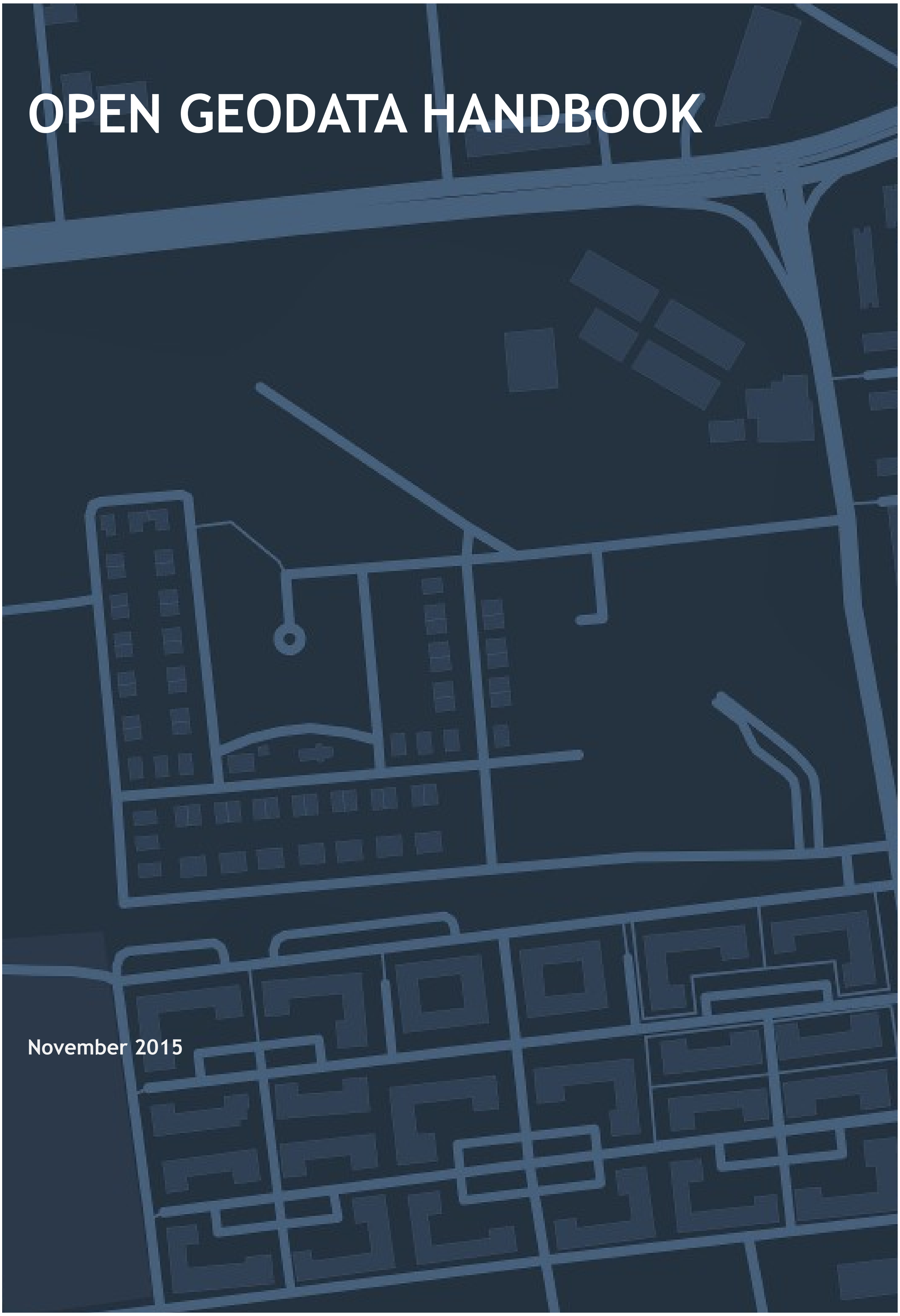
IKG
Institute of Cartography and Geoinformation

ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Universitatea Tehnică de Construcţii Bucureşti
Centrul de Cercetare Ingineria Apelor Subterane

OPEN GEODATA HANDBOOK



November 2015

FNSNF

PROGRAMUL DE COOPERARE ELVEȚIANO-ROMÂN
SWISS-ROMANIAN COOPERATION PROGRAMME

UEFISCDI

Table of Contents

Introduction

About this guide

This guide expands the *Open Data Handbook*[1], developed by *Open Knowledge*[2], to cover technical aspects related to open geodata publishing. This guide discusses the legal, social and technical aspects of open data. It can be used by anyone but is especially designed for those seeking to open up data/geodata. The work was carried out in the framework of *GEOIDEA.RO research project*[3]. The guide is available under the *Creative Commons Attribution 4.0 International License*[4] terms.

