COMPONENT	5. OPEN DATA
PHASE	5.2 GUIDELINES OPEN DATA
PHASE LEADER	INSTITUTO PEDRO NUNES
DELIVERABLE	GUIDELINES OPEN DATA
VERSION	VERSION 1.0 (OCTOBER 9TH, 2014)

TABLE OF CONTENTS

TABLE OF FIGURES

TABLE OF ACRONYMS

ACRONYMS	DESCRIPTION
API	Application Programming Interface
CC	Creative Commons
CC0	Creative Commons Zero
CKAN	Comprehensive Knowledge Archive Network
CSV	Comma-Separated-Values
EU	European Union
IT	Information Technology
JSON	JavaScript Object Notation
OD	Open Definition
ODI	Open Data Institute
OI	Open Institute
OK	Open Knowledge
OKFN	Open Knowledge Foundation
ODaF	Open Data Foundation
PDF	Portable Document Format
PSI	Public Sector Information
RDFa	Resource Description Framework in attributes
URI	Uniform Resource Identifier
USA	United States of America
XML	eXtensible Markup Language

1. INTRODUCTION

Open data refers to the idea that certain data should be freely available to everyone to use and re-use. There has been a continuous increase in people joining the open data movement. This increase brought the topic to political and economical debate and the results are quite encouraging. A recent European Commission Communication on open data [1] predicts that overall economic gains from opening up public data could amount to €40 billion a year in the European Union (EU). Other proof of the increasing importance of the open data is the presence of the theme in the Agenda for Europe – A Europe 2020 Initiative in the area of Content & Media. The main goal of the open data movement is to foster the transparency and accountability of institutions, governments, and other bodies, as well as to contribute to the continuous increase of the integration and interoperability of the Internet.

In this document we'll guide you through the main concepts of open data, its objectives and benefits, and history. Also, we'll describe some of the most important organizations that promote open data and present visible projects on open data in European area and in other continents. Finally, to close the common ground on open data, we'll talk about licensing issues and explain the main steps on how to open data and describe some web spaces where you can publish your data. Next, we will present some valuable products and services developed using open data. In this section we will demonstrate the business opportunities that can arise from the opening of data. Lastly, we'll describe the activities developed in the scope of CITEK project and provide a step-by-step guide on how to publish a dataset in the CITEK's open data web space.

Due to the wide spread focus of the information presented in this document we believe it can be useful for an audience that wants to know what is open data. So, we think this document as a good introduction for developers, educators, decision makers, public bodies, investors, experienced entrepreneurs, innovators, young students, recent graduates, young professionals, or the curious citizen interested in open data and its potential to disrupt markets.

2. COMMON GROUND ON OPEN DATA

2.1. INTRODUCTION

In this section, we'll lay the foundation for readers who are unaware of what open data is and how it can be of use to one's benefit. We'll first settle on a concise and objective definition of open data and we'll take on a journey into the past to understand how the concept emerged and how the movement became what it is today.

The goals of the open data movement are very broad and comprise different domains that we shall summarize and analyse in the light of the benefits that it provides. We'll also see how open data has been transforming governments around the world despite the obstacles that may exist, setting the stage for better innovation and greater collaboration with and among citizens.

Further down the document, we'll take a look at some successful examples of organizations opening data around the world, as well as some applications and services that emerged as a result of it. Finally, we'll guide the reader into the licensing models available regarding open data, which sets the stage for the remaining sections.

WHAT IS OPEN DATA?

The most common definition of open data stands as follows:

«a piece of data or content that can be freely used, re-used and redistributed by anyone – subject only, at most, to the requirement to attribute and share-alike» [2].

Simply put, open data refers to information that is made available and free to use. This definition may apply to domains such as science, education, transportation, or government. The underlying principles of open data are built on the need for transparency, collaboration, and open access to knowledge in a way that makes societies more actionable and stronger. These principles are the also the main drivers of the open data movement, which comprises hundreds of organizations and activists from around the world working on and lobbying for open data.

The Open Knowledge Foundation extended definition of open data [3] defines detailed criteria that must be met, divided into 7 points:

1. ***Access**: The work shall be available as whole and at no more than a reasonable reproduction cost, preferably downloading via the Internet without charge. The work must also be available in a convenient and modifiable form.*
2. ***Redistribution:** The **license** shall not restrict any party from selling or giving away the work either on its own or as part of a package made from works from many different sources. The license shall not require a royalty or other fee for such sale or distribution.*
3. ***Reuse:** The **license** must allow for modifications and derivative works and must allow them to be distributed under the terms of the original work.*
4. ***Absence of Technological Restriction:** The work must be provided in such a form that there are **no technological obstacles** to the performance of the above activities. This can be achieved by the provision of the work **in an open data format**, i.e., one whose specification is publicly and freely available and which places no restrictions monetary or otherwise upon its use.*
5. ***Attribution:** The **license** may require as a condition for redistribution and re-use the attribution of the contributors and creators to the work. If this condition is imposed it must no be onerous. For example if attribution is required a list of those requiring attribution should accompany the work.*
6. ***Integrity:** The **license** may require as a condition for the work being distributed in modified form that the resulting work carry a different name or version number from the original work.*

7. ***No Discrimination Against Persons or Groups:*** *The* ***license*** *must not discriminate against any person or group of persons.*
8. ***No Discrimination Against Fields of Endeavour:*** *The* ***license*** *must not restrict anyone from making use of the work in a specific field of endeavour. For example, it may not restrict the work from being used in a business, or from being used for genetic research.*
9. ***Distribution of License:*** *The rights attached to the work must apply to all to whom it is redistributed without the need for execution of an additional* ***license*** *by those parties.*
10. ***License Must Not be Specific to a Package:*** *The rights attached to the work must not depend on the work being part of a particular package. If the work is extracted from that package and used or distributed within the terms of the work's* ***license****, all parties to whom the work is redistributed should have the same rights as those that are granted in the conjunction with the original package.*
11. ***License Must Not Restrict the Distribution of Other Works****: The* ***license*** *must not place restrictions on other works that are distributed along with the licensed work. For example, the license must not insist that all other works distributed on the same medium are open.*

The aforementioned criteria set the rules for governments, city councils, and companies to publish data that promotes access and reuse.

WHAT DATA CAN I OPEN?

Most types of information are eligible for open access, especially those that originate from within the public sector. In modern societies, data from governments and cities are, in fact, generated and supported by taxpayers' money and, because of that, are of public interest. Relevant data may include information such as how money is spent — for transparency and accountability — or cartography representing the territory people live in.

The Public Sector Information Directive (PSI) identifies a large set of information that European Union (EU) countries are recommended to make available, but there is no known broad definition for what may be considered relevant to open up. However, the Open Knowledge Foundation has identified several kinds of data that might foster re-use:

- Cultural: Data about cultural works and artifacts — for example titles and authors — and generally collected and held by galleries, libraries, archives, and museums.
- Science: Data that is produced as part of scientific research from astronomy to zoology.
- Finance: Data such as government accounts (expenditure and revenue) and information on financial markets (stocks, shares, bonds etc.).
- Statistics: Data produced by statistical offices such as the census and key socioeconomic indicators.
- Weather: The many types of information used to understand and predict the weather and climate.
- Environment: Information related to the natural environment such as water pollution levels in rivers and oceans.
- Transport: Data such as timetables, routes, on-time statistics.

It is important to notice that open data does not include personal data. Although there are open data initiatives that collect data directly from one's smartphone or any other sensors, the data itself ought to be completely dissociated from the individual that played a role in generating it, thus discarding any trace of personal information.

THE VISION OF THE OPEN DATA MOVEMENT

The open data movement strives for stronger societies that leverage open access to information to everyone's benefit. Activists and lobbyists for open data around the world work hard everyday to generate new datasets and enrich the current ecosystem. Collected information expands and enlarges the social and economic opportunities that follow, both, for citizens as well as for companies. These opportunities may be divided into three areas:

- **Transparency**: free access to government data allows citizens to better understand what their representatives do (and how they do it) and provides the means to hold them accountable. It also allows governments to promote transparency towards citizens;
- **Civic engagement:** open data promotes civic participation by providing more accurate information and thus more informed opinions. Data itself is often too complex to understand, but it may be parsed and interpreted in order to generate visualization representations (e.g. infographics);
- **Social and commercial value:** citizens, developers, and companies should be able to find data that unlocks new social or business opportunities. Data itself lacks value, but it may often be an enabler of applications and services that meet specific user needs (e.g. open meteorology data in the US has enabled the emergence of multiple weather forecasting applications).

It can be said that the open data movement aims to accomplish some of the same goals as the open source movement. Both take on the premise that it is possible to create value not by closing code or data but by making them free to reuse as building blocks for entrepreneurial endeavours.

THE OPEN DATA FORMATS

Any piece of information that is made freely available for reuse may be considered open data. However, the more accessible the data, the more likely it will be useful and reused. This principle is the foundation of the "open data five star rating" system, introduced by Sir Tim Berners-Lee (Director of the World Wide Web Consortium and evangelist for Linked open data) at the Gov 2.0 Expo 2010, on 26th of May 2010 [4].

The rating model establishes five different levels of open data quality, as follows:

1. Data is accessible on the web and readable by humans, but not a software agent because it is in a closed format and not easily re-used (e.g. an image of a scanned document);
2. Data is accessible on the web in a structured machine readable format, however still depends on proprietary software to be accessed (e.g. Word or Excel files);
3. Data no longer relies on proprietary software to be used; Users can manipulate data in any way they please (e.g. Comma Separated Values (CSV) file);
4. Data is in the web and may be linked to with a URI. It gives fine-granular control over the data (e.g. Resource Description Framework in attributes (RDFa) file containing 3 Uniform Resource Identifiers (URIs));
5. Data is not only in the web but is also linked to other data. Through interlinking, data gets interconnected and becomes discoverable from other sources (e.g. RDFa file containing URIs and semantic properties).

In summary, the easier it is to access and reuse data, the more stars it gets. This includes data being represented using open standards that require no proprietary tools to use it, such as CSV, JavaScript Object Notation (JSON) or eXtensible Markup Language (XML) files, that both people and machines are able to read for free.
It is also important to notice that the more stars one aims, the more responsibilities one accepts towards those who are using the data. This means that high-quality open data is usually harder to generate, to update and to maintain within the organization providing it, and may pose a difficult organizational challenge.

ORGANIZATIONAL ISSUES

Publishing open data belonging to an organization may range from a simple, automatic, click to a complex coordinated procedure involving several internal departments, depending on the existing Information Technology (IT) infrastructure and overall readiness and alignment with open data strategies. Above all, it must be accompanied by a cultural shift to a new paradigm where there is no control over how data is used once it is published.

Organizations looking to position themselves as frontrunners and open data drivers should be able to coordinate internally in several fronts:

- The board;
- The departments and people that generate or update data;
- The IT department;
- The legal department.

The board must define the data strategy for the organization and drive its members to action. It must also coordinate internally which resources should be involved and how they interact with one another, in order to keep track of its costs and benefits over time. The board should also define the policies for publishing data, clearly identifying what data can be published, how it should be updated and by what mechanism (manual or automatic), data granularity, etc.

Regarding the public sector, the "open data by default" motto from the 2012 "Project open data" initiative by the White House stands as the example to follow. It is defined as follows [5]:

«To promote continued job growth, Government efficiency, and the social good that can be gained from opening Government data to the public, the default state of new and modernized Government information resources shall be open and machine readable.»

Having a mission statement such as the one in the above sentence is invaluable in communicating the organization a long-term vision regarding open data, and is something that goes a long way in engaging people to follow.

The departments where data gets generated and updated should be aligned with the open data strategy of the board and must be trained to implement it on a daily basis. This includes becoming proficient with new tools or processes and learning how to deal with feedback. Listening to open data consumers is an important step towards creating a feedback loop between those who publish data and those who use it, in order to improve the process.

The internal IT department should be able to prepare the infrastructure for making it simple and efficient to publish data. It should also develop mechanisms that allow it to be done without disrupting the existing methodologies and processes, in collaboration with the people who generate or update the data. Training programmes should be promoted, if necessary.

Finally, the internal legal departments must be ready to deal with open licensing, so that licences are chosen wisely for each dataset or data type. This includes being acquainted with the aforementioned open definition, the open data Commons, the Creative Commons, among other licensing models.

2.2. OPEN DATA OBJECTIVES AND BENEFITS

Simply put, the open data movement aims to promote the availability of data and its reuse in a variety of domains such as education, transportation, and government. It consists of researchers, political activists, civil servants, journalists, developers, and entrepreneurs, all having a different perspective on how data is valuable to their activity or craft. Open data is therefore subject of interest for many different reasons and its benefits are found to be very diverse, depending on what data is being used and for what purpose.

On broad terms, the main benefits of open data allude to the potential reuses of data, both social and economical. Making data freely available is simultaneously both a goal in itself as well as a means to an end considering re-use and re-distribution. It is currently impossible to estimate what new combinations of data may create new knowledge, insights, and businesses, and what their impact will be.
The Open Knowledge Foundation identifies several areas where open data – and more specifically, open government data – is an enabler of value creation:

- Transparency and democratic control;
- Participation;
- Self-empowerment;
- Improved or new private products and services;
- Innovation;
- Improved efficiency of government services;
- Improved effectiveness of government services;
- Impact measurement of policies;
- New knowledge from combined data sources and patterns in large data volumes.

On a non-economic perspective, open data benefits relate to the new possibilities regarding social and political activism. Having free and reliable data, citizens are able to better understand how money is spent or even what their representatives are doing. This kind of transparency is invaluable in promoting civic action and in bringing citizens and governments closer together.

Economically, open data benefits range from the emergence of products and services that improve citizens' quality of living to job creation (especially in the IT industry). New products and companies are re-using open data in ways one could have not predicted before. For instance, the fact that meteorological data became open allowed for the development of weather applications that would not have existed otherwise. As a result, at the time of writing of this document (August 2014), the iPhone app store alone returns over 2000 results for weather apps, using the word "weather" as the search term. Scientific research also benefits tremendously from open access to information, which allows more comprehensive analyses over a wider scope of data. The more data is available, the more accurate the conclusions drawn from research.

All in all, open data plays an important role in promoting economic and social innovation, which in turn is a driver of better societies and sustainable development.

2.3. OPEN DATA HISTORY

The concept of open data goes back to the early 1940s, when the Robert King Merton, one of the fathers of sociology of science, came up with the idea that research data should be shared freely for the common good. If every researcher contributed to a "common pot", that way, science as a whole would move faster and produce results more rapidly. While this may have set the beginning of the open access concept, open data as an expression didn't emerge until 1995, in a document from an American scientific agency on geophysical and environmental data. It stated that *«Our atmosphere, oceans and biosphere form an integrated whole that transcends borders»* meaning that global phenomena could only be understood and acted upon if countries exchanged scientific data amongst themselves, towards a scope as large as the study subject. Researchers were the first to understand the potential benefit of openness and of sharing of data.

Fast forward to the 21st century, specifically the year of 2007. Several prominent researchers, activists and technologists gathered in Sebastopol, California, aiming to discuss and come to a common definition of what open public data is. This group included people that mostly came from the free software and culture movements, such as Tim O'Reilly, Lawrence Lessig (Harvard Law School Professor and creator of Creative Commons licences), Adrian Holovaty (web entrepreneur and co-creator of the Django framework) and Aaron Swartz (political Internet activist). In Lawrence Lessig's words:

«*The objective was to find a simple way to express values that a bunch of us think are pretty common, and these are values about how the government could make its data available in a way that enables a wider range of people to help make the government function better*»

The thirty people who gathered in Sebastopol started promoting this idea that open public data could be used as an enabler of new insights, businesses and civic action, based on the principles and success of the open source movement, of which many of them were aficionados. This is considered to be the day the open data movement was born.

The first big step happened later in 2009, when recently elected Barack Obama signed the Memorandum on Transparency and Open Government, which paved the path towards a making open data a reality in the public administration. This was later reinforced, in 2013, when an executive order was issued setting the open data status as the new default regarding government information. Its goal was to create a framework to «*help institutionalize the principles of effective information management at each stage of the information's life cycle to promote interoperability and openness*». This directive also set the beginning of "Project open data", a collection of code, tools, and case studies hosted on github.com «*to help agencies adopt the open data Policy and unlock the potential of government data*».

Infographic by FCW – The Business of Federal Technology [36]

2.4. OPEN GOVERNMENT

Although the open data and open access concepts are rooted in the scientific community, its principles rapidly expanded to other domains such as education, transportation, and government. Government in particular has been one of the most targeted domains from open data movements from all around the world, «*because of the quantity and centrality of the data it collects, but also because most of that government data is public data by law, and therefore could be made open and made available for others to use*» [6].

