



D4.4.6 Pan European Open Data Platform – c: Open Data Platform – Tagging Toolset

Project Acronym Open Cities

Grant Agreement number: 270896

Project Title: OPEN INNOVATION Mechanism in Smart Cities

D4.4.6 Pan European Open Data Platform – c: Open Data Platform – Tagging Toolset

Revision: 0.3

Authors:

Jens Klessmann (Fraunhofer FOKUS)
Benjamin Dittwald (Fraunhofer FOKUS)
Evanela Lapi (Fraunhofer FOKUS)

Project co-funded by the European Commission within the ICT Policy Support Programme		
Dissemination Level		
P	Public	X
C	Confidential, only for members of the consortium and the Commission Services	

Revision History

Revision	Date	Author	Organisation	Description
0.1	17.10.2013	Jens Klessmann	Fraunhofer FOKUS	Update of chapters on “Synergies with other Projects” and “Supporting the 2 nd Open Data Challenge”.
0.2	21.10.2013	Benjamin Dittwald	Fraunhofer FOKUS	Update of chapters “Introduction” and “What is in the OD Platform version 2.0”
0.3	22.10.2013	Evanela Lapi	Fraunhofer FOKUS	Update of Chapter “Dataset statistics for Open Cities data catalogue”

Statement of originality:

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

TABLE OF CONTENTS

1	LIST OF TABLES	4
2	LIST OF FIGURES	4
3	INTRODUCTION	5
4	WHAT IS IN THE OD PLATFORM VERSION 2.0.....	5
4.1	OD PLATFORM NEW FEATURES	5
4.1.1	<i>Thesaurus search ability.....</i>	5
4.1.2	<i>Facetted Search.....</i>	6
4.2	OD PLATFORM NEW ENHANCEMENTS.....	8
4.2.1	<i>New layout</i>	8
4.2.2	<i>New library for connecting with CKAN</i>	10
4.2.3	<i>Updating to CKAN 2.1</i>	10
4.2.4	<i>Integration of apps and documents catalogue</i>	11
4.2.5	<i>Improvement of stability and performance</i>	11
4.2.6	<i>Metadata schema enhancements.....</i>	11
4.2.7	<i>Link checker.....</i>	11
4.2.8	<i>Maven and GIT for code management</i>	11
4.2.9	<i>Automatic code testing with respect to coding conventions and mistakes</i>	12
4.2.10	<i>OD Platform Bug Fixes.....</i>	12
5	DATASET STATISTICS FOR OPEN CITIES DATA CATALOUGE	12
5.1	DATASETS STATISTICS UPDATES.....	13
5.1	DATA.OPENCITIES.NET – PORTAL USAGE INDICATORS IN 2013	14
6	SYNERGIES WITH OTHER PROJECTS.....	15
7	SUPPORTING THE 2ND OPEN DATA CHALLENGE.....	17
8	CONCLUSION.....	18

1 LIST OF TABLES

Table 1 Datasets per Category at data.opencities.net	13
Table 2 Overview of synergies with other projects	15

2 LIST OF FIGURES

Figure 1 Search process with thesaurus	5
Figure 2 Thesaurus example Amsterdam	6
Figure 3 Facetted search	7
Figure 4 New facetted search.....	8
Figure 5 Results page of version 1.0 and 2.0	9
Figure 6 German open data portal (www.govdata.de).....	10
Figure 7 data.opencities.net	12
Figure 8 Number of Datasets per City (Feb 2012 / Dec 2012 / Oct 2013)	14
Figure 9 data.opencities.net – Portal Usage Indicators in 2013.....	15
Figure 10: Screenshot from the Open Cities Hack at Home challenge website	17

3 INTRODUCTION

The goal of this deliverable is mainly to describe what we have done to achieve task 4.4.6 of the Open Cities project. Beside that this deliverable gives an overview of the most recent statistics of the usage of the Open Cities Data Catalogue, how the Open Data Platform have synergies with other EU and national projects and the contribution of Fraunhofer FOKUS to the 2nd Open Data Challenge.

4 WHAT IS IN THE OD PLATFORM VERSION 2.0

Within this task we developed new features and a lot of new enhancements to the Open Data Platform. This chapter shows what is new in version 2.0 and shows the differences to version 1.0.

4.1 OD PLATFORM NEW FEATURES

With version 2.0 of the Open Data Platform two new features were integrated:

- Thesaurus search ability
- Facetted search

Each new feature and its benefits are described below.

4.1.1 Thesaurus search ability

To enhance the capabilities of finding metadata we addressed the issue of having different terms for the same meaning. For example housing, home, dwelling and residential block have all a common meaning. But if one is trying to perform a search with only one of the above mentioned terms, he will only get results which are matching with the search term. The results will not include the entries which are matching to the synonyms of the search term. That's where a thesaurus comes in. The thesaurus takes the search term and looks for synonyms of it. The search term gets enriched with the corresponding synonyms by the thesaurus and the search results has all entries with the matching synonyms included. Figure 1 shows this process in a diagram.