Geodata

Geodata is a broad term that refers to data that has a spatial component, defined through various methods, such as pairs of coordinates, name of location, address identifiers etc. Its usage is wide spread over various domains, such as: natural resources, government, mapping, health services, transportation, communications and utilities, military, public safety and the list may continue.

Open data

The notion of open data and specifically open government data - information, public or otherwise, which anyone is free to access and re-use for any purpose - has been around for some years. In 2009 open data started to become visible in the mainstream, with various governments (such as the USA, UK, Canada and New Zealand) announcing new initiatives towards opening up their public information.

GEOIDEA.RO project

GEOIDEA.RO stands for Geodata Openness Initiative for Development and Economic Advancement in Romania. It is a research project and a collaboration between the *Institute of Cartography and Geoinformation (IKG)*[5] at ETH Zurich, Switzerland and the *Groundwater Engineering Research Center (CCIAS)*[6] at the Technical University of Civil Engineering Bucharest, Romania. It aims at improving the scientific basis for open geodata model adoption in Romania. Led by the belief that publishing government geodata in Romania over the Internet, under an open license and in a reusable format can strengthen citizen engagement and yield new innovative businesses, we are building a collaboration with six selected governmental geodata producers in order to define the best methodology and tools to release geodata.

[1] http://opendatahandbook.org
[2] https://okfn.org
[3] http://geoidea.ro
[4] http://creativecommons.org/licenses/by/4.0/
[5] http://www.ikg.ethz.ch/en
[6] http://www.ccias.utcb.ro/en/home.html

Why Open Data?

Open data, especially open government data, is a tremendous resource that is as yet largely untapped. Many individuals and organisations collect a broad range of different types of data in order to perform their tasks. Government is particularly significant in this respect, both 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. Why is that of interest?

There are many areas where we can expect open data to be of value, and where examples of how it has been used already exist. There are also many different groups of people and organisations who can benefit from the availability of open data, including government itself. At the same time it is impossible to predict precisely how and where value will be created in the future. The nature of innovation is that developments often comes from unlikely places.

It is already possible to point to a large number of areas where open government data is creating value. Some of these areas include:

- 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

Examples exist for most of these areas.

In terms of transparency, projects such as the Finnish ‘tax tree’ and British ‘where does my money go’ show how your tax money is being spent by the government. And there’s the example of how open data saved Canada $3.2 billion in charity tax fraud. Also various websites such as the Danish folketsting.dk track activity in parliament and the law making processes, so you can see what exactly is happening, and which parliamentarians are involved.

Open government data can also help you to make better decisions in your own life, or enable you to be more active in society. A woman in Denmark built findtoilet.dk, which showed all the Danish public toilets, so that people she knew with bladder problems can now trust themselves to go out more again. In the Netherlands a service, vervuilingsalarm.nl, is available which warns you with a message if the air-quality in your vicinity is going to reach a self-defined threshold tomorrow. In New York you can easily find out where you can walk your dog, as well as find other people who use the same parks. Services like ‘mapumental’ in the UK and ‘mapnificent’ in Germany allow you to find places to live, taking into account the duration of your commute to work, housing prices, and how beautiful an area is. All these examples use open government data.

Economically, open data is of great importance as well. Several studies have estimated the economic value of open data at several tens of billions of Euros annually in the EU alone. New products and companies are re-using open data. The Danish husetsweb.dk helps you to find ways of improving the energy efficiency of your home, including financial planning and finding builders who can do the work. It is based on re-using cadastral information and information about government subsidies, as well as the local trade register. Google Translate uses the enormous volume of EU documents that appear in all European languages to train the translation algorithms, thus improving its quality of service.