Open Government is also about enabling new tools, services, and businesses on top of raw data, which in turn empower societies to act and collaborate. The emergence of services – made possible because of open data – developed by citizens or private companies represents a shift in how societies interact with governments and is reshaping our current model of government and associated roles.

Up until now, we have lived «*in a world of 'vending machine' government*» [7]. We pay our taxes and we expect services to be dispensed in return. When we are unsatisfied with the services we get, our 'participation' is limited to protest, which essentially amounts to 'shaking' the 'vending machine' [8]. This model is quickly becoming obsolete. Nowadays, citizens are able to do so much more for themselves and others, thanks to having the necessary skill sets, the connectivity and, of course, the data. The conditions finally exist to radically rethink and redesign the roles of both society and government. Open government no longer consists of a vending machine. Instead, it enables and nourishes a marketplace of applications and services that uncover, unlock and unleash the potential of open data.

2.5. OPEN DATA IN THE WORLD

Open data initiatives from around the world have proliferated in recent years and have covered a wide range of domains such as government, transportation, geographical mapping, etc. Huge steps have been taken especially in the USA, thanks to enforcing legislation on open access and open data. The White House memo of 2012, in particular – stating that government data should open by default – has set a new standard for governments around the world to take a stand in supporting the open data movement and promote innovation.

Some of the most prominent open data initiatives from around the world are listed next, along with a short description of what they intend to achieve.

OPENSTREETMAP

https://www.openstreetmap.org/

OpenStreetMap (OSM) is a global scale initiative to geographically map the planet through crowdsourcing. The OSM community provides tools and support to anyone looking to contribute and edit the map. The data collected is available for API (Application Programming Interface) use or download with an open licence.

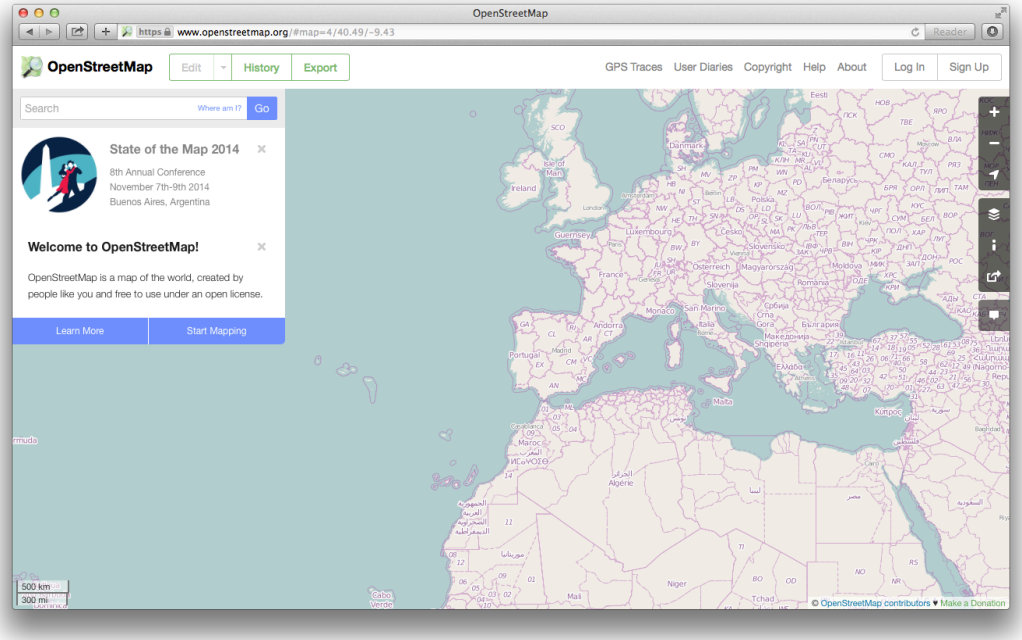


FIGURE 1 – OPENSTREETMAP.ORG

DATA.GOV

https://www.data.gov/open-gov/

Data.gov is «*the central site for U.S. Government data, [...] an important part of the Administration's overall effort to open government*» and it launched in 2009. It contains over 112 594 datasets from 38 USA states and 46 counties covering a wide range of topics such as energy, safety, agriculture, etc. A spin-off initiative was created later in 2013 called "Project open data". It provides a «*collection of code, tools, and case studies to help agencies adopt the open data policy and unlock the potential of government data*».

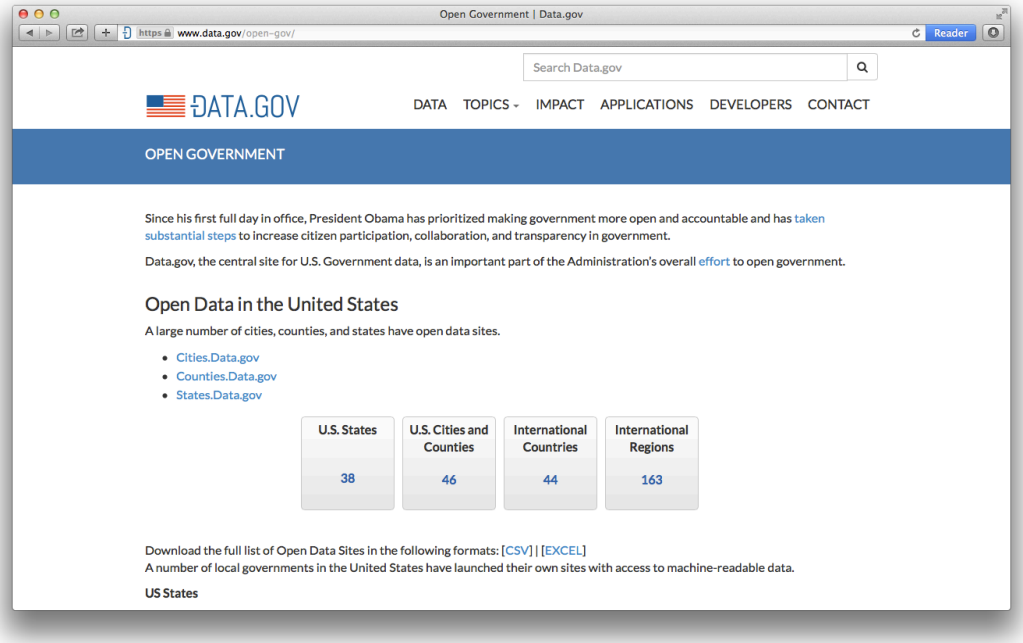


FIGURE 2– DATA.GOV

OPEN GOVERNMENT PARTNERSHIP

http://www.opengovpartnership.org

The OpenGov Partnership (OGP) is an international initiative from all continents aiming to implement open government strategies and promote innovation, civic participation, and transparency. In 2014, more than 60 countries embody the partnership and work together towards designing actionable open government reforms.

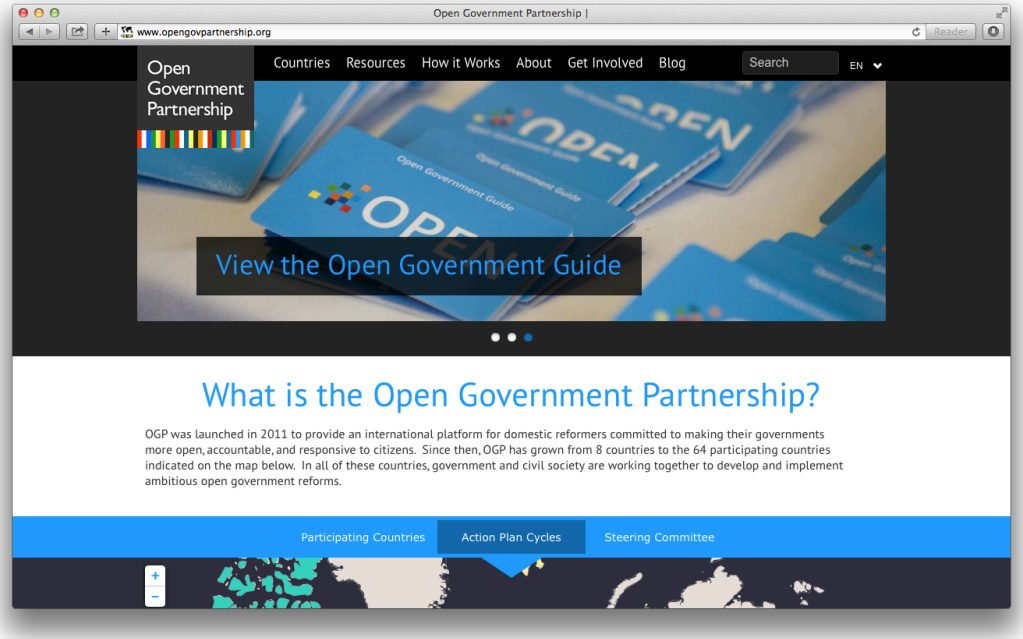


FIGURE 3 – OPENGOVPARTNERSHIP.ORG

DATAHUB

http://datahub.io/about

The datahub is an open data platform maintained by the Open Knowledge Foundation that aggregates static datasets from organizations around the world. It also allows users to search for data and access it freely, either by downloading content or using its API. All in all, it aims to become a collaborative effort that brings together publishers and consumers into the same platform.

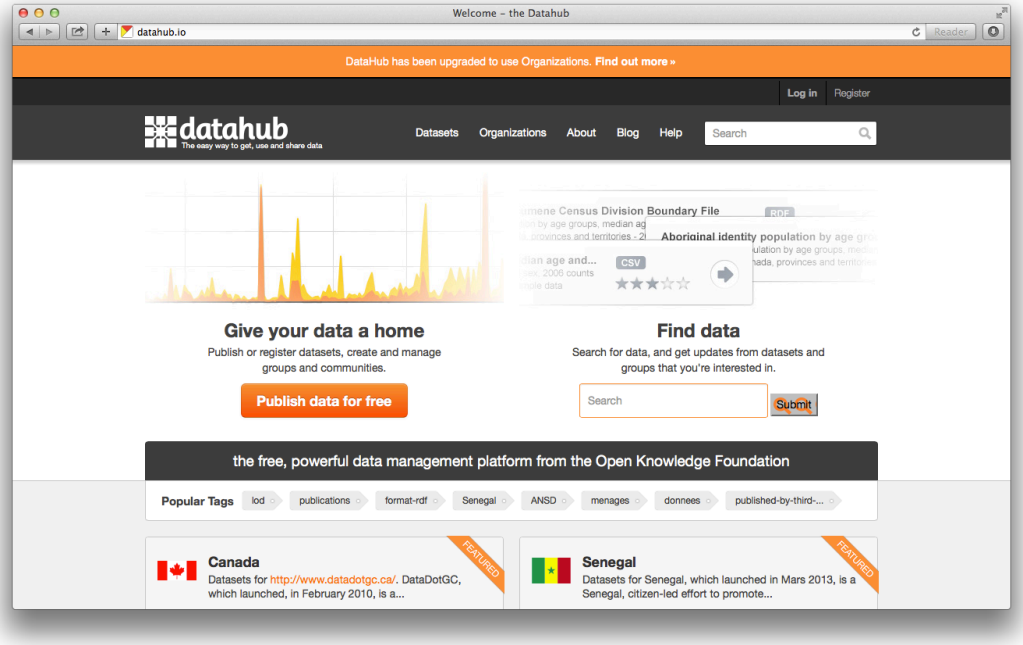


FIGURE 4 – DATAHUB.IO

OPEN BANK PROJECT

http://www.openbankproject.com/en/

The Open Bank Project aims to create an ecosystem of apps based on user banking data. To achieve that, it develops and provides support to an open source infrastructure that banks may use in order to make certain data available to user-facing apps. The user is required to authorize access to data on a per-application basis. Considering that all banking data is private, this does not classify as open data but open access instead.

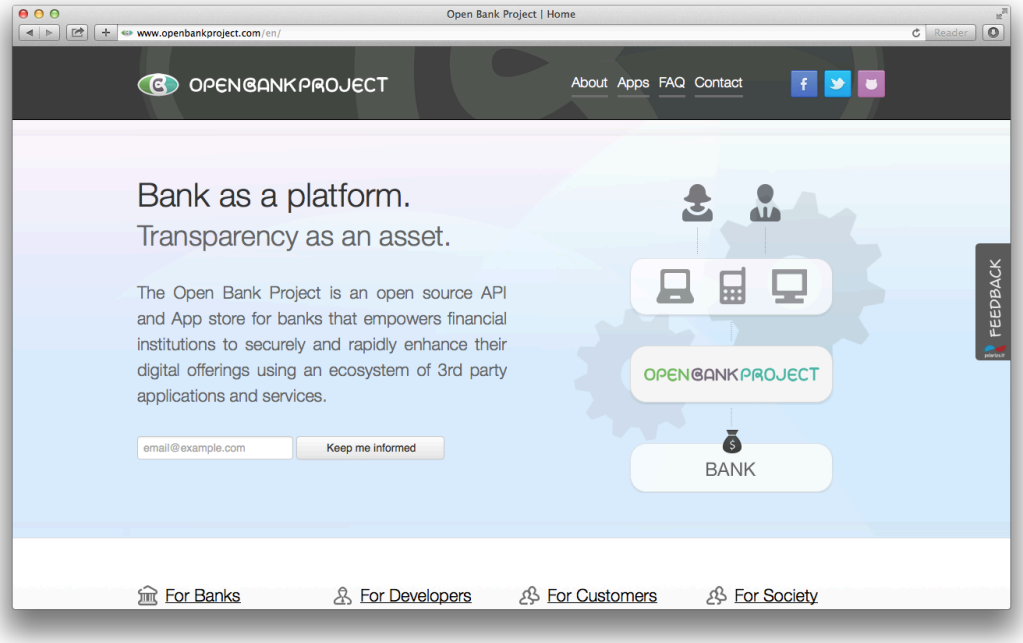


FIGURE 5 – OPENBANKPROJECT.COM

SCIENCEOPEN

https://www.scienceopen.com/home

ScienceOpen is a research and publishing network of scientific information. It consists of a Web platform where researchers around the world are able to share their work for others to access and evaluate.

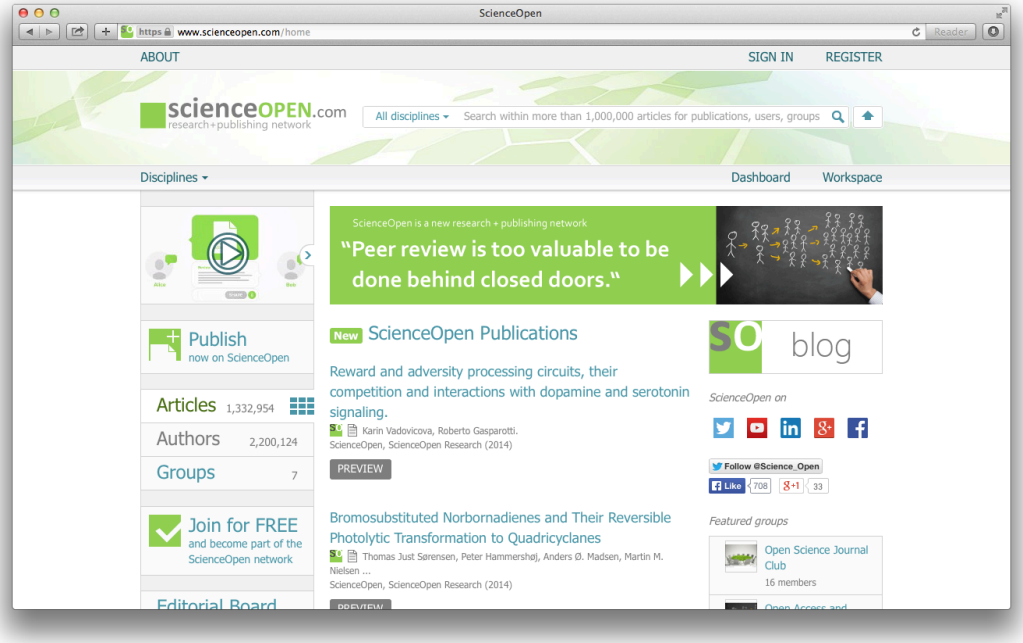


FIGURE 6 – SCIENCEOPEN.COM

OPENSPENDING

https://openspending.org

The OpenSpending project aims to «*track and analyse public financial information globally*» as a means to promote government financial transparency. The premise is that citizens should know how governments spend money on their behalf, which sets the stage for political accountability, corruption fighting, fraud detection, and citizen engagement.
The "Where does my money go" [9] project is part of the OpenSpending initiative and provides nice visualizations for reading and interpreting financial data from the British government.