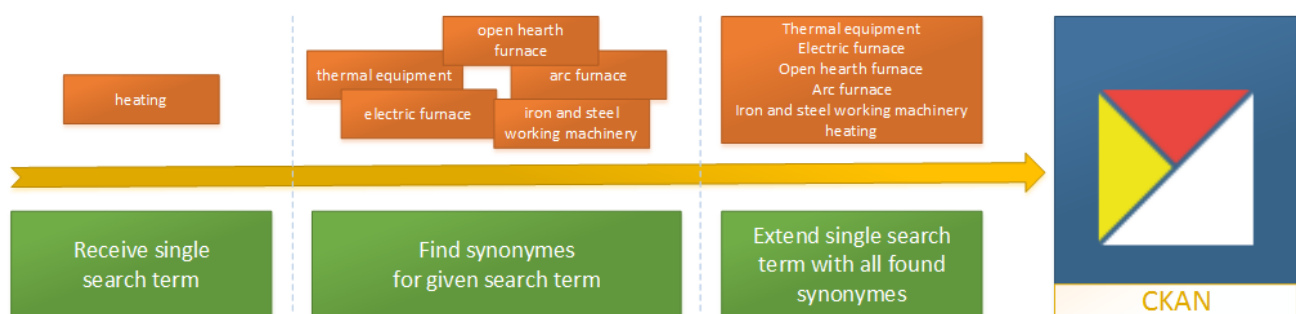


FIGURE 1 SEARCH PROCESS WITH THESAURUS

A thesaurus needs a vocabulary where it can look for synonyms. We use the EuroVoc for the vocabulary for our Open Data Platform. The EuroVoc vocabulary fits very well to the needs of the Open Data Platform. It has many terms in the topic of public sector and is used by many parliaments

in the EU. The vocabulary of the EuroVoc is already available in many languages including all languages of the OpenCities Open Data Platform pilots, meaning Dutch, Spain and German.

As an example for the Amsterdam pilot figure 2 shows the results of a search performed with the terms *woning*, *huisvesting*, *woonblock* and all of them are giving the same 4 results.

4 Results

Sort by Relevance Name Update 1

Tevredenheid over de eigen woning naar buurtcombinaties Rapportcijfer voor de tevredenheid met de eigen woning, 2009-2011 Publishing authority: Onderzoek en Statistiek Categories: Construction, Housing, and Public Works	 Free use Last Update: 11.9.2013
Kerncijfers wonen Amsterdam naar buurten Kerncijfers wonen Amsterdam naar buurten, 1 januari 2011 Publishing authority: Onderzoek en Statistiek Categories: Construction, Housing, and Public Works	 Free use Last Update: 11.9.2013
Vloeroppervlakte x 1.000 m² naar buurtcombinaties en soort gebruik Vloeroppervlakte x 1.000 m2 naar buurtcombinaties en soort gebruik, 1 januari 2010 Publishing authority: Onderzoek en Statistiek Categories: Construction, Housing, and Public Works	 Free use Last Update: 11.9.2013
Woningwaardekaart 2009 Bron via Dienst OenS: alle adrespunten waar in 2009 een woning, appartementen en eengezinswoningen, is verkocht. Zie uitleg.txt Publishing authority: Dienst Ruimtelijke Ordening Categories: Construction, Housing, and Public Works	 Limited use Last Update: 11.9.2013

FIGURE 2 THESAURUS EXAMPLE AMSTERDAM

4.1.2 Facetted Search

The Open Data Platform in version 1.0 has had the ability to search metadata via categories. The start page or even the dataset page offered a set of categories to filter for metadata. In version 1.0 besides the search this was the only option to filter or search for metadata. In version 2.0 this is improved via a facetted search. We have integrated several filters to qualify the results. At the moment there are five filters:

- **Type of License:** The portal differs between metadata with an open source license or a restricted one. With this filter one can see all metadata whatever its license is or only see open source or restricted metadata.
- **Categories:** One can filter for metadata belonging to one or more categories. This of course is new in version 2.0. Version 1.0 only allowed to search metadata in one category.
- **Keywords:** Filter for specific keywords.
- **Formats:** Filter for metadata which has its resources in specific formats.
- **Licenses:** Filter for specific licenses.

If one used the category filter in version 1.0 a new search was performed and all current results were lost. Now the filters are applied to the current search. The benefit of this approach is that the old results are still available, only those entries which are not matching to the new conditions are sort out. They can easily be get back by unchecking the last filter. Figure 3 shows the process of the faceted search.