Open data is also of value for government itself. For example, it can increase government efficiency. The Dutch Ministry of Education has published all of their education-related data online for re-use. Since then, the number of questions they receive has dropped, reducing work-load and costs, and the remaining questions are now also easier for civil servants to answer, because it is clear where the relevant data can be found. Open data is also making government more effective, which ultimately also reduces costs. The Dutch department for cultural heritage is actively releasing their data and collaborating with amateur historical societies and groups such as the Wikimedia Foundation in order to execute their own tasks more effectively. This not only results in improvements to the quality of their data, but will also ultimately make the department smaller.

While there are numerous instances of the ways in which open data is already creating both social and economic value, we don't yet know what new things will become possible. New combinations of data can create new knowledge and insights, which can lead to whole new fields of application. We have seen this in the past, for example when Dr. Snow discovered the relationship between drinking water pollution and cholera in London in the 19th century, by combining data about cholera deaths with the location of water wells. This led to the building of London's sewage systems, and hugely improved the general health of the population. We are likely to see such developments happening again as unexpected insights flow from the combination of different open data sets.

This untapped potential can be unleashed if we turn public government data into open data. This will only happen, however, if it is really open, i.e. if there are no restrictions (legal, financial or technological) to its re-use by others. Every restriction will exclude people from re-using the public data, and make it harder to find valuable ways of doing that. For the potential to be realized, public data needs to be open data.

What is Open Data?

Open data is data that can be freely used, re-used and redistributed by anyone - subject only, at most, to the requirement to attribute and sharealike.

The *full Open Definition*[7] gives precise details as to what this means. To summarize the most important:

- **Availability and Access**: the data must be available as a whole and at no more than a reasonable reproduction cost, preferably by downloading over the internet. The data must also be available in a convenient and modifiable form.
- **Re-use and Redistribution**: the data must be provided under terms that permit re-use and redistribution including the intermixing with other datasets.
- **Universal Participation**: everyone must be able to use, re-use and redistribute - there should be no discrimination against fields of endeavour or against persons or groups. For example, 'non-commercial' restrictions that would prevent 'commercial' use, or restrictions of use for certain purposes (e.g. only in education), are not allowed.

If you're wondering why it is so important to be clear about what open means and why this definition is used, there's a simple answer: **interoperability**.

Interoperability denotes the ability of diverse systems and organizations to work together (inter-operate). In this case, it is the ability to interoperate - or intermix - different datasets.

Interoperability is important because it allows for different components to work together. This ability to componentize and to 'plug together' components is essential to building large, complex systems. Without interoperability this becomes near impossible — as evidenced in the most famous myth of the Tower of Babel where the (in)ability to communicate (to interoperate) resulted in the complete breakdown of the tower-building effort.

We face a similar situation with regard to data. The core of a "commons" of data (or code) is that one piece of "open" material contained therein can be freely intermixed with other "open" material. This interoperability is absolutely key to realizing the main practical benefits of "openness": the dramatically enhanced ability to combine different datasets together and thereby to develop more and better products and services (these benefits are discussed in more detail in the section on 'why' open data).

Providing a clear definition of openness ensures that when you get two open datasets from two different sources, you will be able to combine them together, and it ensures that we avoid our own 'tower of babel': lots of datasets but little or no ability to combine them together into the larger systems where the real value lies.

Readers have already seen examples of the sorts of data that are or may become open - and they will see more examples below. However, it will be useful to quickly outline what sorts of data are, or could be, open – and, equally importantly, what won't be open.

[7] http://opendefinition.org/okd/

The key point is that when opening up data, the focus is on non-personal data, that is, data which does not contain information about specific individuals.

Similarly, for some kinds of government data, national security restrictions may apply.

How to Open up Data

This section forms the core of this handbook. It gives concrete, detailed advice on how data holders can open up data/geodata. We'll go through the basics, but also cover the pitfalls. Lastly, we will discuss the more subtle issues that can arise.