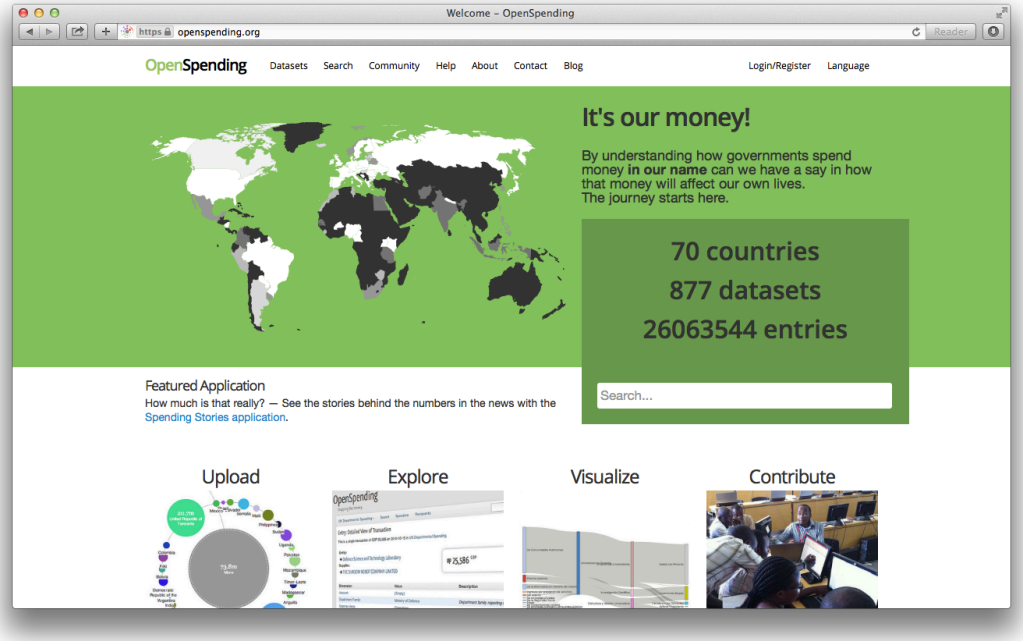


FIGURE 7 – OPENSPENDING.ORG

2.6. OPEN DATA IN EUROPE

Open data adoption in Europe has advanced slowly but at a steady pace. Northern European countries in particular have been taken important steps in promoting innovation and transparency in government and are currently the model to follow in this regard. Some of the most prominent European data initiatives are listed next.

OPEN METEO FOUNDATION

http://openmeteofoundation.org

OpenMeteoData is an open data European initiative for meteorological data. The project launched in 2012 and already provides tools and services for developers to reuse weather data for third-party applications.

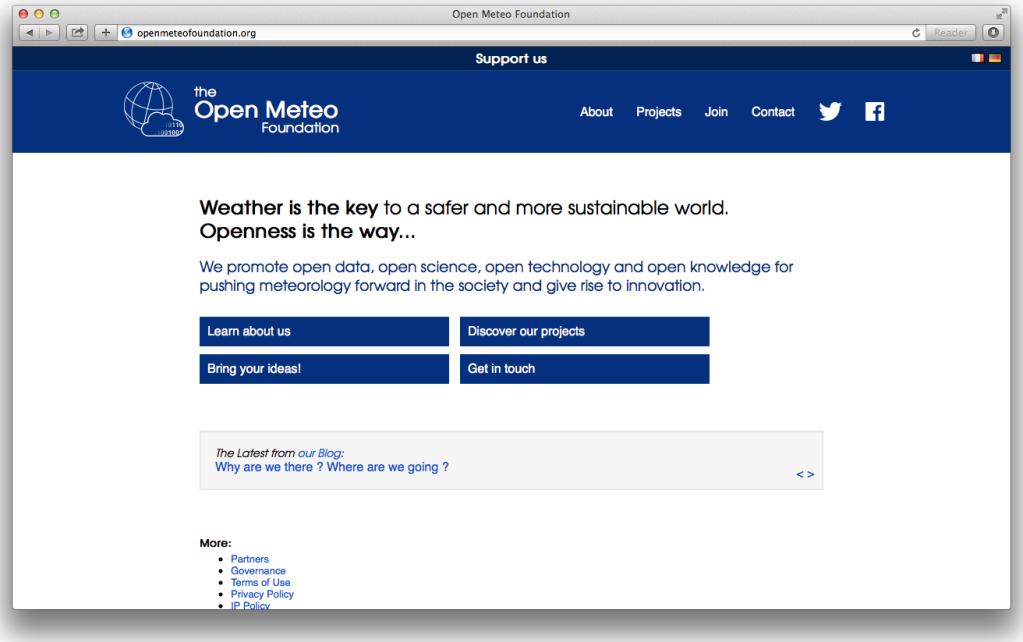


FIGURE 8 – OPENMETEOFOUNDATION.ORG

PUBLICDATA.EU

http://publicdata.eu

PublicData.eu is a Pan European data portal that collects and aggregates open and freely reusable datasets from public bodies across Europe. The portal itself is a federation of other official data portals from several countries, whose data is indexed and made available to search and access. PublicData.eu is currently maintained by the Open Knowledge Foundation.

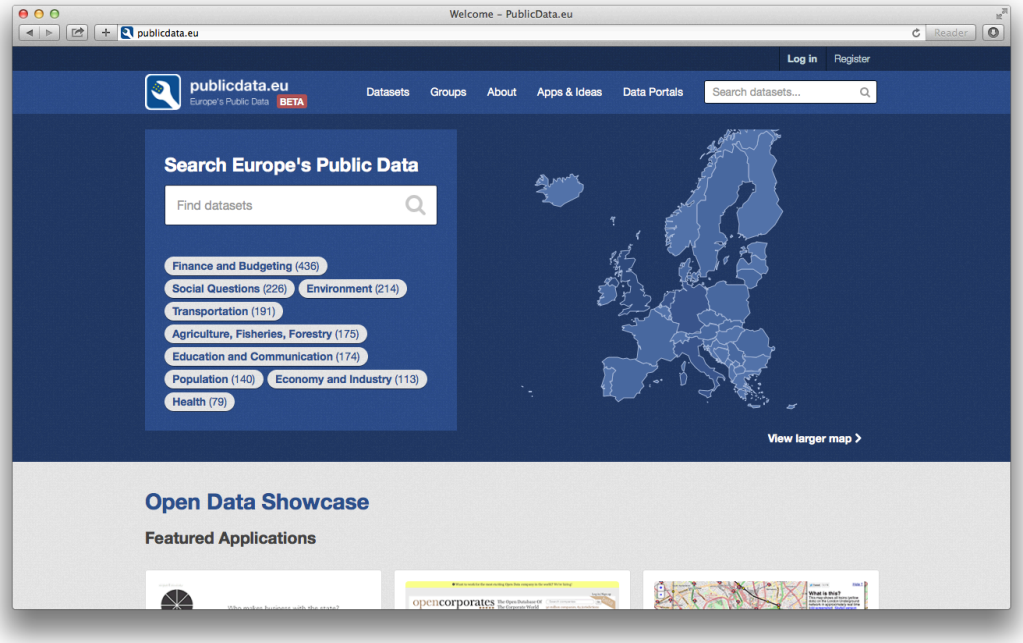


FIGURE 9 – PUBLICDATA.EU

DATA.GOV.UK

http://data.gov.uk

Data.gov.uk is the data.gov counterpart for the US data.gov portal, which pioneered the concept worldwide. It contains a large number of datasets, which range from public roles and salaries to spending reports and even near real-time transportation data for highways. The portal is based on Comprehensive Knowledge Archive Network (CKAN), an open source tool for publishing open data from the Open Knowledge Foundation.

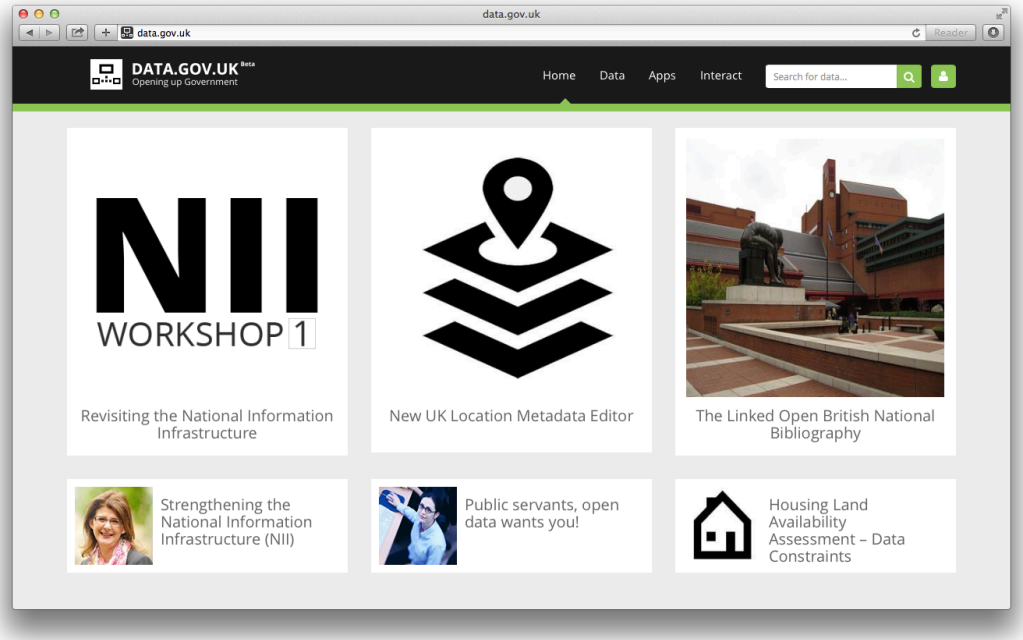


FIGURE 10 – DATA.GOV.UK

CITYSDK

http://www.citysdk.eu

CitySDK is a European project aiming to design and creating a unified service toolkit for smart cities. It comprises three different domains: smart mobility, smart tourism, and civic participation. To that effect, «*CitySDK provides better and easier ways for the cities [...] to release their data in a format that is easy for the developers to re-use*». This includes publishing mass transit static data, points of interest, events, amongst other relevant geographic information.

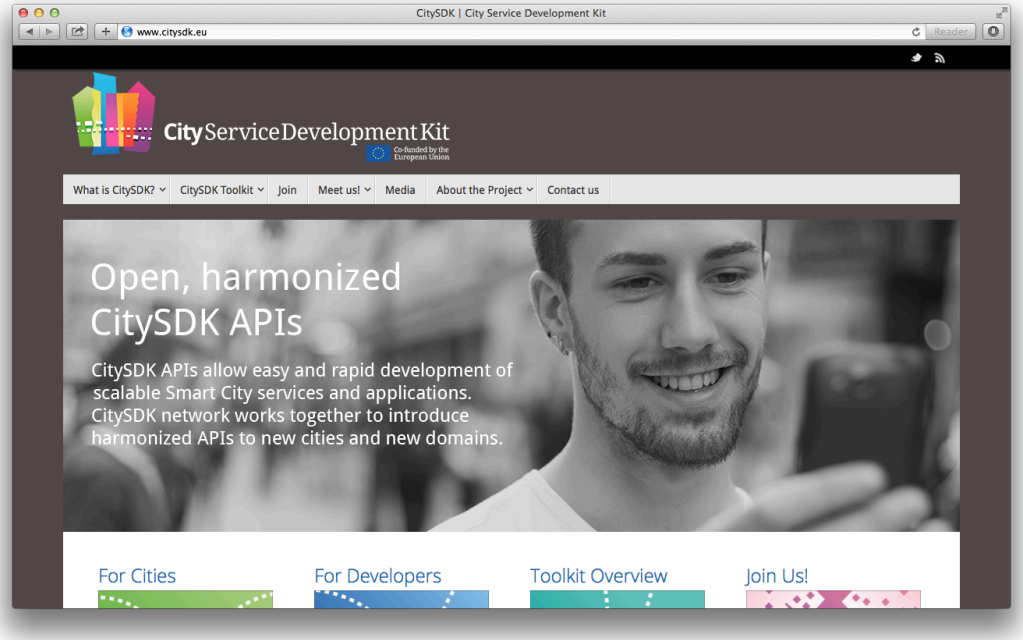


FIGURE 11 – CITYSDK.EU

HOMER PROJECT

http://homerproject.eu

HOMER [10] is a project focused on making Public Sector Information (PSI) in the Mediterranean space easily available and reusable, as encouraged by open data trends. It aims to increase governmental transparency and create economic value by setting the foundation for a competitive and sustainable transnational digital market.
Firstly centred on five strategic sectors (Agriculture, Tourism, Environment, Energy, and Culture), Homer will provide the necessary help with the adoption of open data policies in several countries (Spain, Italy, France, Malta, Greece, Slovenia, Cyprus and Montenegro).

In truth, existing cultural and legal barriers threatens Homer institutional partners and the desired outcome of the project. Hence, in a first phase, required action and strategies were established in order to accomplish the main objectives and effectively tackle the identified challenges. The aggregation of initially isolated and fragmented open data initiatives played an important part in this stage. It enhanced cooperation and facilitated the sharing of experiences, guidelines, and good practices.
Due to the importance of these partnerships, the Homer Federation [11] pursues to grow in numbers, appealing to other independent open data initiatives to join in. Together, each with its complementary skills and knowledge, the projects' potential is maximised and the data catalogue is expanded.

In order to enter the federation, the candidate stakeholder must meet the partnership requirements listed in the "Memorandum of Understanding". If so, the entity may fill the corresponding application form [11] and wait for the e-mail response from the Homer project staff. After a positive reply, the stakeholder must sign the concerning Memorandum to finalize the federation process.

As of today, Homer has democratised hundreds of inaccessible datasets to the average citizen, within its federation. This motivated the development of several innovative applications, as well as the materialisation of ideas and scenarios, in organised hackathon events [12].
The HOMER project lives on, as described, raising awareness and promoting inter-operable solutions built upon open data.

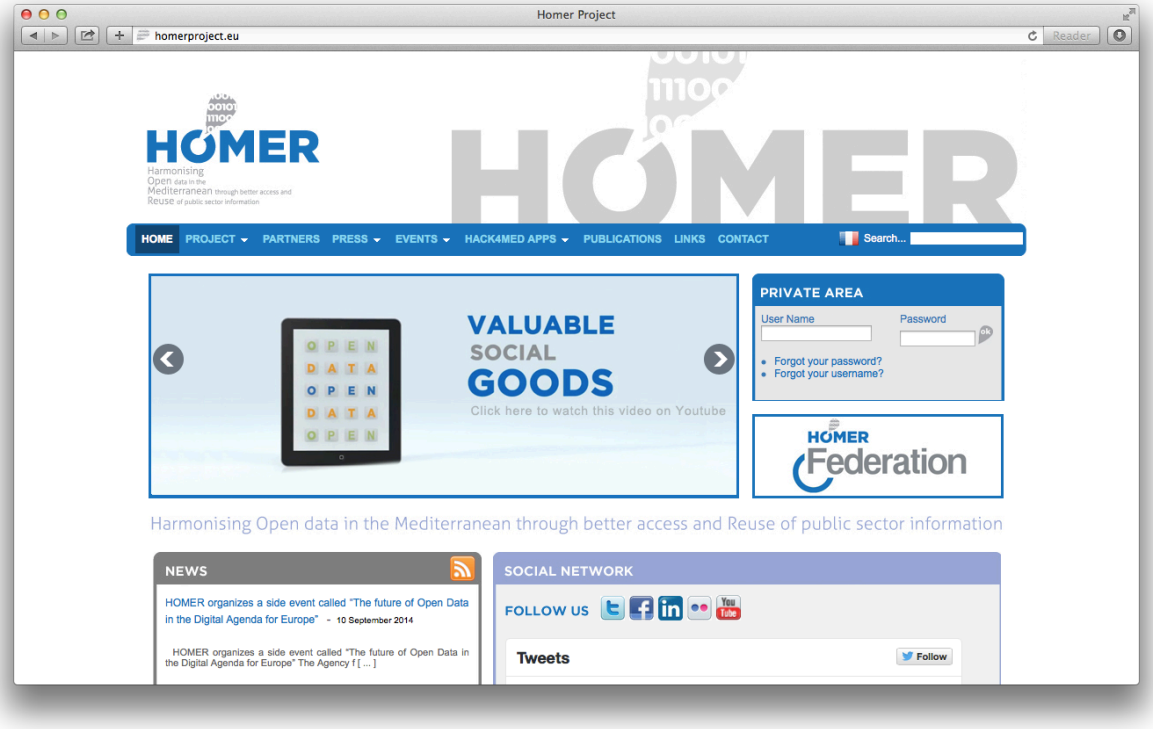


FIGURE 12 – HOMERPROJECT.EU

EUROPEAN UNION OPEN DATA PORTAL

http://open-data.europa.eu/en/data/

The European Union Data Portal is an aggregation portal of open data from «*institutions and other bodies of the European Union*». It aims to promote data reuse for commercial initiatives as well as to «help foster the transparency and the accountability» of institutions providing the data. The portal is currently maintained by the Publications Office of the European Union [13].