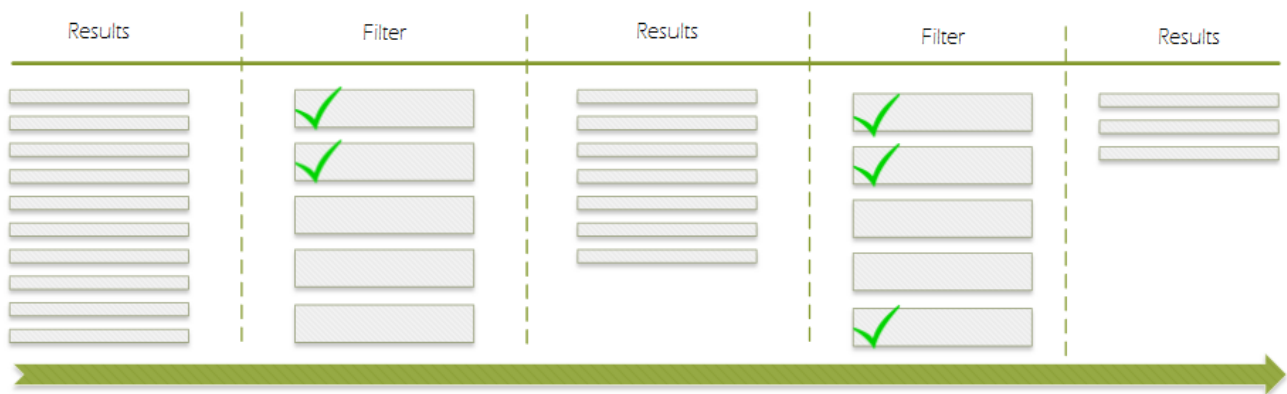


FIGURE 3 FACETTED SEARCH

Each filter show the expected count of results if one decides to use this filter. Figure 4 shows how the new faceted search looks like.

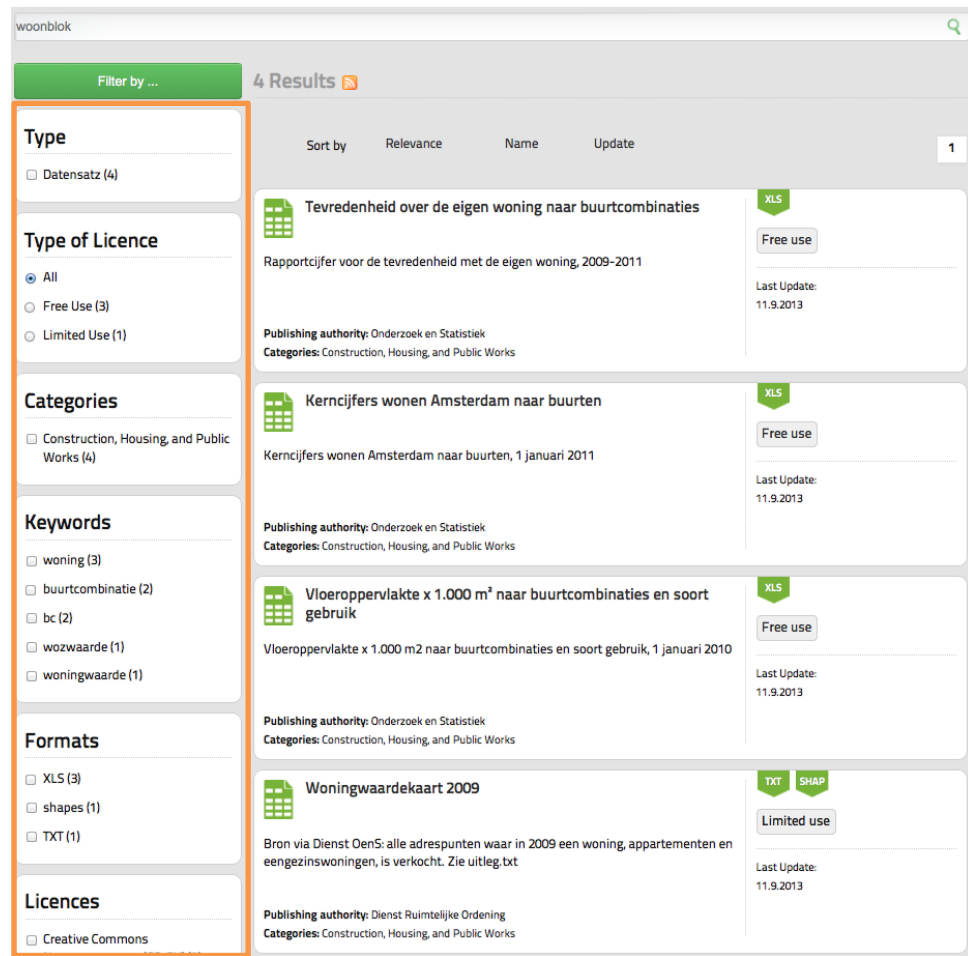


FIGURE 4 NEW FACETTED SEARCH

Due to the system architecture it's easily possible to add more filters.

4.2 OD PLATFORM NEW ENHANCEMENTS

Besides the new features the Open Data Platform version 2.0 has a lot of enhancements. Which are described in this chapter.