There are three key rules we recommend following when opening up data:

- **Keep it simple**. Start out small, simple and fast. There is no requirement that every dataset must be made open right now. Starting out by opening up just one dataset, or even one part of a large dataset, is fine – of course, the more datasets you can open up the better. Remember this is about innovation. Moving as rapidly as possible is good because it means you can build momentum and learn from experience – innovation is as much about failure as success and not every dataset will be useful.
- **Engage early and engage often**. Engage with actual and potential users and re-users of the data as early and as often as you can, be they citizens, businesses or developers. This will ensure that the next iteration of your service is as relevant as it can be. It is essential to bear in mind that much of the data will not reach ultimate users directly, but rather via 'info-mediaries'. These are the people who take the data and transform or remix it to be presented. For example, most of us don't want or need a large database of GPS coordinates, we would much prefer a map. Thus, engage with infomediaries first. They will re-use and repurpose the material.
- **Address common fears and misunderstandings**. This is especially important if you are working with or within large institutions such as government. When opening up data you will encounter plenty of questions and fears. It is important to (a) identify the most important ones and (b) address them at as early a stage as possible.

There are four main steps in making data open, each of which will be covered in detail below. These are in very approximate order - many of the steps can be done simultaneously.

1. **Choose your dataset(s)**. Choose the dataset(s) you plan to make open. Keep in mind that you can (and may need to) return to this step if you encounter problems at a later stage.
2. **Apply an open license**.
    1. Determine what intellectual property rights exist in the data.
    2. Apply a suitable 'open' license that licenses all of these rights and supports the definition of openness discussed in the section above on 'What Is Open Data'
    3. NB: if you can't do this go back to step 1 and try a different dataset.
3. **Make the data available -** in bulk and in a useful format. You may also wish to consider alternative ways of making it available such as via an API.
4. **Make it discoverable -** post on the web and perhaps organize a central catalog to list your open datasets.

Choose Dataset(s)

Choosing the dataset(s) you plan to make open is the first step – though remember that the whole process of opening up data is iterative and you can return to this step if you encounter problems later on.

If you already know exactly what dataset(s) you plan to open up you can move straight on to the next section. However, in many cases, especially for large institutions, choosing which datasets to focus on is a challenge. How should one proceed in this case?

Creating this list should be a quick process that identifies which datasets could be made open to start with. There will be time at later stages to check in detail whether each dataset is suitable.

There is no requirement to create a comprehensive list of your datasets. The main point to bear in mind is whether it is feasible to publish this data at all (whether openly or otherwise).

Asking the community

We recommend that you ask the community in the first instance. That is the people who will be accessing and using the data, as they are likely to have a good understanding of which data could be valuable.

1. Prepare a short list of potential datasets that you would like feedback on. It is not essential that this list concurs with your expectations, the main intention is to get a feel for the demand. This could be based on other countries' open data catalogs.
2. Create a request for comment.
3. Publicise your request with a webpage. Make sure that it is possible to access the request through its own URL. That way, when shared via social media, the request can be easily found.
4. Provide easy ways to submit responses. Avoid requiring registration, as it reduces the number of responses.
5. Circulate the request to relevant mailing lists, forums and individuals, pointing back to the main webpage.
6. Run a consultation event. Make sure you run it at a convenient time where the average business person, data wrangler and official can attend.
7. Ask a politician to speak on your agency's behalf. Open data is very likely to be part of a wider policy of increasing access to government information.

Cost basis

How much money do agencies spend on the collection and maintainence of data that they hold? If they spend a great deal on a particular set of data, then it is highly likely that others would like to access it.

This argument may be fairly susceptible to concerns of freeriding. The question you will need to respond to is, "Why should other people get information for free that is so expensive?". The answer is that the expense is absorbed by the public sector to perform a particular function. The cost of sending that data, once it has been collected, to a third party is approximately nothing. Therefore, they should be charged nothing.

Ease of release
Sometimes, rather than deciding which data would be most valuable, it could be useful to take a look at which data is easiest to get into the public's hands. Small, easy releases can act as the catalyst for larger behavioural change within organisations.

Be careful with this approach however. It may be the case that these small releases are of so little value that nothing is built from them. If this occurs, faith in the entire project could be undermined.

Observe peers
Open data is a growing movement. There are likely to be many people in your area who understand what other areas are doing. Formulate a list on the basis of what those agencies are doing.