Apart from the open data structural projects themselves, there has also been a great effort in engaging citizens and the developer communities in order to use available data. This has mostly taken the form of application challenges and hackathons that award the best ideas. In this context, the most prominent initiative is Apps for Europe [14] that organizes competitions on a yearly basis and helps promoting the projects in complement to the financial prize.

Another good example is the Open Data Institute (ODI) [15], based in London and founded by Sir Tim Berners-Lee. As part of its mission, the ODI coordinates a startup incubation program focused on open data based ventures, counting 17 companies at the time of writing [16].

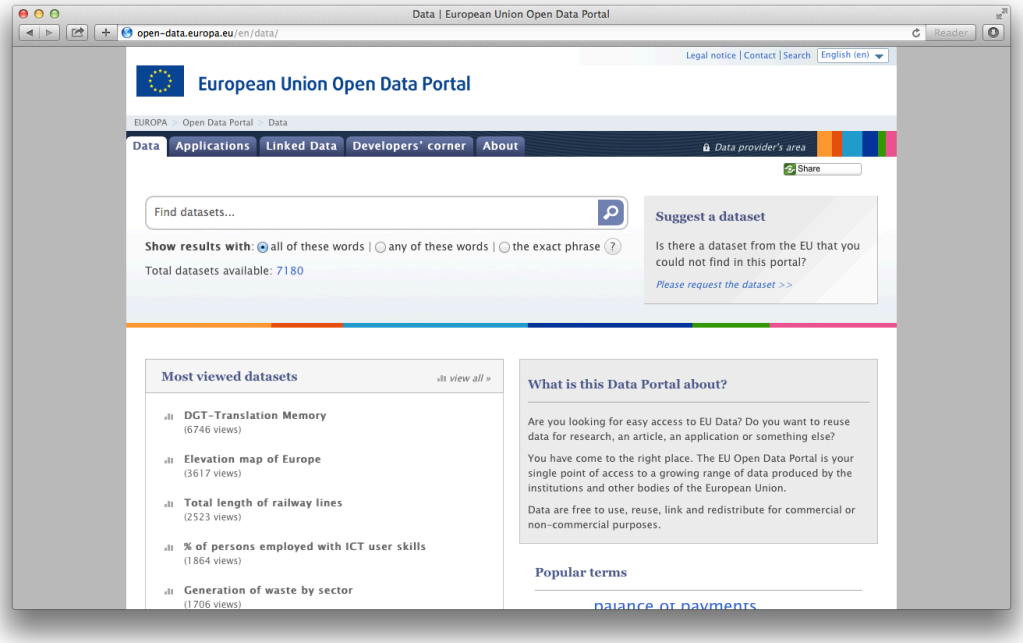


FIGURE 13 – OPEN-DATA.EUROPA.EU

2.7. ORGANIZATIONS PROMOTING OPEN DATA

The open data movement comprises many prominent organizations around the world that strive to make open data a reality. This includes lobbying, speaking in events about the benefits of open data as well as developing tools to make it easy and accessible, among other related activities. Some of the most renowned organizations promoting open data are listed next.

THE OPEN KNOWLEDGE FOUNDATION

https://okfn.org

The Open Knowledge Foundation (OKFN) «*is a nonprofit organization that promotes open knowledge, including open content and open data*» [17]. It was founded in 2004 by Rufus Pollack and it has been promoting dissemination activities ever since. These activities include participating and public speaking in events about open data, catalyzing open data projects, event hosting such as hackathons and conferences, developing tools (CKAN being the most famous) and providing infrastructure for independent projects. In parallel, the OKFN coordinates several mailing lists and IRC channels for collaboration and general networking.

THE OPEN DATA INSTITUTE

http://theodi.org

The Open Data Institute (ODI) is a non-profit London-based organization founded by Sir Tim Berners-Lee and Nigel Shadbolt in 2012, aiming to promote a culture of open data as an enabler of «*economic, environmental, and social value*». As part of its strategy of «spreading and developing the principles of open data technology», the ODI coordinates several training programs that comprise basic topics such as open data concepts and data quality, as well as advanced topics such as open data technology and licensing. The target audience of these courses consists of «*open data practitioners, information managers and architects, ICT suppliers, knowledge managers, statisticians and analysts, technical strategists, entrepreneurs, business developers, developers, research and intelligence*».

The Open Data Institute also coordinates a startup incubation program for early-stage ventures that create businesses out of open data.

THE OPEN INSTITUTE

http://openinstitute.com

The Open Institute (OI) is an organization founded in 2012 in Nairobi, Kenya, that is mainly active in Africa. Its mission is to empower and engage citizens to participate and take an active role in government and society, in order to induce bigger and faster change. The activities carried out by the OI focus on three different strategic areas:

- Open Government interventions for better outcomes;
- Local implementation of international initiatives;
- Multi-stakeholder engagement in governance.

The OI works with civil society organizations to promote civic engagement so that a more responsive and inclusive form of government emerges. In this context, open data is seen as an actionable good and an enabler of tools that help this vision to be implemented. Apart from developing tools, the OI works on knowledge creation and policymaking, in a global effort to foster social innovation and an open society.

THE OPEN DATA FOUNDATION

http://www.opendatafoundation.org

The Open Data Foundation (ODaF) is a non-profit organization «*dedicated to the adoption of global metadata standards and the development of open-source solutions*» that help promotes and understands statistical data. It aims to bring together different communities of open data practitioners, researchers, and ICT agents, among other industry players, in order to develop technological standards and tools that may shape the way open data is discovered, accessed, and distributed. This is regarded as a means to provide better decision-making, supported by higher-quality data and metadata.

You may find a more comprehensive list of organizations promoting open data in [18].

2.8. LICENSING ISSUES

THE IMPORTANCE OF LICENSING

No data should ever be published without a licence. Licensing data is of vital importance in communicating to potential users how exactly it may be reused to avoid legal problems down the road. Moreover, it encourages the reuse itself by laying out the terms clearly and creating visibility within the community of users and reusers. It is also important to notice that the default legal position in the absence of an explicit licence is that nothing can be done without first contacting the author (or rights owner) every time it is meant for reuse.

In its "Open Data Handbook" [19], the Open Knowledge Foundation addresses this need by stating that «*even in places where the existence of rights is uncertain, it is important to apply a license simply for the sake of clarity. Thus, if you are planning to make your data available, you should put a license on it – and if you want your data to be open, this is even more important.*»

In what comes to open data in particular, one must take the open definition into consideration: «*a piece of data or content is open if anyone is free to use, reuse and distribute it – subject only, at most, to the requirement to attribute and/or share-alike*». This means that only a few licences described ahead qualify for open data as described by the Open Definition. Licensing open data that is compliant to the OD is the subject of the first star in the five star open data model [4] suggested by Sir Tim Berners-Lee.

The PSI directives do not impose any particular licence but do include some recommendations. The 2013/37/EU specifically states the following:

«*Any licences […] should […] place as few restrictions on reuse as possible […]. Open licences available online, which grant wider reuse rights without technological, financial or geographical limitations and relying on open data formats, should play an important role in this respect. Therefore, Member States should encourage the use of open government licences […].*»

Recital (26)

And also:

«*Public sector bodies may allow reuse without conditions or may impose conditions, such as indication of source, where appropriate through a licence. These conditions shall not unnecessarily restrict possibilities for reuse and shall not be used to restrict competition*».

Article 8, paragraph 1

All in all, the directive pushes towards having transparent pre-established conditions that are applicable to everyone, preferably open licences that should be made explicit and associated with the available data.

AVAILABLE LICENCES AND GOOD PRACTICES

In this sub-section we shall look into what licences are available for publishing data and how to choose one. In this regard, there are a few different options available that should be carefully analyzed when licensing. Some criteria for analysis include:

- Some data(sets) may be required to be openly available;
- Some data(sets) may be subject to restrictions (e.g. privacy, national security, third party rights);
- Some data(sets) may be available for reuse but not for modification (e.g. legal texts, public budgets);
- Some data(sets) may be published allowing derivations with attribution of authoritative source (e.g. legal commentary, translations).

In the end, the process of picking a licence is a thoughtful one as there is no magic formula or one size that fits all. For this very reason, one may feel inclined towards writing a new one that fully adapts to the situation at hand. However valid this option may be, one should always consider using standard licences to help avoiding redundancy and other mistakes at the very least. In "Licensing Open Data: A Practical Guide" [20], Naomi Korn and Professor Charles Oppenheim address the bespoke licensing versus standard licences dilemma:

«*it is often more beneficial to use standard licences rather than bespoke ones. Apart from the benefits of enhanced organizational efficiency and cost saving, the use of standard licensing terms can lead to greater interoperability of data as well as increased user awareness of the licence terms, thereby enabling better compliance.*»

Open Data Licences, page 4

Next we're going to describe the most popular standard licences, each in its own sub-section:

- Creative Commons;
- CC-Zero;
- Public domain.

For each of the above we present it's purpose, use cases, and some considerations when using them.

CREATIVE COMMONS

Creative Commons (CC) is a «*nonprofit organization that enables the sharing and use of creativity and knowledge through free legal tools*» [21]. CC also develops Creative Common copyright licences, that are «*fast becoming one of the most used and recognized standard licenses for providing access to data and other resources*».

Simply put, CC licences «*permit the free of charge copying, reuse, distribution and, in some cases, the modification of the initial creator's creative work, without having to obtain permission every single time from the rights holder*». This happens because the rights holder grants this permission beforehand through the licence that classifies the data. CC licences offer a series of "baseline rights" that may be picked according to the rights holder preferences and conditions, and range from the default "all rights reserved" to "some rights reserved". It is important to notice that CC licences «*are not an alternative to copyright. They work alongside copyright and enable you to modify your copyright terms to best suit your needs*», as stated in the Creative Commons website [21].

Creative Commons licences are based on four basic clauses (building blocks), which the licensor can pick individually and match:

Attribution (BY)
Short description: «*You must attribute the work in the manner specified by the author or licensor (but not in any way that suggests that they endorse you or your use of the work).*» This clause is a feature in every licence. It states that every time we use the work we must clearly indicate who the author is.

Non Commercial (NC)
Short description: «*You may not use this work for commercial purposes.*» This means that if we distribute copies of the work, we cannot do it in any way that is primarily intended for or directed toward commercial advantage or private monetary compensation. To do this, we have to ask the licensor for specific permission.

No Derivatives (ND)
Short description: «*You may not alter, transform, or build upon this work.»* If we want to modify, to correct, to translate, or to remix the work, we have to ask the licensor for specific permission.

Share Alike (SA)
Short description: «*If you alter, transform, or build upon this work, you may distribute the resulting work only under the same or similar license to this one.*» This clause – as in the Free Software model – grants that the "freedoms" conceded by the author will be also kept on the derivative works and on the derivative ones of the derivative ones, with a persistent effect.

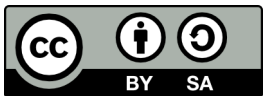
These four building blocks may be combined into 6 different licences, described next. However, it is important to observe that:

- Every licence includes the "Attribution" clause;
- A licence cannot include both the "No Derivatives" and the "Share Alike" clauses, because they are incompatible with one another. While the former denies the possibility of modifying the work, the latter implies that one holds the permission to do so.

Every one of these six licences is thoroughly described on the Creative Commons website, but we'll provide a summarized description for better understanding:

Attribution (BY)
Others can distribute, remix, tweak, and build upon your work, even commercially, as long as they credit you for the original creation.

Licence deed: http://creativecommons.org/licenses/by/4.0

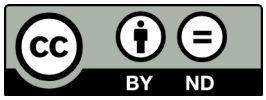


Attribution -Share Alike (BY-SA)
Others can remix, tweak, and build upon your work even for commercial purposes, as long as they credit you and licence their new creations under the identical terms.

Licence deed: http://creativecommons.org/licenses/by-sa/4.0

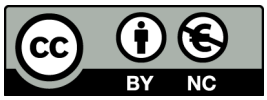
Attributions - Non Commercial - Share Alike (BY-NC-SA)
Others can remix, tweak, and build upon your work non-commercially, as long as they credit you and licence their new creations under the identical terms.

Licence deed: http://creativecommons.org/licenses/by-nc-sa/4.0



Attribution - No Derivatives (BY-ND)
Allows for redistribution, commercial and non-commercial, as long as it is passed along unchanged and in whole, with credit to you.

Licence deed: http://creativecommons.org/licenses/by-nd/4.0



Attribution - Non Commercial (BY-NC)
Others can remix, tweak, and build upon your work non-commercially, and although their new works must also acknowledge you and be non-commercial, they don't have to license their derivative works on the same terms.

Licence deed: http://creativecommons.org/licenses/by-nc/4.0

Attribution - Non Commercial - No Derivatives (BY-NC-ND)
Only allows others to download your works and share them with others as long as they credit you, but they can't change them in any way or use them commercially.

Licence deed: http://creativecommons.org/licenses/by-nc-nd/4.0

The following table makes it clear how the different licences stack up against each other, making for a good cheat sheet to carry around:

	Can someone use it commercially?	Can someone create new versions of it?
Attribution		
Share Alike		Yup, AND they must license the new work under a Share Alike license.
No Derivatives		
Non-Commercial		Yup, AND the new work must be non-commercial, but it can be under any non-commercial license.
Non-Commercial Share Alike		Yup, AND they must license the new work under a Non-Commercial Share Alike license.
Non-Commercial No Derivatives		

Apart from the descriptions and licence deeds available on the CC website, there is an extra tool which allows picking the right licence in a wizard-like manner. On that wizard [22] the licensor is able to answer a few questions and get the appropriate CC licence, which can be further embedded on another website.

CC-ZERO

CC-Zero is a tool (thus not a licence) developed by Creative Commons that allows work authors to licence their work as public domain «*through a waiver of all copyright to the extent permitted by law*». This is contrary to other CC licences that allow rights holders to grant permissions while retaining their copyright. CC-Zero allows rights holders to renounce the copyright and establishes a "no rights reserved" policy regarding the work.

This also means that once an owner applies CC0 to a work, it no longer belongs to him/her in legal terms. It then becomes truly public domain, and free to use and distribute «*in any way and for any purpose, including commercial purposes*».

According to Creative Commons, «*CC0 enables scientists, educators, artists and other creators and owners of copyright-protected content to waive copyright interests in their works and thereby place them as completely as possible in the public domain, in order for others to freely build on, enhance and reuse the works for any purposes without restriction under copyright*».
This allows us to summarize CC0 as follows:

CC-Zero
«*CC0 is a tool created by Creative Commons to facilitate the release of content, data, datasets and databases into the public domain [...] Where this is not possible, a CC0 licence provides the means for the rights holder [...] to provide instead an irrevocable, royalty-free and unconditional licence for anyone to use the resource, etc., for any purpose.*» [20]

Licence deed: http://creativecommons.org/publicdomain/zero/1.0/

PUBLIC DOMAIN

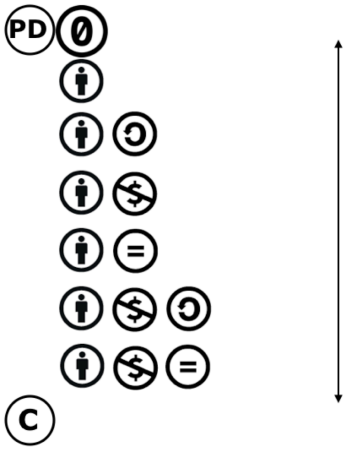
Apart from the standard licences, Creative Commons defines an extra licence for works that are free of known copyright around the world. According to the CC website, *«the Public Domain Mark enables works that are no longer restricted by copyright to be marked as such in a standard and simple way, making them easily discoverable and available to others. [...] The Public Domain Mark operates as a tag or a label, allowing institutions like those and others with such knowledge to communicate that a work is no longer restricted by copyright and can be freely used by others».*

Public Domain
Our Public Domain Mark enables works that are no longer restricted by copyright to be marked as such in a standard and simple way, making them easily discoverable and available to others.