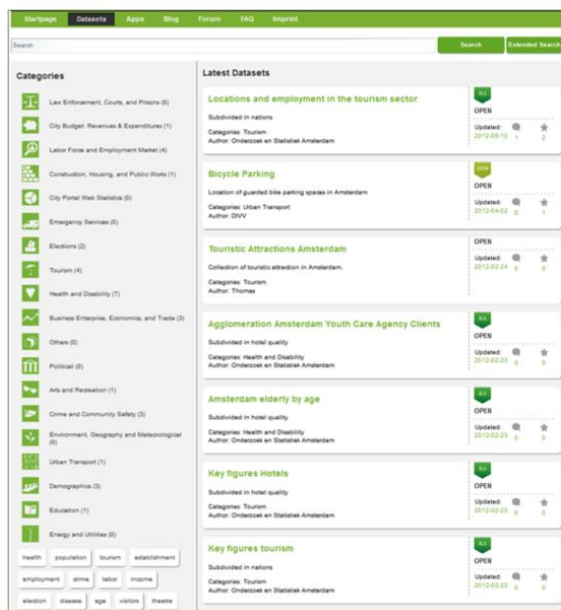
4.2.1 New layout

Every part of the portal has now a new design which is adopted from version 1.0. The design has a more clean appearance and looks much more intuitive. During the implementation phase of version 2.0 we had two community workshops concerning the usability of the portal. The outcome of these workshops was taking into account for the further development.

Most changes can be found on the search results page. In version 1.0 the left side was used for categories and a tag cloud. One result of the community workshops was that a tag cloud isn't useful in that way version 1.0 used it, so we dropped it in version 2.0. The category selection is now replaced by the more useful faceted search. A simple selection for categories can still be found on the start page. The list of results got some enhancements also. A new intuitive symbol indicates now whether this metadata is describing data, a document or an app. The user can now see how many results was found and if it's a huge amount of results the portal now offers pagination for the result

pages now. It's also possible to sort the results by relevance (default), name or update. A very new and wanted feature is the capability of saving a RSS feed for the current search. A user can easily keep track of the search now. The flags on the right side of each result have a slightly new meaning. As we had a flag for each resource in version 1.0, now the flags are indicating that there is at least one resource attached to this metadata with this format.

Version 1.0



Version 2.0

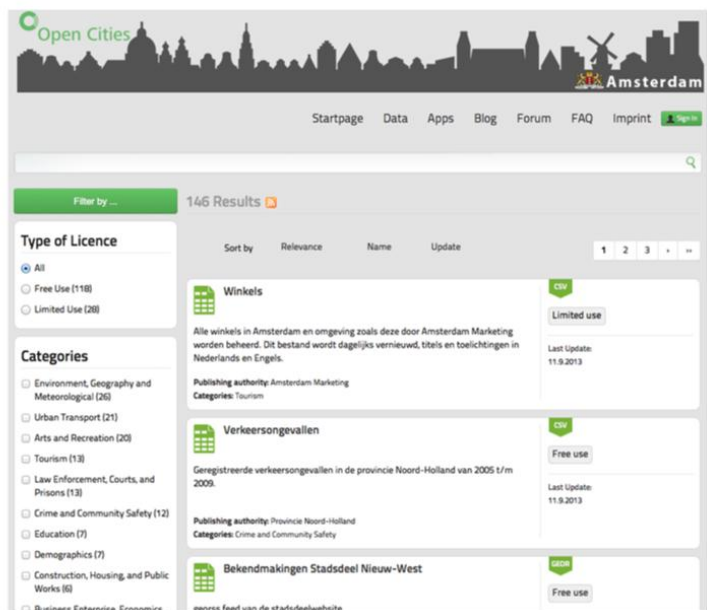


FIGURE 5 RESULTS PAGE OF VERSION 1.0 AND 2.0

Other improvements goes to accessibility. Each image of the platform now have an alt tag. This gives special software the capability to read the whole platform out. It's now possible to navigate through the whole portal only by keyboard. We reduced the usage of javascript so that it is not necessary to keep javascript activated in order to use the portal.

The new theme is more flexible as the old one and can easily be adopted for other portals. Figure 6 shows how the Open Data Portal for GovData the official German Open Data Portal looks like.

