Belgium	National	Royal Meteorological Institute	royal-meteor	dcat_rdf	BIWEEKLY	https://opendata.meteo.be/geonetwork/srv/eng/rdf	done	27	1	
France	Versailles	Versailles Grand	vgp	dcat_rdf	BIWEEKLY	https://vgp-esrifrance.opendata.arcgis.com/api/feed/	done	25	26	
Germany	Berlin	Geodata Infrastruktur	gdi-be	csw	MANUAL	https://gdi.berlin.de/geonetwork/srv/eng/csw?REQU	done	3010	0	
Italy	Milan (Lomb)	Geographic Portal	milano-geop	dcat_json	MANUAL	https://metadata.comune.milano.it/geoportale/rest/d	not successful	0	1	
International	National	Navitia	navitia	dcat_rdf	MANUAL	https://navitia.opendatasoft.com/api/v2/catalog/exp	done	19	0	
Germany	Berlin	OpenDataBerlin	opendataber	dcat_rdf	MANUAL	https://datenregister.berlin.de/catalog_no_fb.rdf	is not the pri	615	1	
Greece	Trikala (Thess)	Open data portal	open-data-tr	dcat_json	MANUAL	https://data.trikalacity.gr/data.json	is not the pri	26	0	
Italy	National	GeoDatiGovIt RN	geodatigovit	csw	MONTHLY	https://geodati.gov.it/RNDT/csw	done	21732	24209	
Netherlands	National	Nationaal Georegister	ngr	csw_inspire	MONTHLY	http://nationaalgeoregister.nl/geonetwork/srv/eng/c	done	8054	252	
Italy	Milan (Lomb)	Open Data of Lombardia	dati-lombard	dcat_json	MONTHLY	https://www.dati.lombardia.it/resource/xmqv-vx2s.js	done	3,315	554	
Belgium	Leuven (Flem)	Flemish region - Open Data	open-data-vl	csw	MONTHLY	https://metadata.vlaanderen.be/srv/eng/csw	done	7061	392	
Netherlands	Eindhoven (N)	Overheid.nl	overheid-nl	ckan	MONTHLY	https://data.overheid.nl/data/	done	16295	6815	
Germany	National	Mcloud - BMDV	mcloud-bmd	dcat_rdf	MONTHLY	https://www.mcloud.de/export/datasets?pageSize=1	done	3231	17041	
France	Paris	AirParif	air-parif	dcat_json	WEEKLY	https://data-airparif-asso.opendata.arcgis.com/api/fe	done	540	0	
France	National	CEREMA	cerema	csw	WEEKLY	https://catalogue.cdata.cerema.fr/geonetwork/srv/eng	done	460	0	
France	Paris	Ile-de-France Mobilités	idf-mobilites	dcat_json	WEEKLY	https://data.iledefrance.fr/api/explore/v2.1/catalog/ent	done	53	0	
France	Paris	Ile-de-France Open Data	data-iledefra	dcat_rdf	WEEKLY	exports/dcat?lang=fr	done	548	0	
Italy	Milan (Lomb)	Municipality of Milan	comune-di-mi	ckan	WEEKLY	https://dati.comune.milano.it	done	2167	0	
France	Paris	Open data platform	paris-region-	arcgis	WEEKLY	https://data-iau-idf.opendata.arcgis.com/	done	97	22	
Belgium	Leuven (Flem)	Flemish region - Open Data	open-data-vl	dcat_rdf	WEEKLY	https://metadata.vlaanderen.be/metadataserver/srv	done	661	0	
France	Paris	Paris open Data	opendata-pa	dcat_rdf	WEEKLY	https://opendata.paris.fr/api/v2/catalog/exports/dca	done	393	0	
France	Paris	RATP	ratp	dcat_rdf	WEEKLY	https://data.ratp.fr/api/explore/v2.1/catalog/exports	done	19	0	

Portals
successfully
harvested : 34 (27
active jobs for the
codagon)

Scraping

CKAN hosts a vast collection of publicly available datasets, but it is not always possible to harvest or retrieve all datasets from CKAN and other sources for several reasons:

1. Permissions and Licensing
2. API and Rate Limiting
3. Data Size and Complexity
4. Selective Data Publishing
5. Technical Constraints