Licence deed: http://creativecommons.org/about/pdm

CONCLUSION

Using CC licences has many advantages, including their widespread adoption and flexibility. The different licences cover a wide range of use cases, for both, those looking to retain copyright and those who don't. Regarding the ones presented earlier, here's how they rank in the spectrum of "copyrightness":



There are, however, some considerations when using CC licences. For instance, they do not cover third-party rights, just those of the owner. They also do not guarantee to provide any info about the content itself, which leaves the risk to the licensee. They are irrevocable and provide no means to go back and undo licensing. Creative Commons licensing may also cause some interoperability issues, due to the fact that some licences are not compatible, as well attribution stacking, if licences are poorly chosen.

These issues may be avoided by following a set of good practices. The OpenData Support initiative [23] makes the following recommendations on how to license open data [24]:

- For metadata use CC0. Most copyright regimes do not give protection to factual data, CC0 ensures that no party can claim copyright on metadata;
- For content choose between CC0, PDM, CC BY or CC BY-SA;
- If the original data is in the public domain (e.g. by law), the recommendation is to keep it there – use, for example, the Creative Commons Zero Public Domain Dedication or the Open Data Commons Public Domain Dedication and License (PDDL);
- For some documentation where integrity needs to be protected, use a No-Derivatives licence, for example "Creative Commons Attribution- No Derivatives", but only if really necessary;
- Avoid Non-Commercial licences if at all possible, as these seriously restrict reuse.

All in all, it all comes down to promoting unrestricting licences, protecting against liability, avoid issues such as attribution stacking and gaining visibility within the community of users.

Last but not least, CC licences are just among the most popular options available today to data publishers, but there are others. We recommend taking a look at Open Data Commons [25] and The Open Government License [26] for alternative ways to license your data.

2.9. HOW TO OPEN EXISTING DATA

The process of opening up data should be iterative and seen as a means to an end, not the end itself. Taking this perspective, the Open Knowledge Foundation has put together a handful of advice in this regard in its Open Data Handbook [19], which have been included and summarized in this section.
The iterative approach should stand on the "release early, release often" principle, which is very common in the software development industry. It consists of starting simple so that the first iteration is completed fast and results may be presented early on. To keep momentum going, one should plan short iterations, slowly adding complexity and tackling obstacles on the way. This helps to keep the "game running" and will incentivize quicker and more regular feedback from data consumers, which will, in turn, provide new insights and ideas on how to act and refine the process from there onwards.
Regarding the process itself, the steps of publishing data will be briefly explained in the following sub-sections.

CHOOSE DATASETS TO PUBLISH

The first step consists of picking the datasets to make available. This should be a result of short process of identification of possible datasets that can be made public and a quick assessment of the benefit of reuse / cost to publish ratio for each. Remember, one should start simple and add complexity over time.

APPLY THE APPROPRIATE LICENCE

This step is often neglected but is of extreme importance when releasing data. The licence determines how it can be used by third-party agents, such as users, companies or developers, and accounts for the only legal defence against potential data misuse. It is recommended to involve the legal department in the process, who should be able to understand all the rights and obligations that each licence states and therefore advise on which better suits the situation.
This is even more important in scenarios where the data is used for commercial activities. Having companies doing business that rely on your data and infrastructure is a sensitive situation that should be handled both in the technical field as well as the legal one. Update policy, data quality standard, and service level agreements are some of the topics that should be looked at thoughtfully when designing the appropriate licence.

The licence to apply should always be compliant with the Open Definition [2].

MAKE DATA AVAILABLE

Picking the right datasets and the correspondent appropriate licences are the first steps towards publishing data. Having done both, it becomes necessary to make the technical arrangements that allow data to be released. This includes preparing the data so that it complies with open data quality criteria (such as the format being used) as well as the infrastructure that supports it. In this context, a few general rules of thumb apply:

- Data should be available online and «*priced at no more than a reasonable cost of reproduction*» [27], preferably free via download or via an API;
- Data should be available in bulk and self-contained;
- Data should be published in a machine-readable format that is not subject to patents nor requires commercial software to be opened.

The method for publishing strongly depends on what type of data is at hand, as well as its potential relevance for re-use. For instance, common citizens are accustomed to downloading data such as documents and lists. Developers, however, tend to prefer APIs that enable machine-to-machine seamless communication.
This will also depend on what mechanism will be used for data publishing, what features it has, and how versatile it is. The main options may be summarized as follows:

- Use an existing website;
- Use a specialized data publication tool (such as CKAN);
- Use third-party online services (Github, Socrata, etc);
- Use an FTP server;
- Use torrents;
- Use an API.

Of these, the most developer-friendly options are using an API or an FTP server, because data fetching can then be made totally automatic. Torrents are usually harder to get, due to a variety of reasons (network or firewall restrictions, QoS, etc.) but valid nonetheless, especially for large datasets. Using an existing website or a third-party service is a complex decision to make, which we cover in detail in section 2.10.

MAKE DATA DISCOVERABLE

The last step consists of making sure that the published data reaches its intended audience and that feedback is received and processed. Consumer engagement is the step that makes all the three previous steps worthwhile and the motivation for future iterations as well.

Making data discoverable can be done in different ways. For instance, one may choose to publish data in more than just one platform, to ensure it reaches a far wider audience and therefore creating a higher chance data will be useful for building tools.

Using some of the existing online platforms that allow publishing of data is a good way of getting users aware of the released data, especially popular ones such as Github (github.com), Morph (morph.io), Datahub (datahub.io), among others.
Regardless of the tool being used, there are a few factors that should be considered in what comes to making your data discoverable and easy to use:

- Allow and promote data publishing from other publishers into the same platform, as a means to enrich it with other complementary data;
- Avoid registration or access restriction policies;
- Allow feedback mechanisms from data consumers, such as dataset quality rating.

All in all, the easier it is for consumers to browse, download and use data, the more discoverable it is likely to become.

At this point, another recommendation can be added to reinforce the iterative approach mentioned earlier: the assessment period. It is of utmost importance to collect feedback from users and monitor the results of publishing data, and use this information as an input to future iterations, so that there is a continuous refinement of the process and, consequently, better results are achieved. The following questions should be asked:

- Was data useful, how, and to whom?
- Was the quality good enough? What is it lacking?
- Were there any identified constraints or limitations on fetching data?
- How frequently was data downloaded / accessed?
- How much feedback was given, and how was user support rated?

Besides the data publishing itself, one should also focus on assessing the ecosystem that is being created as a whole. It should comprise the staff responsible for creating, updating, and providing support for data,

developers (often called info-mediaries) and users (those who use the tools based on data). Open data has the potential to create and foster a win-win-win situation for all, which should be subject to analysis:

- Are the users getting new tools?
- Are the developers getting the data that they need?
- Are the providers getting some kind of benefit from it? (e.g. happier users)

For a more comprehensive guide on how to publish open data, check the Open Data Handbook by the Open Knowledge Foundation [28].

2.10. WHERE TO PUBLISH OPENED DATA

When facing the need to publish data, one must look for alternative methods for hosting the data as well as the underlying platform where uploads and updates are made. On an abstract level, there are three available choices:

- Use an online service;
- Install and configure an existing platform;
- Developing a new platform from scratch.

The implications of choosing one of the above intersect both the strategic and the technical perspectives. Using an online service such as Github (github.com) or Socrata (socrata.com) implies that the publisher has no control over the infrastructure where data is being held. This also means compliance with the terms and conditions stated by either service regarding its usage, security, etc. All in all, one must consider the pros and cons of relying on an external service, just as one would if it were source code, documents or any other digital good.

Examples of online data hosting platforms include:

- Github.com
- Datahub.io (hosted by Open Knowledge)
- PublicData.eu (hosted by the European Commission)
- Socrata.com

The most obvious alternative to the previous scenario is to pick a (preferably open source) data platform, such as CKAN or DKAN, and host it within one's IT infrastructure. This option gives one the control over every step of the process of publishing data but it doesn't come without drawbacks. It implies a set of technical tasks that one must be willing and able to carry out, such as updates to the platform, ensuring quality of service and service level agreements (if applicable), caring for security and overall IT maintenance. These tasks may be outsourced to specialized and certified IT companies to save money, but, in general, it will cost a lot more compared to using existing online services.

Examples of open source self-hosting data platforms:

- CKAN (http://www.ckan.org)
- DKAN (http://nucivic.com/dkan/)
- The Data Tank (http://thedatatank.com)
- Dataverse (http://datascience.iq.harvard.edu/dataverse)

Finally, there is a third option available, which is to develop a new open data platform from scratch tailored to the specific needs at hand. This would be the most expensive of all three options because of software development costs. Taking into account the investment needed, this scenario is hardly justifiable considering the open source choices available nowadays that can be used to publish data. Plausible explanations to follow this route may include the fact that none of the available options have the necessary features for the specific type of data that one is trying to publish - for instance, real-time data – or the right settings – for instance, if there is no support for your language and localization settings. It may also be the case that none of the open source alternatives perform adequately for a specific domain, in regard to design or base architecture. Other legitimate

justifications might be technological (legacy systems integration, closed source policy, security through obscurity, etc.), legal or even political.

3. THE REUSE OF OPEN DATA AS A BUSINESS OPPORTUNITY IN EUROPE

In the following sections we will present some examples of platforms and applications that use open data for business in the EU area.

CITYMAPPER – THE ULTIMATE TRANSPORT APP

https://citymapper.com

Citymapper is a trip planning multi-platform application that aims to make daily urban commuting within the cities as easy as possible.
As one of the most popular and promising applications of its kind, the main objective of Citymapper is too make complicated metropolis simple and easy to use. It provides route planning, as well as real-time alerts and notifications regarding weather and transit disruptions.

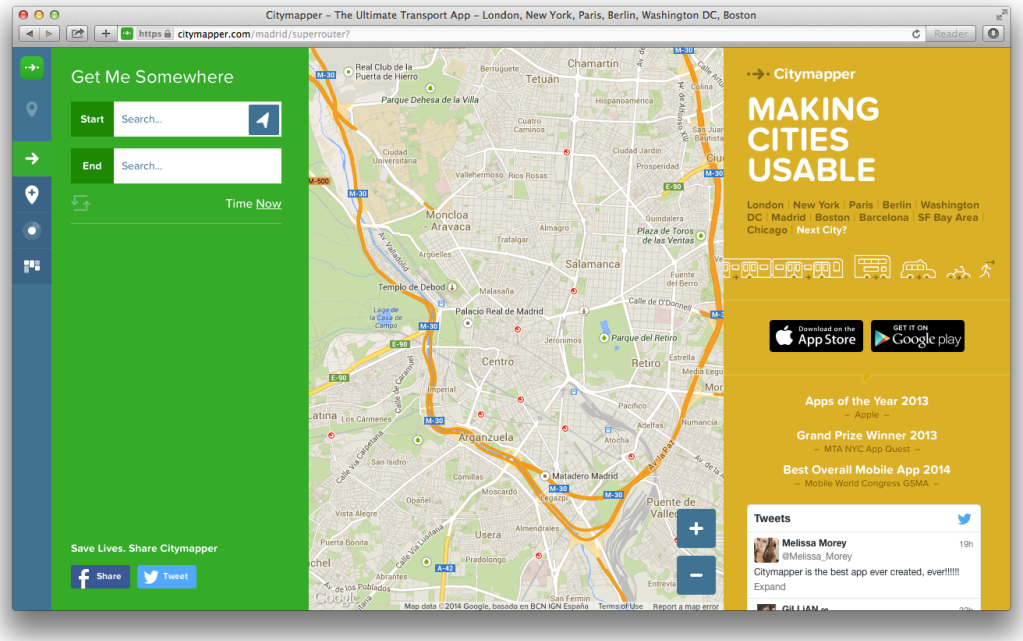


FIGURE 14 – CITYMAPPER – THE ULTIMATE TRANSPORT APP

OST – ONE.STOP.TRANSPORT

https://www.ost.pt

One.Stop.Transport platform (OST) is a software system for data aggregation, transformation, and delivery, following modern open data principles, standards and directives. It is available online and comprises three main components:

- **Users**: centralized authentication using the OAuth open standard to authorization protocol;
- **Applications**: includes an app marketplace for the end-user, integration and hosting mechanisms for the developer;
- **Data and APIs**: open data for fuelling the engine.

The type of data OST currently manages regards Transportation (mass transit), Traffic data (highways), Points of interest and Territory (mainly OSM), which are all supposed to be open data. The data APIs may be built for static or for real-time data, although data providers currently contribute with static data only. For the last two

years, OST has been working with several data providers, mainly related to public transportation, such as the following:

- SMTUC (a Coimbra-based mass transit agency)
- CP (a major Portuguese nationwide railroad agency)
- STCP (Oporto mass transit bus agency)
- MoveAveiro (Aveiro mass transit agency)
- Carris (Lisbon mass transit bus agency)
- Transdev (Portuguese nationwide intercity bus agency)

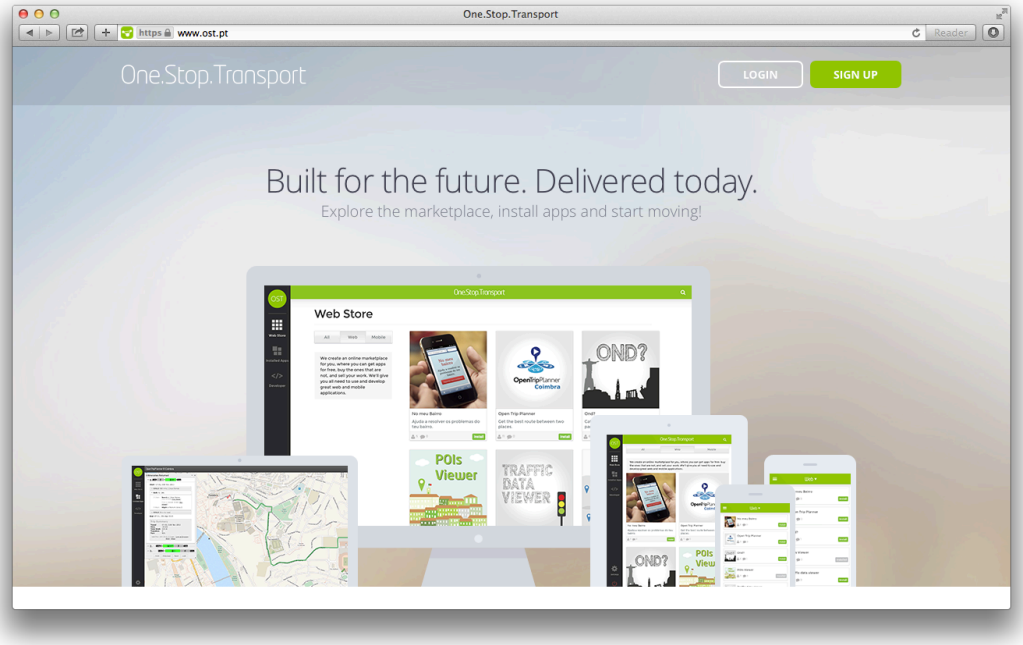


FIGURE 15 – ONE.STOP.TRANSPORT PLATFORM

MOOVIT

http://moovitapp.com/

Moovit is a mobile application that aims to give all the information necessary for the user to plan his daily commutes using public transportation.
This application allows the user to check schedules and get the fastest and most convenient route to a given destination, with step-by-step map navigation. Moreover, it empowers users with real-time information like arrival times and alerts.

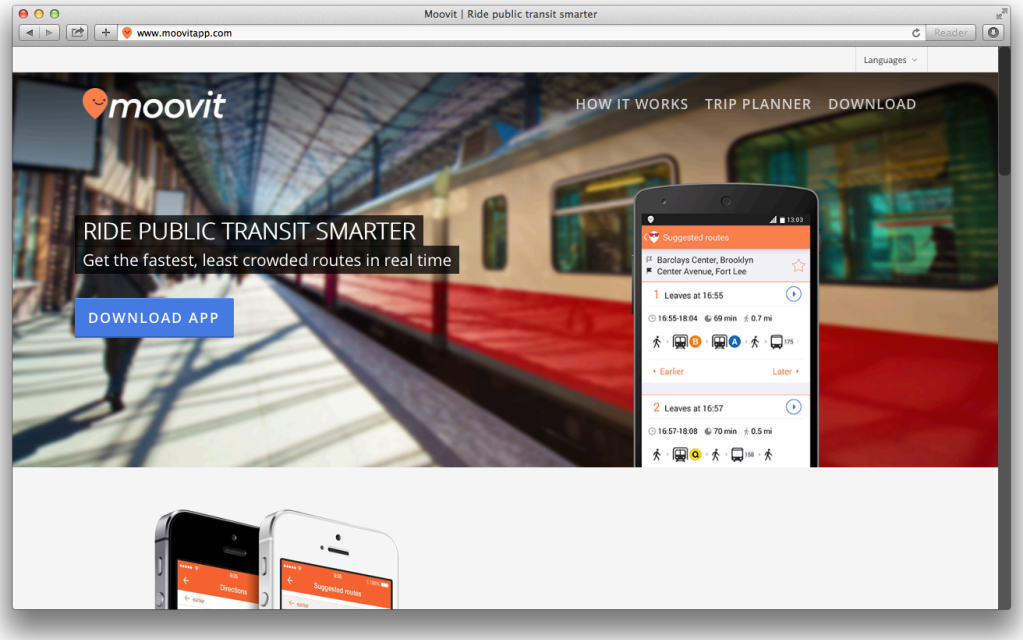


FIGURE 16 – MOOVIT APPLICATION

MAPNIFICENT

http://www.mapnificent.net

Mapnificent is an open web application that seeks to promote the use of public transportation. It intends to be an invaluable help in finding habitable places for a better daily commute.
Rather than being a complex trip planning service, Mapnificent simply shows the area you can go to from any point within a given period of time, using public transportation only. It primarily aims to be a decision-making tool to explore possible habitable locations according to the working place and desired travel time.