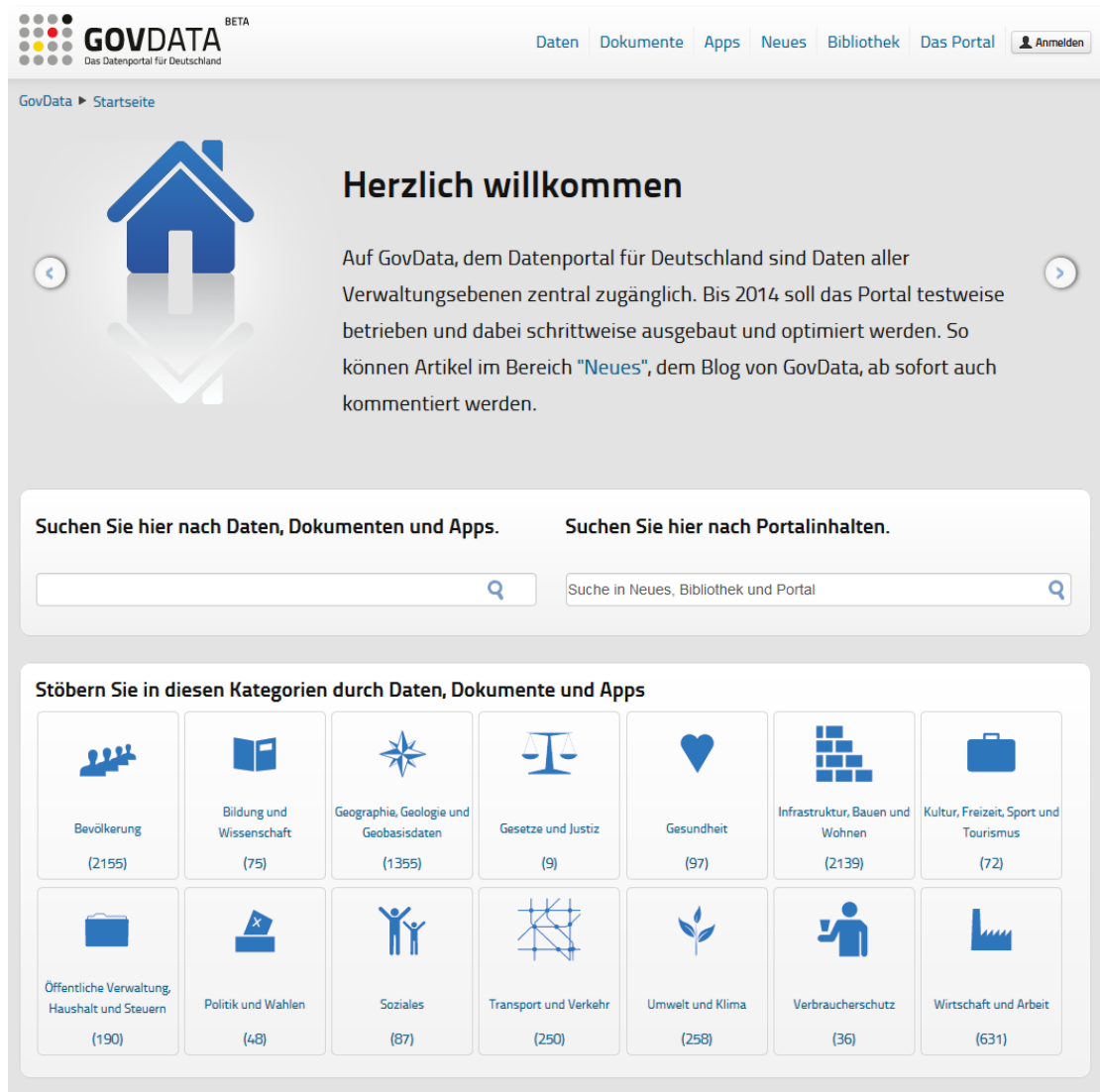


FIGURE 6 GERMAN OPEN DATA PORTAL (WWW.GOVDATA.DE)

4.2.2 New library for connecting with CKAN

We replaced our former simple middleware with a new library which we implemented for our Open Data Platform. This library abstracts from the JSON schema CKAN is delivering to java objects which can be easily used by java developer. The new library reduces requests to CKAN for performing the tasks. This is increasing the performance of the platform.

4.2.3 Updating to CKAN 2.1

CKAN was updated to version 2.1 due to security and performance issues. CKAN 2.1 provides some new API functions which we are using in our new middleware. A new promising feature of newer CKAN versions is the linking between metadata. For example a metadata describing an app can say which data it uses. We are looking forward to integrate those feature in our platform in the future.

4.2.4 Integration of apps and documents catalogue

Besides data version 2.0 is now able to handle metadata describing apps and documents. We implemented a new web form which is able to create or edit apps and documents as well as data. For adding new apps the user don't need any privileges except that it has to be a registered user. Due to spam issues we have implemented a moderated process for publishing apps. Each app announcement needs to be verified by a moderator before it is visible for all users.

4.2.5 Improvement of stability and performance

An issue with version 1.0 was its stability and performance. We addressed that for version 2.0. The new library for CKAN and the usage of newer web frameworks gives us the chance to implement a clean Model View Controller pattern and a cleaner code. This heavily increased the stability of the platform. We also invested time to use new caching mechanisms which are increasing the performance of the portal.

4.2.6 Metadata schema enhancements

With the support of German municipalities we did some small enhancements to the metadata schema. The new schema now fits better to public sector issues. Some changes were necessary to support apps and documents.

The whole new JSON schema can be viewed or downloaded under the following link.

https://github.com/fraunhoferfokus/ogd-metadata/blob/master/OGPD_JSON_Schema.json

4.2.7 Link checker

As in our platform the responsibility for each dataset remains at the data provider we have no influence on the availability on the resources shown in each metadata description. This was an issue because some links could be broken after some time or they had changed. The solution was to develop a link checker. This link checker analysis and tests all resources of all metadata description periodically. After each run it creates a log file showing which resources are not working and the responsible data provider. One have now to contact the data provider to clarify the reasons why his resource is not working.

4.2.8 Maven and GIT for code management

For Version 2.0 we changed from ANT to Maven as a build tool and from SVN to GIT for version control. Maven is used because of its very good dependency management and the seamlessly integration to a continuously integrated environment. GIT is used because of its submodule feature.