Apply an Open License (Legal Openness)

In most jurisdictions there are intellectual property rights in data that prevent third-parties from using, reusing and redistributing data without explicit permission. 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.

What licenses can you use? We recommend that for 'open' data you use one of the licenses conformant with the Open Definition and marked as suitable for data. This list (along with instructions for usage) can be found at:

- http://opendefinition.org/licenses

A short 1-page instruction guide to applying an open data license can be found on the Open Data Commons site:

- http://opendatacommons.org/guide

Make Data Available (Technical Openness)

Open data needs to be technically open as well as legally open. Specifically, the data needs to be available in bulk in a machine-readable format.

- **Available**. Data should be priced at no more than a reasonable cost of reproduction, preferably as a free download from the Internet. This pricing model is achieved because your agency should not undertake any cost when it provides data for use.
- **In bulk**. The data should be available as a complete set. If you have a register which is collected under statute, the entire register should be available for download. A web API or similar service may also be very useful, but they are not a substitutes for bulk access.
- **In an open, machine-readable format**. Re-use of data held by the public sector should not be subject to patent restrictions. More importantly, making sure that you are providing machine-readable formats allows for greatest re-use. To illustrate this, consider statistics published as PDF (Portable Document Format) documents, often used for high quality printing. While these statistics can be read by humans,

they are very hard for a computer to use. This greatly limits the ability for others to re-use that data.

Here are a few policies that will be of great benefit:

- Keep it simple,
- Move fast
- Be pragmatic.

In particular it is better to give out raw data now than perfect data in six months' time.

There are many different ways to make data available to others. The most natural in the Internet age is online publication. There are many variations to this model. At its most basic, agencies make their data available via their websites and a central catalog directs visitors to the appropriate source. However, there are alternatives.

When connectivity is limited or the size of the data extremely large, distribution via other formats can be warranted. This section will also discuss alternatives, which can act to keep prices very low.

Online methods

Via your existing website

The system which will be most familiar to your web content team is to provide files for download from webpages. Just as you currently provide access to discussion documents, data files are perfectly happy to be made available this way.

One difficulty with this approach is that it is very difficult for an outsider to discover where to find updated information. This option places some burden on the people creating tools with your data.

Via 3rd party sites

Many repositories have become hubs of data in particular fields. For example, pachube.com is designed to connect people with sensors to those who wish to access data from them. Sites like Infochimps.com and Talis.com allow public sector agencies to store massive quantities of data for free.

Third party sites can be very useful. The main reason for this is that they have already pooled together a community of interested people and other sets of data. When your data is part of these platforms, a type of positive compound interest is created.

Wholesale data platforms already provide the infrastructure which can support the demand. They often provide analytics and usage information. For public sector agencies, they are generally free.

These platforms can have two costs. The first is independence. Your agency needs to be able to yield control to others. This is often politically, legally or operationally difficult. The second cost may be openness. Ensure that your data platform is agnostic about who can access it. Software developers and scientists use many operating systems, from smart phones to supercomputers. They should all be able to access the data.

Via FTP servers

A less fashionable method for providing access to files is via the File Transfer Protocol (FTP). This may be suitable if your audience is technical, such as software developers and scientists. The FTP system works in place of HTTP, but is specifically designed to support file transfers.

FTP has fallen out of favour. Rather than providing a website, looking through an FTP server is much like looking through folders on a computer. Therefore, even though it is fit for purpose, there is far less capacity for web development firms to charge for customisation.

As torrents

BitTorrent is a system which has become familiar to policy makers because of its association with copyright infringement. BitTorrent uses files called torrents, which work by splitting the cost of distributing files between all of the people accessing those files. Instead of servers becoming overloaded, the supply increases with the demand increases. This is the reason that this system is so successful for sharing movies. It is a wonderfully efficient way to distribute very large volumes of data.

As an API

Data can be published via an Application Programming Interface (API). These interfaces have become very popular. They allow programmers to select specific portions of the data, rather than providing all of the data in bulk as a large file. APIs are typically connected to a database which is being updated in real-time. This means that making information available via an API can ensure that it is up to date.