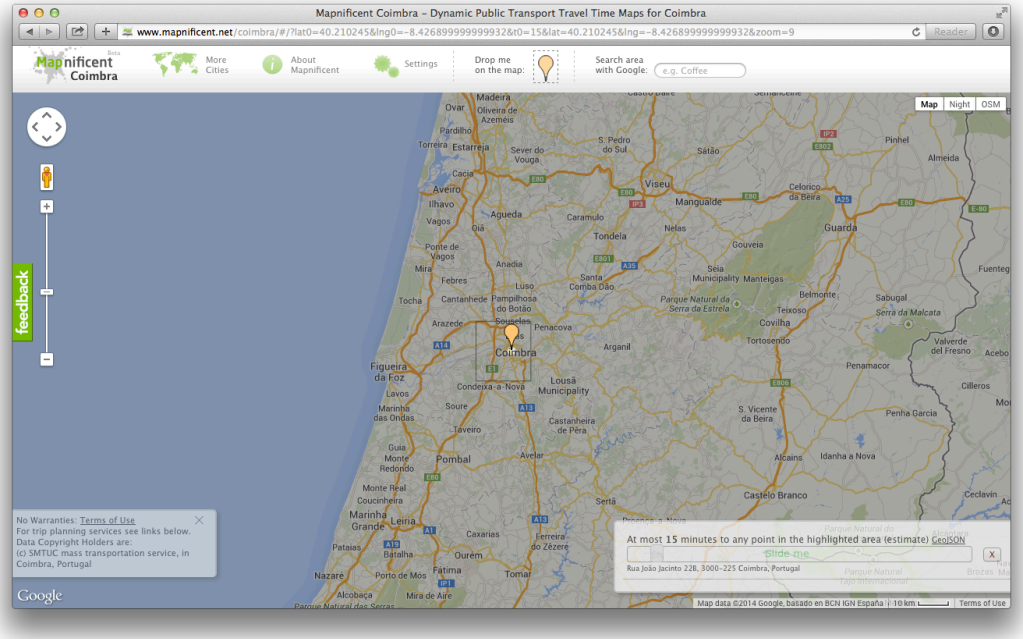


FIGURE 17 – MAPNIFICENT APPLICATION

OPENTRIPPLANNER

http://www.opentripplanner.org

OpenTripPlanner is a complex trip planning web and mobile application that uses public transit data and transportation network information.
Instead of limiting its engine to public transportation as the only way to commute, OpenTripPlanner combines bus, pedestrian, bike, and car related information, for finding itineraries and provide a more complete service to the user.

FIGURE 18 – OPENTRIPPLANNER PLATFORM

SKILLS ROUTE

http://skillsroute.com

This online application helps young students and their parents to discover which nearby schools and colleges are most suitable to them.
Skills Route is designed to identify education and career options that might be available in the future, according to current achievements and goals. Although it is still in beta version, it strives to be a life decision tool, providing valuable information, such as where the desired subjects are schooled and what are the expectations and outcomes for each option.

FIGURE 19 – SKILLS ROUTE APPLICATION

ILLUSTREETS

http://illustreets.co.uk

Illustreets helps citizens to identify the best places to live in England, filtered by various criteria, including standard of living, crime rates, house prices, and education.
This web application aims to remove the hassle out of finding the right property to rent or buy, by narrowing down the list of possible places considering to user's budget and desired location.

FIGURE 20 – ILLUSTREETS APPLICATION

CARAMBLA

http://carambla.com

Carambla is a web-based application whose purpose is to locate the nearest parking spaces available.
As a way to address the parking problem in urban areas, Carambla offers information regarding availability, opening hours and tariffs, for each geo-located parking space, enabling the commuter with all the necessary information to plan ahead and parking smart.

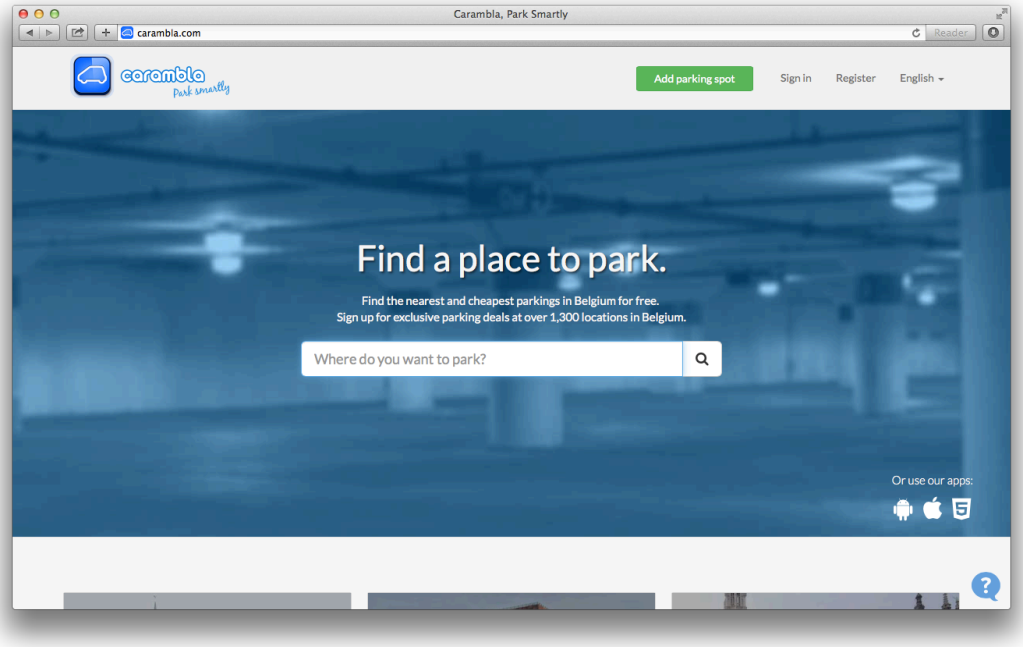


FIGURE 21 – CARAMBLA APPLICATION

TOTALCARCHECK

http://totalcarcheck.co.uk/

TotalCarCheck is an online tool that helps people check information and history of a vehicle.
This application aims to help those who are looking to buy a vehicle but don't want to fall victim to scams or bad deals, hence exposing information and history regarding the vehicle (e.g. original color, engine number, inspection, and police reports).

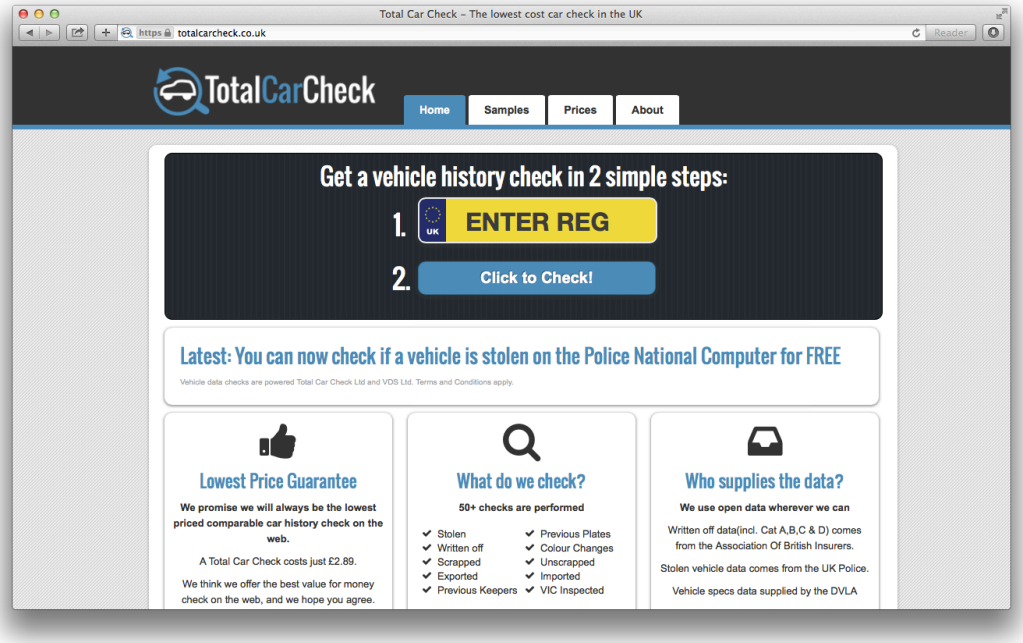


FIGURE 22 – TOTALCARCHECK APPLICATION

WHEREDOESMYMONEYGO

http://wheredoesmymoneygo.org

WhereDoesMyMoneyGo is a project that aims to entrust the common citizen with knowledge regarding public spending through interactive analysis and visualization of information. It serves as means of raising accountability and transparency towards the general public. In fact, it enables the citizen to explore into government affairs and finances, promoting participation, collaboration, and civic engagement.

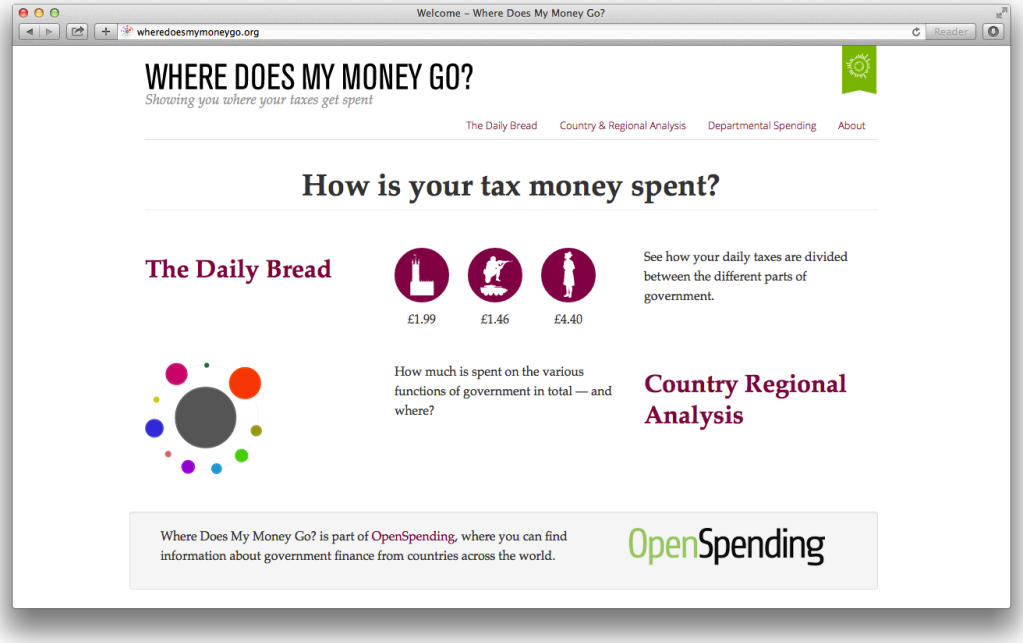


FIGURE 23 – WHEREDOESMYMONEYGO APPLICATION

BIKECITYGUIDE

http://www.bikecityguide.org/app/

BikeCityGuide is a mobile application that provides a navigation and routing system designed for the needs of cyclists in urban areas.
This application encourages people to use the bicycle as means of transportation rather than a form of recreation. It offers cycle-friendly routes so the user may enjoy cycling in urban areas, in a safer and more comfortable way.

FIGURE 24 – BIKECITYGUIDE APPLICATION

4. OPEN DATA IN CITEK PROJECT

The main objective of the CITEK project [29] is to promote the setting up of a MED transnational and trans-local system of innovation involving EU-wide national and regional governments, research centres, innovation drivers, associations, and small and medium enterprises (SMEs) – the triangle of innovation – thus capitalizing the results of ICS, IKTIMED, and R&D Industry (members of the first MED Programme cluster), and setting up synergies and a knowledge exchange with projects funded by other programs (Central Europe Program, IVC, Atlantic Area, FP7).

The main project activities are: collecting and analysing Med smart specialization strategies, developing and implementing an on-line observatory for RIS3, setting up a trans-national knowledge based innovation network platform and online tools, setting up a transnational online business community, scouting and analysing the internationalisation potentialities of the project area, and setting up a pilot open data system.

Regarding the pilot open data system, the aim is threefold: (a) to leverage the knowledge from projects being capitalized, thus showing its potential to contribute to European social and economic sustainability; (b) to disseminate the benefits of making such knowledge available as open data; (c) to take advantage of a diversity of partner nationalities to bring together people from different backgrounds to discuss the potential of open data and knowledge innovation, providing a networking experience to foster not only ideas but also further cooperative projects.

In the open data phase of the CITEK project the following activities have been pursued:

- Organization of four laboratories on open data;
- Development of a web space with datasets from selected territorial projects on innovation and enterprise support;
- Organization of a transnational event on open data;
- Development of a guidelines document on open data.

These activities are listed next, along with a short description of the work done and the intended results.

ORGANIZATION OF FOUR LABORATORIES ON OPEN DATA

The main goal of these laboratories on open data is the analysis and discussion of the open data theme. The laboratories are used to collect information from experts that can then use it to structure and develop the guidelines document to introduce project partners to the topic.
During the project implementation two open data laboratories were organized and there was a participation in Malta's event of the EU HOMER Project. The open data laboratories were organized in the following sequence:

- November, 28th in Rome, Italy: discussion and common ground about open data;
- January, 9th in Barcelona, Spain: introduction to the CKAN open data platform, practical examples on CKAN management, and a brief tutorial on how to use the tool;
- March, 11th in Silema, Malta: discussion about the opening of data from selected territorial cooperation projects and participation in the HOMER Project meeting. The complete agenda of this meeting can be found at http://homerproject.eu/en/events/calendar/icalrepeat.detail/2014/03/11/445/-/malta-meeting.

The last open data laboratory was merged with the transnational event on open data held on October 16th and 17th in Coimbra, Portugal. This event capitalizes on the information and contacts gathered during the organization of laboratories to share ideas, raise awareness of open data, and distribute the guidelines document developed in this project.

DEVELOPMENT OF A WEB SPACE WITH DATASETS FROM SELECTED TERRITORIAL PROJECTS ON INNOVATION AND ENTERPRISE SUPPORT

One of the goals of this phase is to implement a pilot open data system for making available databases, information, and outputs of CITEK and its capitalized projects, as well as other data that can be considered useful by project partners.

By the end of the project, databases from IKTIMED, R&D Industry, and ICS Business Community projects will be opened. The opening of these databases means that, instead of it being locked inside one system, they can be published in a manner accessible to the general public.

To this end, the CITEK open data web space [30] based on CKAN's platform is being set up. Datasets opened during the project can be browsed in this web space.

Section 7 of this document contains a tutorial about how to add new datasets to CITEK's open data platform. For now, this feature is only accessible to project partners of CITEK project, but those whishing to belong to this community to open some datasets can contact the consortium [31].

ORGANIZATION OF A TRANSNATIONAL EVENT ON OPEN DATA

The aim of this activity is to organize an event to share experiences and raise awareness of open data. This materialized into the event Raising AWereness of Open Data (RAW Open Data), taking place in Coimbra, Portugal on 16th and 17th of October 2014.