Each portlet and library we developed for the Open Data Platform has now its own GIT repository. Beside that there is a main repository which has all other repositories included as sub modules. This works a bit similar like the dependency management of Maven. We are now able to easily create new platforms with our Open Data Platform technology. Not every platform needs the same portlets or features like the others. Therefore we can now create a new parent repository and select those specific portlets and features we need for the new platform. We then just need to add them as GIT submodules.

4.2.9 Automatic code testing with respect to coding conventions and mistakes

During the development of version 1.0 two issues rises up for multiple times. Each developer has his own understanding of coding convention or don't know about common ones. In addition to this experienced developers as well can write common Java mistakes in their code. We addressed this with version 2.0. Thanks to Maven we can run several tests on coding conventions and common mistakes. Because of our continuous integration with Jenkins we are able to run those tests automatically after each push to the GIT remote origin.

4.2.10 OD Platform Bug Fixes

A big issue of version 1.0 was a bug regarding to the old caching mechanism. This caused the platform to crash periodically. To avoid this issue a workaround was used to run a cronjob to restart the platform periodically. Finally this issue is fixed with version 2.0. Besides that many other minor bugs was fixed regarding stability and the user interface

5 DATASET STATISTICS FOR OPEN CITIES DATA CATALOUGE

In addition to the OD platform enhancements, to support the 2nd Open Cities Open Data challenge, the already established Open Cities data catalogue (data.opencities.net) was further maintained and updated with new datasets. Launched in February 2012, the Open Cities data catalogue contained 786 datasets spread over 18 data categories. Currently (October 2013), as shown in Figure 1 the Open Cities data catalogue has 1600 datasets, more than twice the number of datasets of its launch time.

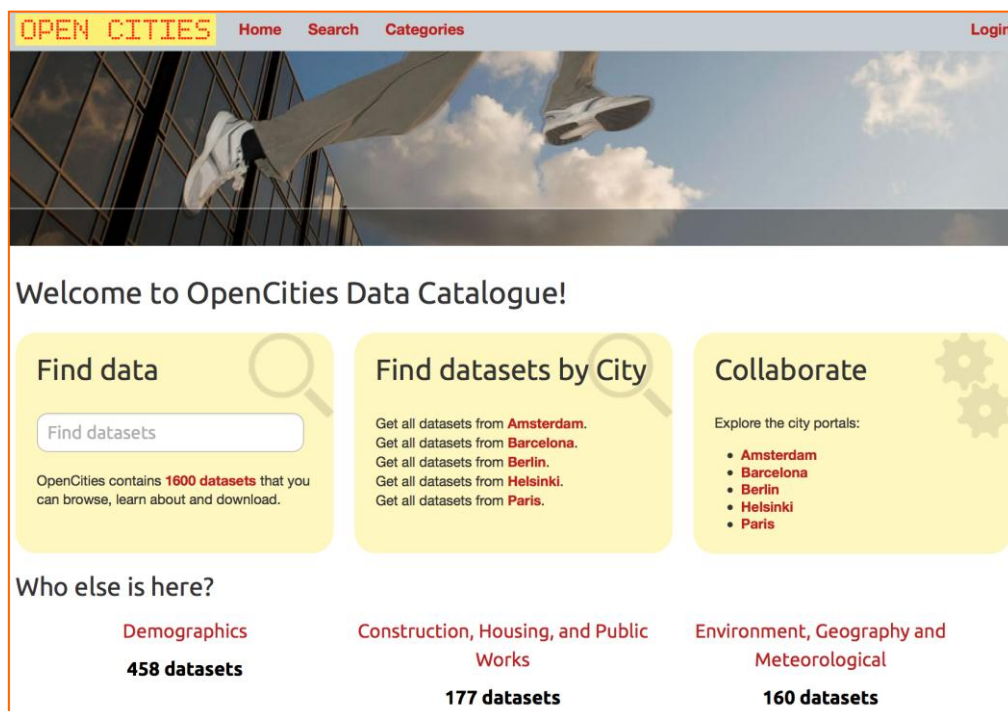


FIGURE 7 DATA.OPENCITIES.NET

Hundreds of new datasets have been harvested from the existing city portals such as:

- Amsterdam -----146 Datasets (out of 304 available at the city portal)
- Berlin -----276 Datasets (out of 276 available at the city portal)
- Paris ----- 54 Datasets (out of 96 available at the city portal)
- Helsinki -----1044 Datasets (out of 1044 available at the city portal)
- Barcelona -----78 Datasets (out of 347 available at the city portal)

Table 1 below shows the current distribution of datasets per category at data.opencities.net. Categories such as Demographics, Construction, Housing, and Public Works have the highest number of datasets.