Publishing raw data in bulk should be the primary concern of all open data intiatives. There are a number of costs to providing an API:

- The price. They require much more development and maintainence than providing files.
- The expectations. In order to foster a community of users behind the system, it is important to provide certainty. When things go wrong, you will be expected to incur the costs of fixing them.

Access to bulk data ensures that:

- there is no dependency on the original provider of the data, meaning that if a restructure or budget cycle changes the situation, the data are still available.
- anyone else can obtain a copy and redistribute it. This reduces the cost of distribution away from the source agency and means that there is no single point of failure.
- others can develop their own services using the data, because they have certainty that the data will not be taken away from them.

Providing data in bulk allows others to use the data beyond its original purposes. For example, it allows it to be converted into a new format, linked with other resources, or versioned and archived in multiple places. While the latest version of the data may be made available via an API, raw data should be made available in bulk at regular intervals.

For example, the Eurostat statistical service has a bulk download facility offering over 4000 data files. It is updated twice a day, offers data in Tab-separated values (TSV) format, and includes documentation about the download facility as well as about the data files.

Another example is the District of Columbia Data Catalog, which allows data to be downloaded in CSV and XLS format in addition to live feeds of the data.

Make data discoverable

Open data is nothing without users. You need to be able to make sure that people can find the source material. This section will cover different approaches.

The most important thing is to provide a neutral space which can overcome both inter-agency politics and future budget cycles. Jurisdictional borders, whether sectorial or geographical, can make cooperation difficult. However, there are significant benefits in joining forces. The easier it is for outsiders to discover data, the faster new and useful tools will be built.

Existing tools

There are a number of tools which are live on the web that are specifically designed to make data more discoverable.

One of the most prominent is the DataHub and is a catalog and data store for datasets from around the world. The site makes it easy for individuals and organizations to publish material and for data users to find material they need.

In addition, there are dozens of specialist catalogs for different sectors and places. Many scientific communities have created a catalog system for their fields, as data are often required for publication.

For government

As it has emerged, orthodox practice is for a lead agency to create a catalog for the government's data. When establishing a catalog, try to create some structure which allows many departments to easily keep their own information current.

Resist the urge to build the software to support the catalog from scratch. There are free and open source software solutions (such as CKAN) which have been adopted by many governments already. As such, investing in another platform may not be needed.

There are a few things that most open data catalogs miss. Your programme could consider the following:

- Providing an avenue to allow the private and community sectors to add their data. It may be worthwhile to think of the catalog as the region's catalog, rather than the regional government's.
- Facilitating improvement of the data by allowing derivatives of datasets to be cataloged. For example, someone may geocode addresses and may wish to share those results with everybody. If you only allow single versions of datasets, these improvements remain hidden.
- Be tolerant of your data appearing elsewhere. That is, content is likely to be duplicated to communities of interest. If you have river level monitoring data available, then your data may appear in a catalog for hydrologists.

- Ensure that access is equitable. Try to avoid creating a privileged level of access for officials or tenured researchers as this will undermine community participation and engagement.

For civil society

Be willing to create a supplementary catalog for non-official data.

It is very rare for governments to associate with unofficial or non-authoritative sources. Officials have often gone to great expense to ensure that there will not be political embarrassment or other harm caused from misuse or overreliance on data.

Moreover, governments are unlikely to be willing to support activities that mesh their information with information from businesses. Governments are rightfully skeptical of profit motives. Therefore, an independent catalog for community groups, businesses and others may be warranted.

Standards and file formats for online geodata publishing

The Open Geospatial Consortium (OGC) gathers stakeholders in the area of geo-enable Web, such as industries, government agencies and universities. They develop publicly available interface standards through consensus building (Open Geospatial Consortium 2013). The following standards are the most relevant ones for open geodata publishing:

Catalogue Service for the Web (CSW)

The OGC Catalogue Service interface standard specifies a design pattern for defining interfaces to publish and search collections of descriptive information (metadata) about geospatial data, services and related information objects. Providers of resources, such as content providers, use catalogues to register metadata that conform to the provider's choice of an information model; such models include descriptions of spatial references and thematic information. Client applications can then search for geospatial data and services in very efficient ways.