RAW Open Data is a joint transnational event focused on open data and knowledge as critical resources to successfully engage current social and economic challenges. It is organized by IPN – Instituto Pedro Nunes, a Portuguese Association for innovation and research and development in Science and Technology – and it will bring together around 100 people to share and discuss ideas, experiences, and projects. The first day starts by establishing some common ground on open data and a brief overview of the world's leading open-source open data portal platform. Then, the topic moves to the implications of the PSI EU directive that requires the public sector to progressively open its data. Next, there will be information about the intricacies of licensing the data from PSI. The second part of the day is dedicated to open data projects, namely efforts to unlock the full potential of the public sector information in the Mediterranean space; about the experience of the Portuguese public administration in opening its data; about how Grenoble is using open data to promote social innovation; and about how real-time open data can foster value creation and a higher quality of service to users of public transport. The first day closes with a presentation about using innovative visualizations for data. Experts from around Europe will deliver the talks. They come from Open Knowledge Foundation, W3C, LAPSI 2.0 Thematic Network, Junta de Andalucia, Agency for the Modernisation of Administrative Bodies, Scop la Péniche, and Politecnico di Milano.

In the second day of the event, a Hackathon will take place, where participants can create their own applications using real open data and knowledge from different sources.

Last but not least, during these two days, all participants can enjoy fruitful networking moments in our coffee-break areas.

RAW Open Data results from the cooperation between two complementary European Projects, CITEK [29] and Transcreativa [32], involving partners from Italy, France, Slovenia, Portugal and Spain.

More information about this event can be found in RAW Open Data Website [33].

DEVELOPMENT OF A GUIDELINES DOCUMENT ON OPEN DATA THEME

The aim of this activity is the development of guidelines about open data, to raise awareness and give an introduction about the topic. The result of this activity is the present document. It serves as a tutorial/introduction about open data for those that are new to the topic. It can also be used by those who want to open their data and need to understand how to proceed.

The guidelines were developed by a technical working group of ICT experts, by reviewing the topic and using the outputs of the open data laboratories organized in the scope of CITEK Project. The guidelines will be disseminated to all relevant stakeholders, so they can act as catalysts for new open data initiatives in their regions.

5. CITEK'S OPEN DATA PORTAL

In this section we describe how to add a new dataset to CITEK's open data platform. The following five steps will guide you in the process:

STEPS FOR PUBLISHING A DATASET

STEP 0 – YOU NEED TO REGISTER AT HTTP://CKAN.CITEK.IPN.PT AND BE A MEMBER OF ONE OF THE ORGANIZATIONS LISTED AT HTTP://CKAN.CITEK.IPN.PT/ORGANIZATION TO ADD AND EDIT DATASETS.

STEP 1 – LOG IN AT THE WEBSITE HTTP://CKAN.CITEK.IPN.PT.

STEP 2 – ACCESS THE "CREATE DATASET" SCREEN.

STEP 3 – FILL IN THE "CREATE DATASET" FORM.

WHEN YOU HAVE FILLED IN THE INFORMATION IN THIS PAGE, SELECT THE "NEXT: ADD DATA" BUTTON.

STEP 4 – FILL IN THE "ADD DATA" FORM AND ADD RESOURCES (FILES OR LINKS) TO THE DATASET.

WHEN YOU HAVE FINISHED ADDING RESOURCES, SELECT THE "NEXT: ADDITIONAL INFO" BUTTON.

STEP 5 – FILL IN THE "ADDITIONAL DATA" FORM.

WHEN YOU HAVE FILLED IN THE INFORMATION IN THIS PAGE, SELECT THE "FINISH" BUTTON.

These steps, above, are aligned with the CKAN's user guide for adding new dataset in the CKAN platform [34]. Bellow you can find more details and screen shots that illustrate the process.

STEP 0 – YOU NEED TO REGISTER AT HTTP://CKAN.CITEK.IPN.PT AND BE A MEMBER OF ONE OF THE ORGANIZATIONS LISTED AT HTTP://CKAN.CITEK.IPN.PT/ORGANIZATION TO ADD AND EDIT DATASETS.

The pre-condition to add and edit datasets in CITEK's CKAN platform is to be a member of one of the organizations presented in the page http://ckan.citek.ipn.pt/organization. The organizations that have permissions to add and edit datasets are shown in the following picture:

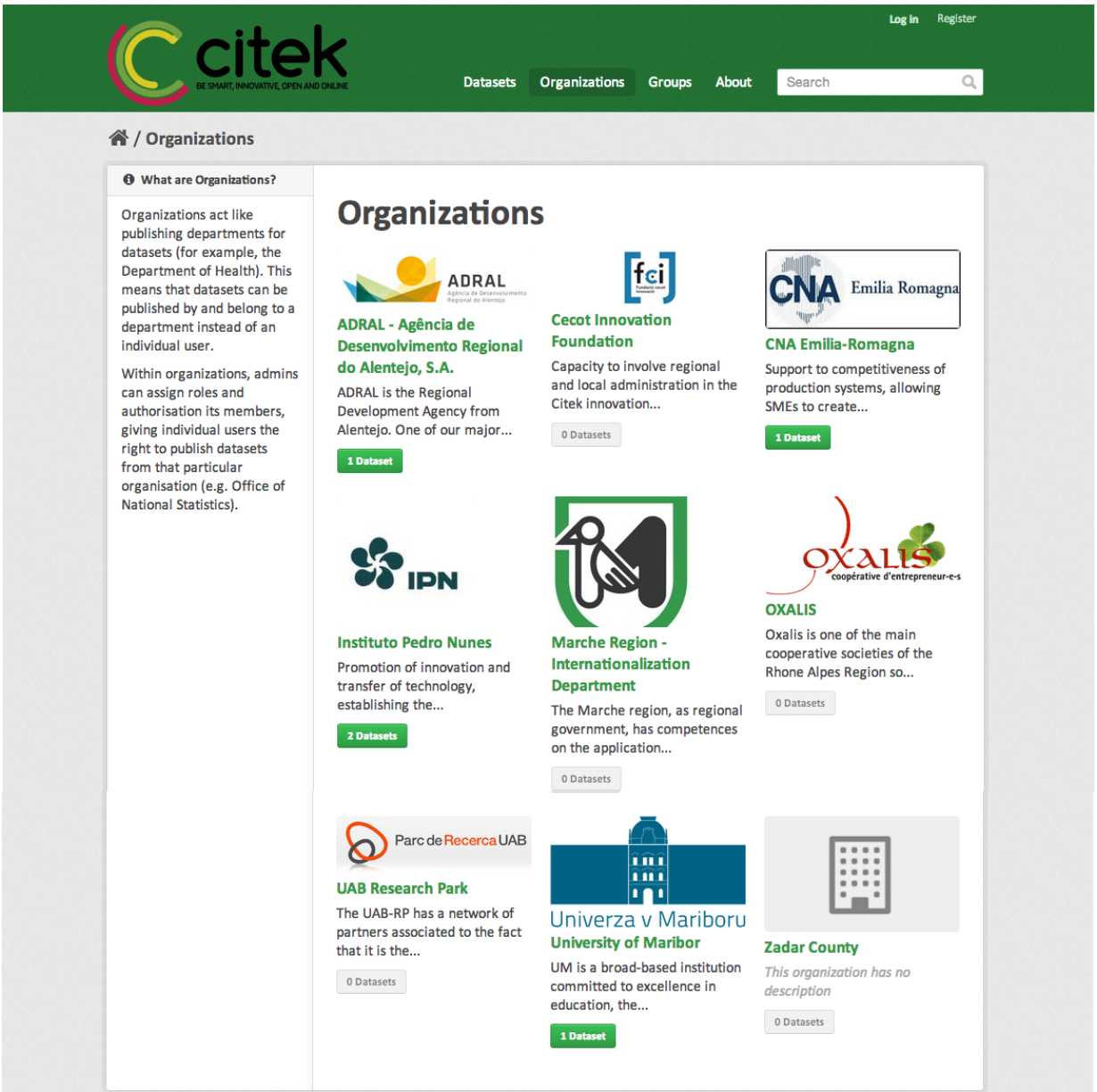


FIGURE 25 – CITEK'S OPEN DATA PLATFORM - HOMEPAGE

The organizations listed are the partners of the CITEK Project. If you don't belong to any of these organizations, you can contact us to add your organization through the contacts available at [35].

STEP 1 – LOG IN AT THE WEBSITE HTTP://CKAN.CITEK.IPN.PT

To accomplish the first step of the process you need to access the website http://ckan.citek.ipn.pt and click in the "Log In" link in the website top-right corner. The "Login" form will be presented to you. Next, you need to correctly insert your credentials (username and password) and click on the "Log in" button.

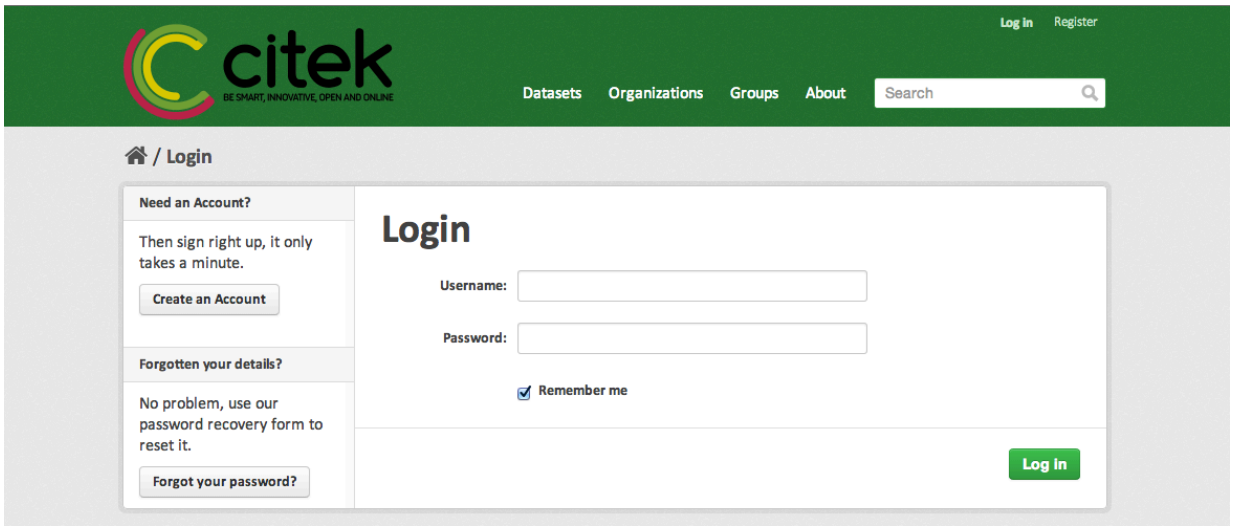


FIGURE 26 – CITEK'S OPEN DATA PLATFORM - LOGIN PAGE

You will be redirected to your "Dashboard" page where you can see the "News feed – Activity from items that you follow" screen.

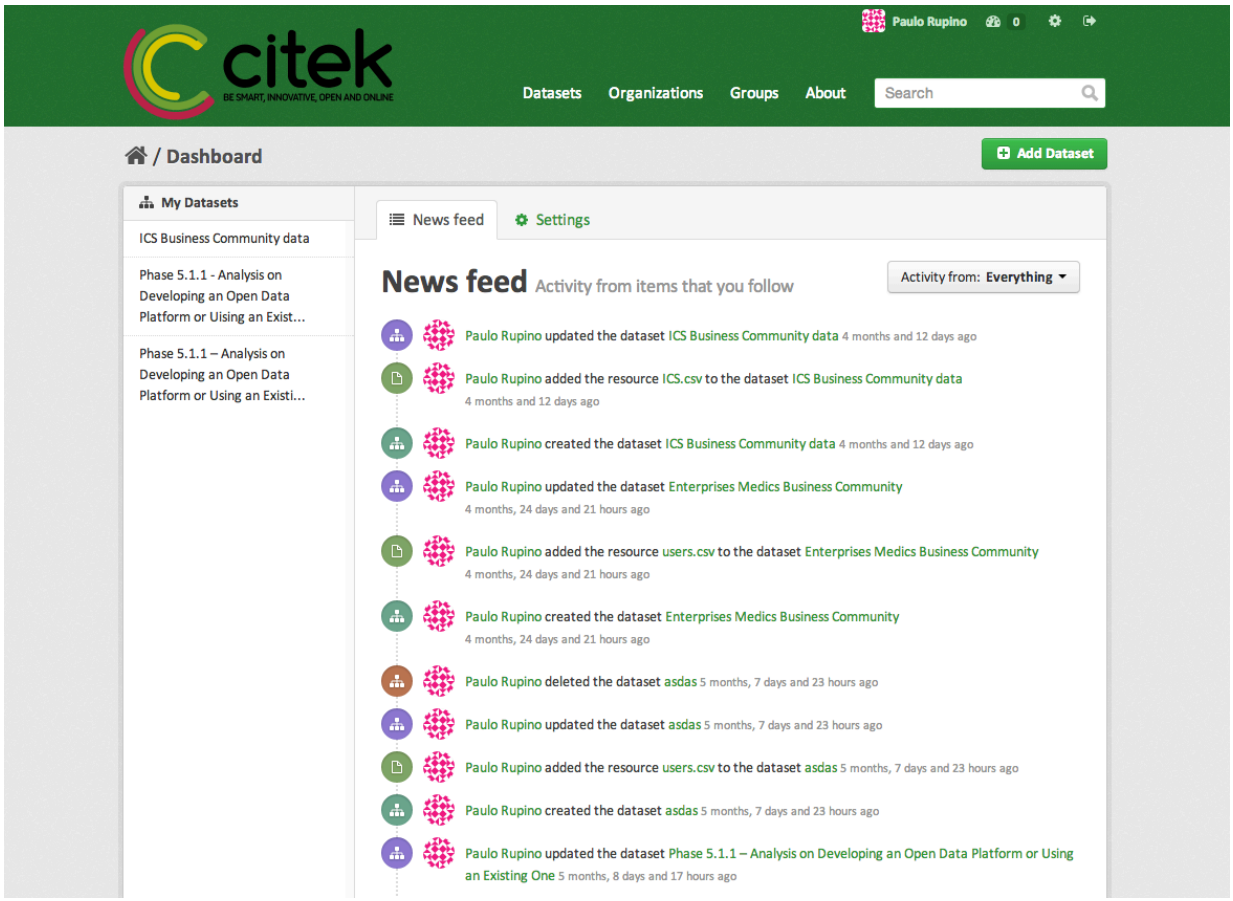


FIGURE 27 – CITEK'S OPEN DATA PLATFORM - DASHBOARD PAGE

STEP 2 – ACCESS THE "CREATE DATASET" SCREEN

Assuming that you are already logged in, you have to access the "Create dataset" screen. You can access it in three ways:

a) In your "Dashboard" page click in the green button "Add Dataset" below the "Search" box;
b) Click in the "Datasets" link at the top main menu. The "Datasets" link is accessible in any page of the platform. Click in the green button "Add Dataset";
c) Alternatively, click in the "Organizations" link in the top main menu. Select the page of your organization that should contain your new dataset. If you are a member with permissions to add a new dataset, you can now click in the white button "Add dataset" below the "Search" box.

STEP 3 – FILL IN THE "CREATE DATASET" FORM

At this phase of the process you start to fill in information about the data you want to make open. The platform will ask you for the following information:

- **Title –** insert a brief but specific title for the dataset. This title will be unique across the entire platform. (e.g. Portuguese population density by region);
- **Description –** insert a longer description of the dataset, including information such as where the data is from and any information that people will need to know when using the data;
- **Tags –** insert tags that will help people find the data and link it with the other related data (e.g. "population", "Centro region");
- **License –** select a licence for your data. As discussed previously in this document, it is important to include licence information so that people know how they can use the data;
- **Organization –** select your organization name. Ensure the default chosen is the correct one before you proceed.

A screen shoot of the form is shown in the following picture.