TABLE 1 DATASETS PER CATEGORY AT DATA.OPENCITIES.NET

Category Name	No. of Datasets
Arts and Recreation	110
Business Enterprise, Economics, and Trade	98
City Budget: Revenues & Expenditures	102
City Portal Web Statistics	1
Construction, Housing, and Public Works	177
Crime and Community Safety	36
Demographics	458
Education	88
Elections	51
Emergency Services	3
Energy and Utilities	55
Environment, Geography and Meteorological	160
Health and Disability	116
Labor Force and Employment Market	150
Law Enforcement, Courts, and Prisons	14
Others	55
Political	21
Tourism	23
Urban Transport	123

5.1 DATASETS STATISTICS UPDATES

Figure 8 below shows the number of published datasets per city that were made available at data.opencities.net, in February 2012, December 2012 and October 2013. As shown here, Helsinki continues to have the highest number of datasets, following by Berlin, Amsterdam, Paris and Barcelona.

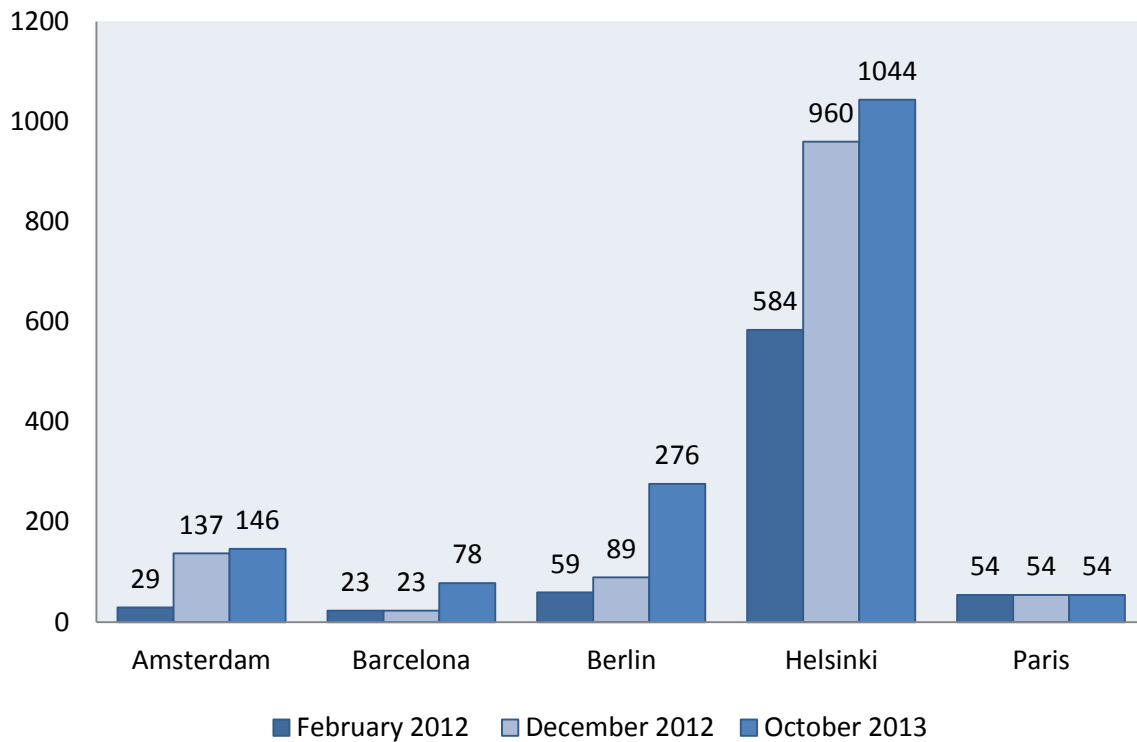


FIGURE 8 NUMBER OF DATASETS PER CITY (FEB 2012 / DEC 2012 / OCT 2013)

5.1 DATA.OPENCITIES.NET – PORTAL USAGE INDICATORS IN 2013

As a total number of visits, the portal has reached 4,293 visits in 2013. There have been 53,721 pages served to 1,711 unique visitors. Figure 9 below shows graphically the monthly distribution of data.opencities.net portal usage statistics in 2013, such unique visitors, number of visits, pages, hits and served bandwidth. As shown in **Error! Reference source not found.**, it has been a peak of sage of data.opencities.net from May till September 2013. These numbers can be very well explained with the fact that the 2nd Open Data challenge has been running at those months.