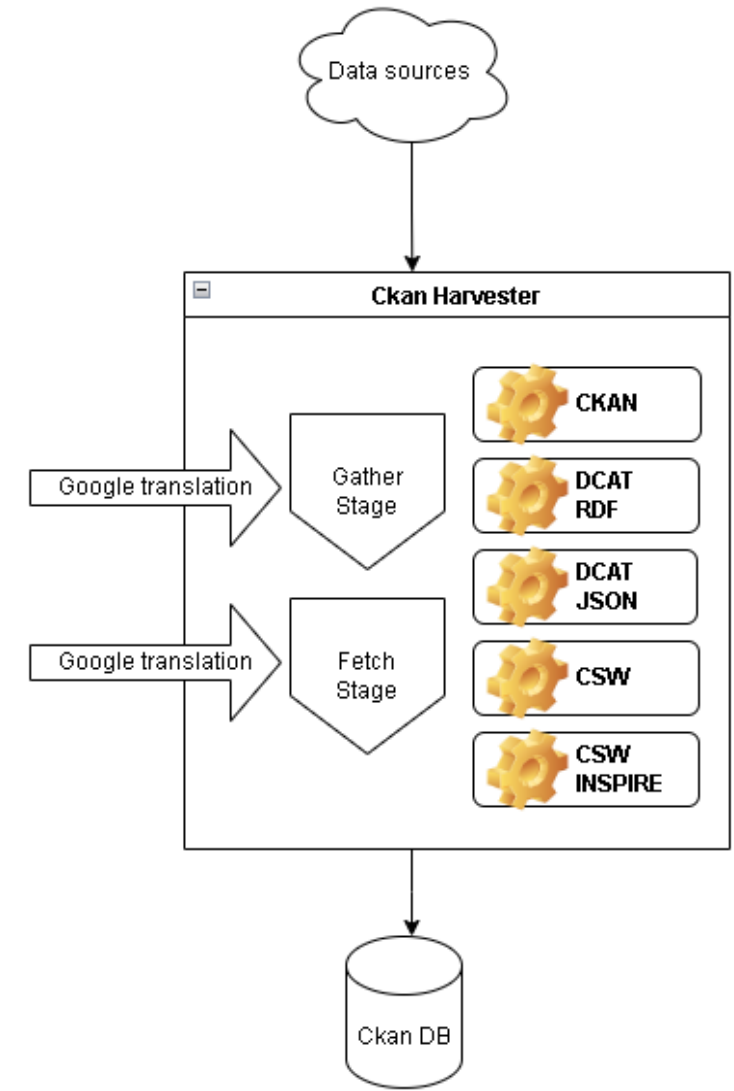
Web scraping is the process of extracting data or information from websites.

```
1 # import libraries
2 from bs4 import BeautifulSoup
3 import urllib.request
4 import csv
5 import pandas as pd
6
7 # specify the url
8 urlpage = 'https://opendata.ndw.nu/'
9
10
11 # query the website and return the html to the variable 'page'
12 page = urllib.request.urlopen(urlpage)
13 # parse the html using beautiful soup and store in variable 'soup'
14 soup = BeautifulSoup(page, 'html.parser')
15
16 # find results within table
17 table = soup.find('table')
18 results = table.find_all('tr')
19 print('Number of results', len(results))
```

Datasets added by other means on CKAN		
Sources	Country / Municipality	Count
Ndov-Locket	Netherlands	55
NDW	Netherlands	26
Milano Geoportale metadati	Milan	90
GEOFABRIK	International	6
Ovapi	Netherlands	4
MobilityDataBase	International	164
Eindhoven municipality	Eindhoven	24
Leuven municipality	Leuven	11
Milan municipality	Milan	8
Navitia/Hove/URV	France	2
Singlespot	France	2
Total		392

Translation

- Google Translation API was used to translate the dataset's metadata while harvesting
 - it translates all the dataset's descriptions into a common language (the translation is done directly in the database).



| CKAN API

- API functions for searching for and getting data from CKAN
- More details about the full usage of CKAN API can be found here:
<https://docs.ckan.org/en/2.9/api/index.html#api-guide>

The CKAN's Action API is a powerful, RPC-style API that exposes all CKAN's core features to API clients. All the CKAN website's core functionalities can be used by external code that calls the CKAN API (meaning that everything that you can do with the web interface and more can be also done with the API). For example, by using the CKAN API your app can:

- Get JSON-formatted lists of a site's datasets, groups or other CKAN objects:
 - https://ckan.mobidatalab.eu/api/3/action/package_list
 - https://ckan.mobidatalab.eu/api/3/action/group_list
 - https://ckan.mobidatalab.eu/api/3/action/tag_list
- Get a full JSON representation of a dataset, resource or other object:
 - https://ckan.mobidatalab.eu/api/3/action/package_show?id=traffic-events
 - https://ckan.mobidatalab.eu/api/3/action/tag_show?id=transport
 - https://ckan.mobidatalab.eu/api/3/action/group_show?id=accessibility
- Search for packages or resources matching a query:
 - https://ckan.mobidatalab.eu/api/3/action/package_search?q=bike
 - https://ckan.mobidatalab.eu/api/3/action/resource_search?query=name:transport
- Create, update and delete datasets, resources and other objects
 - https://ckan.mobidatalab.eu/api/3/action/package_create
- Get an activity stream of recently changed datasets on a site:
 - https://ckan.mobidatalab.eu/api/3/action/recently_changed_packages_activity_list

Thank you for listening

Renee Obregon, AKKODIS,
renee.obregon-gonzalez@akkodis.com



MobiDataLab is funded by the EU under the H2020 Research and Innovation Programme (grant agreement No 101006879).

The content of this document reflects solely the views of its authors. The European Commission is not liable for any use that may be made of the information contained therein. The MobiDataLab consortium members shall have no liability for damages of any kind that may result from the use of these materials.

AKKODIS

AETHON
ENGINEERING



F&S

here



hove

KU LEUVEN

POLIS

UNIVERSITAT ROVIRA I VIRGILI