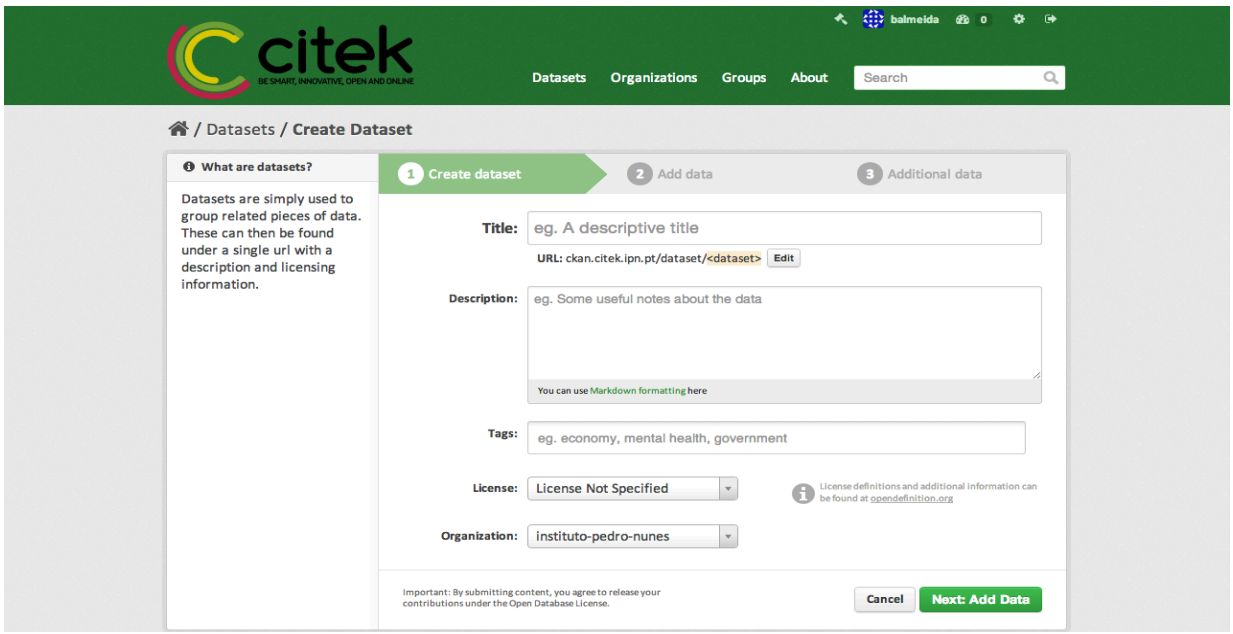


FIGURE 28 – CITEK'S OPEN DATA PLATFORM – "CREATE DATASET" FORM

In this form, the Title is the only required field, but the potential users of your data will benefit a lot if you provide the other requested information.

Please take special care in ensuring the organization you select is the correct one, since this cannot be changed later. All of the other fields can be edited later.

If you are satisfied with the information you inserted, you can click in the green button "Next: Additional info" below all the fields of the form. If not, you can leave the form by clicking in the white button "Cancel" below all the fields of the form.

STEP 4 – FILL IN THE "ADD DATA" FORM AND ADD RESOURCES (FILES OR LINKS) TO THE DATASET

In the following step of the "Create dataset" form you will be able to add one or more resources to your dataset. The platform will ask you for the following information about your resources:

- **Resource –** add a file or link for your data resource after selecting the appropriate choice at the top of the screen:
  - **Link to file** – link to a file with your data available somewhere on the Internet;
  - **Link to an API** – link to an API with your data available on the Internet;
  - **Upload file** – add a file available in your computer that contains the data. CKAN will provide you with a file browser to select it;
- **Name –** insert a name for this resource (e.g. "Population density 2011.CSV". Different resources in the dataset should have different names;
- **Description –** insert a short description of the resource;
- **Format** – select the file format of the resource (e.g. "CSV" – Comma Separated Values, XLS, JSON, PDF, etc.).

The form is illustrated in the screen shoot bellow.

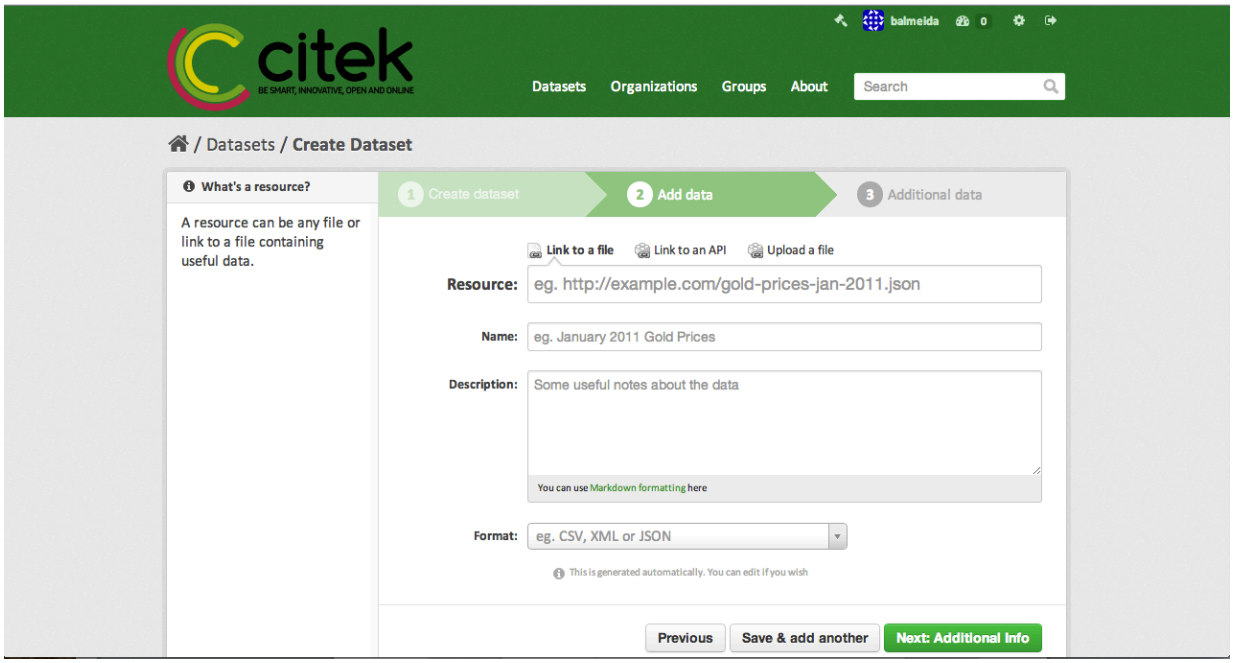


FIGURE 29 – CITEK'S OPEN DATA PLATFORM – "ADD DATA" FORM

In this form, the only required field is Resource. However, filling in the rest of the data adds value to the potential user. All fields can be edited later.

If you are satisfied with the information you inserted, you can click in the green button "Next: Additional info" below all the fields of the form. If not, you can leave the form by clicking in the white button "Previous" below all the fields of the form.

If you want to add more resources to your dataset, you can click in the white button "Save & add another". You will be able to fill in an additional form to describe the second data resource.

STEP 5 – FILL IN THE "ADDITIONAL DATA" FORM

In the last step of the "Add dataset" form you will be asked to provide additional information about your dataset, namely:

- **Visibility –** select the type of visibility of your dataset. Everyone who accesses the platform can see a "Public" dataset. A "Private" dataset can only be seen by members of the organization that was selected in the "STEP 3" of this guide. The "Private" datasets will not appear in searches by other users;
- **Author –** insert the name of the person or an organization responsible for producing the data;
- **Author e-mail –** insert an e-mail address for the author. This is the contact to which users can send questions or comments about the data;
- **Maintainer –** insert the name of the person or organization responsible for maintaining the data (may or may not be the author);
- **Maintainer e-mail –** insert the email of the maintainer of the data;
- **Custom fields –** insert any custom fields that you think will be useful to characterize your data. Here you can add a name and a value for three custom fields.

The form for "Additional data" can be seen bellow.

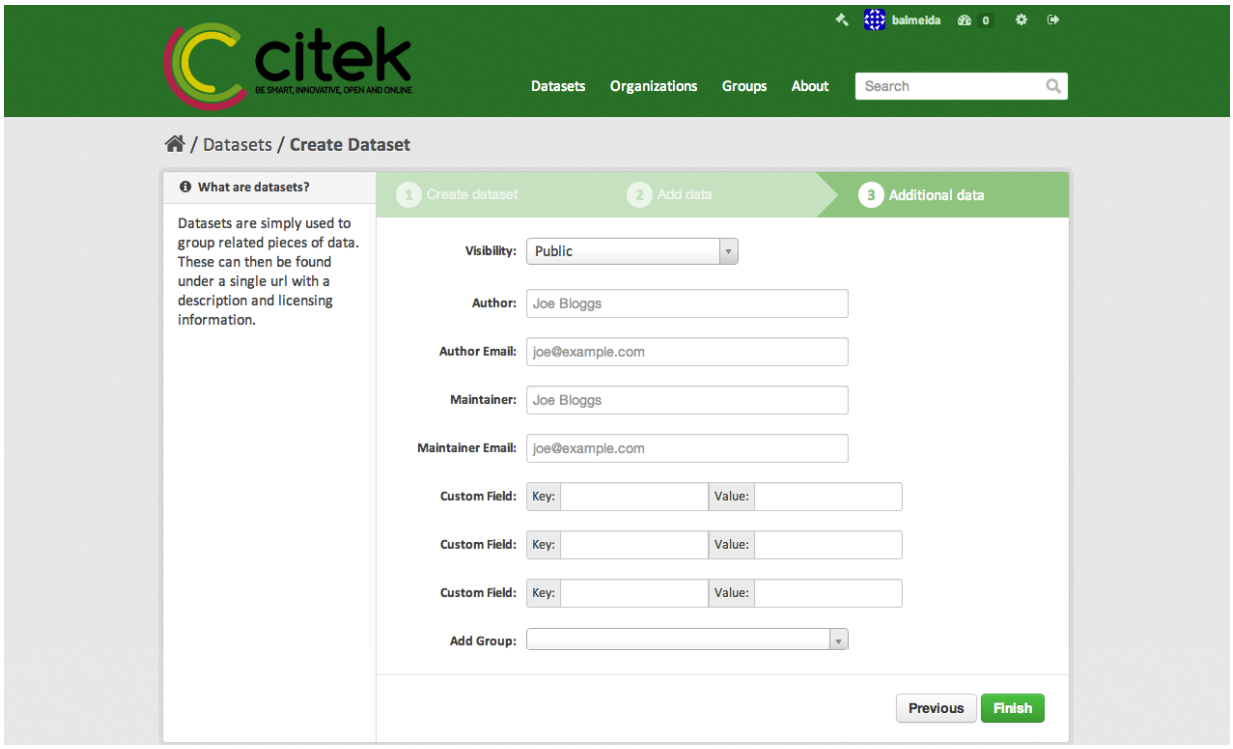


FIGURE 30 – CITEK'S OPEN DATA PLATFORM – "ADDITIONAL DATA" FORM

In this form, there aren't any required fields. However, it is important that you provide the most information that you can.

If you want to go to the previous form you can achieve that by clicking in the white button "Previous" below all the fields of the form.

If you finished filling in the fields and you want to make your dataset available for other users you should click in the green button "Finish". After that, you have finished the publication of your dataset in CITEK's open data platform.

6. REFERENCES

[1] European Comission, “Communication from the Commission to the European Parliament, The Council, The European Economic and Social Committee and the Committee of the Regions - Open data an engine for innovation, growth and transparent governance.” [Online]. Available: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2011:0882:FIN:EN:PDF. [Accessed: 06-May-2014].

[2] Open Knowledge Foundation, “Opendefinition.org - Open data definition.” [Online]. Available: http://opendefinition.org/. [Accessed: 06-May-2014].

[3] Open Knowledge Foundation, “opendefinition.org - Full Open Data definition.” [Online]. Available: http://opendefinition.org/od/. [Accessed: 20-May-2014].

[4] “5-star open data.” [Online]. Available: http://5stardata.info/. [Accessed: 15-Sep-2014].

[5] The White House - President Barack Obama, “Executive Order -- Making Open and Machine Readable the New Default for Government Information.” [Online]. Available: http://www.whitehouse.gov/the-press-office/2013/05/09/executive-order-making-open-and-machine-readable-new-default-government-. [Accessed: 15-Sep-2014].

[6] Open Knowledge Foundation, “Open Data Handbook - Why Open Data?” [Online]. Available: http://opendatahandbook.org/en/why-open-data/index.html. [Accessed: 15-Sep-2014].

[7] D. F. Kettl, *The Next Government of the United States: Why Our Institutions Fail Us and How to Fix Them*. 2008, p. 288.

[8] T. O’Reilly, “Gov 2.0: The Promise Of Innovation (Forbes.com),” 2009. [Online]. Available: http://www.forbes.com/2009/08/10/government-internet-software-technology-breakthroughs-oreilly.html. [Accessed: 15-Sep-2014].

[9] Open Knowledge Foundation, “Wheredoesmymoneygo.org.” [Online]. Available: http://www.wheredoesmymoneygo.org/. [Accessed: 16-Sep-2014].

[10] “HOMER Project Website.” [Online]. Available: http://homerproject.eu. [Accessed: 23-Sep-2014].

[11] “HOMER Project Federation.” [Online]. Available: http://opendata-federation.csi.it/. [Accessed: 23-Sep-2014].

[12] “Hack4MED Website.” [Online]. Available: (http://hack4med.homerproject.eu/. [Accessed: 23-Sep-2014].

[13] “Publications Office of the European Union.” [Online]. Available: http://publications.europa.eu/index_en.htm. [Accessed: 16-Sep-2014].

[14] “AppsForEurope.eu.” [Online]. Available: http://appsforeurope.eu/. [Accessed: 16-Sep-2014].

[15] Open Data Institute, “Open Data Institute.” [Online]. Available: http://theodi.org/. [Accessed: 16-Sep-2014].

[16] Open Data Institute, “Open Data Institute - Startup Programme.” [Online]. Available: http://theodi.org/start-ups. [Accessed: 16-Sep-2014].

[17] Open Knowledge Foundation, “Open Knowledge.” [Online]. Available: https://okfn.org/. [Accessed: 20-May-2014].

[18] “Wikipedia - Open Data.” [Online]. Available: https://en.wikipedia.org/wiki/Open_Data. [Accessed: 16-Sep-2014].

[19] Open Knowledge Foundation, “Open Data Handbook Documentation - Release 1.0.0,” 2012. [Online]. Available: http://opendatahandbook.org/pdf/OpenDataHandbook.pdf. [Accessed: 21-May-2014].

[20] N. Korn and C. Oppenheim, “Licensing Open Data: A Practical Guide - June 2011 version 2.0,” 2010. [Online]. Available: http://discovery.ac.uk/files/pdf/Licensing_Open_Data_A_Practical_Guide.pdf. [Accessed: 16-Sep-2014].

[21] “Creativa Commons - About.” [Online]. Available: https://creativecommons.org/about. [Accessed: 16-Sep-2014].

[22] Creatice Commons, “Create Commons - Wizard to choose Licenses.” [Online]. Available: http://creativecommons.org/choose. [Accessed: 16-Sep-2014].

[23] “Open Data Support Iniative.” [Online]. Available: www.opendatasupport.eu. [Accessed: 16-Sep-2014].

[24] “Open Data Support iniative - Data & Metadata Licensing.” [Online]. Available: http://www.slideshare.net/OpenDataSupport/licence-your-data-metadata. [Accessed: 16-Sep-2014].

[25] Open Knowledge Foundation, “Open Data Commons - Legal tools for Open Data.” [Online]. Available: http://opendatacommons.org/. [Accessed: 06-May-2014].

[26] “Wikipedia - Open Government Licence.” [Online]. Available: http://en.wikipedia.org/wiki/Open_Government_Licence. [Accessed: 16-Sep-2014].

[27] “Open Data Handbook - Make Data Available (Technical Openness).” [Online]. Available: http://opendatahandbook.org/en/how-to-open-up-data/make-data-available.html. [Accessed: 16-Sep-2014].

[28] Open Knowledge Foundation, “Open Data Handbook - How to Open up Data.” [Online]. Available: http://opendatahandbook.org/en/how-to-open-up-data/index.html. [Accessed: 16-Sep-2014].

[29] “CITEK Project.” [Online]. Available: http://your-innovation.eu. [Accessed: 16-Sep-2014].

[30] CITEK, “CITEK Open Data Platform.” [Online]. Available: http://ckan.citek.ipn.pt/. [Accessed: 06-May-2014].

[31] Marche Region - Internazionalization and Innovation Department (IT), “Your Innovation website - Contacts.” [Online]. Available: http://www.your-innovation.eu/contancts. [Accessed: 12-Sep-2014].

[32] “Transcreativa Project.” [Online]. Available: http://www.transcreativa.eu/. [Accessed: 16-Sep-2014].

[33] Instituto Pedro Nunes, “RAW Open Data website.” [Online]. Available: http://rawopendata.ipn.pt. [Accessed: 12-Sep-2014].

[34] Open Knowledge Foundation, “CKAN - User Guide.” [Online]. Available: http://ckan.readthedocs.org/en/latest/user-guide.html. [Accessed: 06-May-2014].

[35] CITEK, “CITEK - Project partners contacts.” [Online]. Available: http://www.your-innovation.eu/contancts. [Accessed: 21-May-2014].

[36] “FCW - The Business of Federal Tecnology - ‘A brief history of open data.’”[Online]. Available: http://fcw.com/Articles/2014/06/09/Exec-Tech-brief-history-of-open-data.aspx?Page=2. [Accessed: 08-Oct-2014].