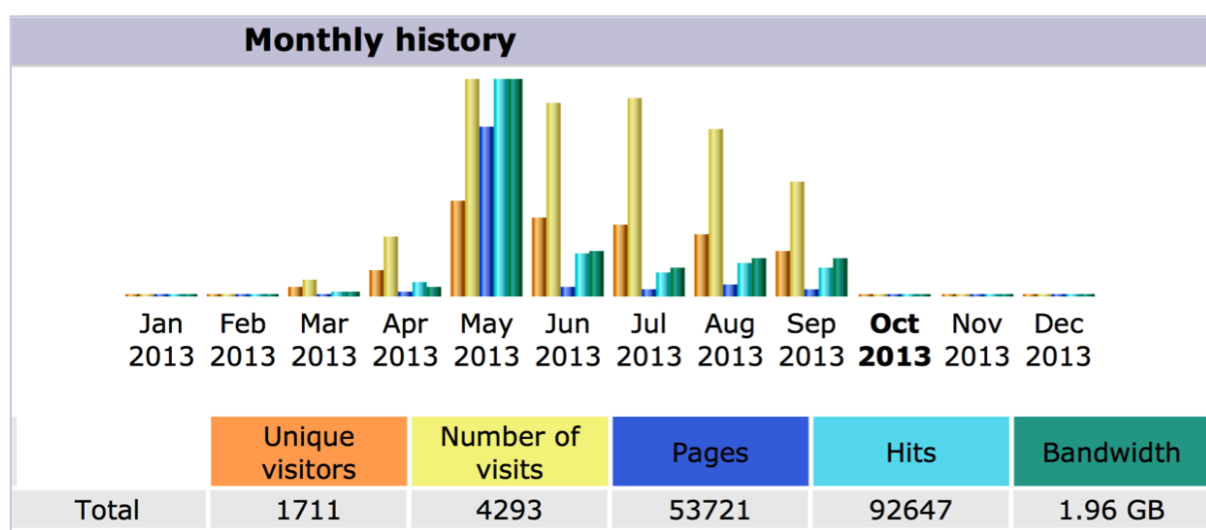


FIGURE 9 DATA.OPENCITIES.NET – PORTAL USAGE INDICATORS IN 2013

6 SYNERGIES WITH OTHER PROJECTS

The OD Platform or parts thereof as described in this and previous deliverables could be re-used in different contexts. The platform was introduced in projects at the local, national and EU level. The following chart provides an overview of these projects.

TABLE 2 OVERVIEW OF SYNERGIES WITH OTHER PROJECTS

Project title	Project short description	Administrative level	Project runtime
Policy Compass, FP7-ICT-2013-10, ICT for Governance and Policy Modelling	The main goal of Policy Compass is to develop an easy-to-use tool for social networks and e-participation platforms, enabling citizens and public officials to easily create and discuss causal models of historical data from trusted open data sources. The aim is to develop methods and tools that facilitate more factual, evidence-based, policy evaluation and analysis.	EU	Fall 2013 – Fall 2016

Streetlife, SMARTCITIES-2013(ICT), Integrated personal mobility for smart cities	FP7-	The goal of the STREETLIFE project is the reduction of carbon emissions and accomplishment of intelligent mobility systems in cities through advanced Future Internet and Cloud Computing techniques.	EU	Fall 2013 – Fall 2016	
iCity, programme, Policy Support	CIP ICT	Opening-up public infrastructures in urban spaces promoting the co-creation of services of public interest	EU	Begin 2012 – End 2014	
GovData – National Platform Germany	The Data for	Conception and development of the German national data portal on behalf of the Ministry of the Interior	National	Summer 2012 – Spring 2013	
Open Hamburg	Data	Conception and support of the development of the Hamburg Open Data Portal on behalf of the Hamburg Senate of Finance	Local	Fall 2012 – Spring 2013	

Additionally experience from conception and implementation of the OD Platform within the Open Cities project and its partaking cities provided valuable experience for further projects in Germany, where Fraunhofer FOKUS was asked to analyse and evaluate local conditions and develop scientific concepts for implementation of local or regional Open Data platforms.

7 SUPPORTING THE 2ND OPEN DATA CHALLENGE

The Open Cities project undertook in 2013 the 2nd Open Data Challenge. Fraunhofer FOKUS again supported the challenge.

The Open Cities Data Catalogue (data.opencities.net) was further maintained so that the challenge participants could fully utilize its functionality as a basis for creating their apps. In addition where possible new datasets were harvested from the existing city portals. This was realized in cooperation with the respective representatives from the project municipalities.



FIGURE 10: SCREENSHOT FROM THE OPEN CITIES HACK AT HOME CHALLENGE WEBSITE

The 2nd Open Data Challenge was adapted on the basis of the experience from the first challenge. One of the objectives was to provide more support to the contest participants. For this mentors from the project partners were selected in order to present contact persons. Fraunhofer FOKUS supported this with two mentors, which gave feedback to app developers and helped evaluating the submitted apps together with the other mentors of the project.

8 CONCLUSION

We presented the new features and enhancements of the Open Data Platform which we implemented to achieve the goal of task 4.4.6. Most namely the usage of a thesaurus with the EuroVoc vocabulary and the faceted search on the results page. Together they improved the search capability of the platform and increases the metadata quality. As mentioned in chapter 4.2.3 a next step could be to link metadata to each other.

The statistics of the Open Data Catalogue shows that almost all cities increased their number of datasets in their portals. This is especially the case for Helsinki and Berlin. The usage statistics shows that there was a lot of activity in May 2013 with decreasing tendency.

Due to the technical architecture we are able to use the Open Data Platform in several other national and EU level projects. The knowledge Fraunhofer FOKUS gained through the work in Open Cities is highly requested in Germany and we will use this experience to support or create new Open Data Platforms.

For the 2nd Open Data Challenge we increased the number of datasets in the Open Data Catalogue and supported the challenge with two mentors.