There are several profiles of the current OGC CSW model. These include:

- ISO 19115/19139 Metadata profile (http://www.iso.org/iso/catalogue_detail.htm?csnumber=32557). This document specifies an application profile for ISO metadata with support for XML encoding per ISO 19139 (http://www.iso.org/iso/catalogue_detail.htm?csnumber=32557) and HTTP protocol binding. This CSW profile is widely implemented in Europe, such as in the Spatial Data Infrastructure for North Rhine Westphalia (federal state of Germany).
- CSW-ebRIM Registry Service. (http://portal.opengeospatial.org/files/?artifact_id=31137) This profile applies the CSW interfaces to the OASIS ebXML registry information model (ebRIM 3.0) so as to provide a general and flexible web-based registry service that enables users—human or software agents—to locate, access, and make use of resources in an open, distributed system; it provides facilities for retrieving, storing, and managing many kinds of resource descriptions. An extension mechanism permits registry content to be tailored for more specialized application domains.
- CSW 39.50: The Z39.50 Protocol binding uses a message-based client server architecture implemented using the ANSI/NISO Z39.50 Application Service Definition and Protocol Specification3. This protocol binding maps each of the general model operations to a corresponding service specified in the ANSI/NISO/ISO standard http://lcweb.loc.gov/z3950/agency/document.html.

Much of the current work on this standard has to do with restructuring the catalogue standard so that there is a well defined, easy to implement core coupled with a well defined mechanism for expressing a variety of extensions (previously known as application profiles).

Web Map Service (WMS)

The OGC Web Map Service Interface Standard (WMS) provides a simple HTTP interface for requesting geo-registered map images from one or more distributed geospatial databases.

The response to the request is one or more map images (returned as JPEG, PNG, etc) that can be displayed a browser and desktop applications

The WMS standard defines three operations:

- GetCapabilities (required): Obtain service-level metadata, which is a machine-readable (and human-readable) description of the WMS's information content and acceptable request parameters.
- GetMap (required): Obtain a map image whose geospatial and dimensional parameters are well defined.
- GetFeatureInfo (optional): Ask for information about particular features shown on a map.

This standard is now implemented in hundreds of products, many of which are registered at (http://www.opengeospatial.org/resource/products/compliant).

Web Coverage Service (WCS)

The OGC Web Coverage Service Interface Standard (WCS) defines a standard interface and operations that enable interoperable access to geospatial "coverages". The term "grid coverages" typically refers to content such as satellite images, digital aerial photos, digital elevation data, and other phenomena represented by values at each measurement point. (http://www.opengeospatial.org/standards/wcs).

The OGC Web Coverage Service (WCS) Interface Standard is a data service. The WCS standard defines a data access service that enables coverages, such as digital elevation models, to be queried using an HTTP based interface. The response to a WCS request includes coverage metadata and an output coverage whose pixels are encoded in a specified binary image format, such as GeoTIFF or NetCDF.

Web Feature Service (WFS)

The OGC Web Feature Service Interface Standard (WFS) defines web interface operations for querying and editing vector geographic features, such as roads or lake outlines. (http://www.opengeospatial.org/standards/wfs).

- The WFS standard defines operations that enable clients to:
- Discover which feature collections are available (GetCapabilities)
- Describe the attribute fields available for features (DescribeFeatureType)
- Query a collection for a subset of features based on a provided filter (GetFeature)
- Add, edit or delete features (Transaction)

All WFSs support input and output data using Geography Markup Language (GML). Some WFSs also support other encodings, such as GeoRSS or shapefiles.

Users typically interact with WFSs through browser based or desktop geospatial clients, which allows them to access vector map layers from external agencies, over the Internet.

Geography Markup Language (GML)

GML (http://www.opengeospatial.org/standards/gml) is an XML grammar defined to express and communicate geographical features. GML serves as a modeling language for geographic systems as well as an open interchange format for geographic transactions on the Internet. Note that the concept of feature in GML is a very general one and includes not only conventional “vector” or discrete objects, but also coverages and some elements of sensor data. The ability to integrate all forms of geographic information is key to the utility of GML.

GML contains a rich set of primitives that are used to build application specific schemas or application languages. These primitives include:

- Feature
- Geometry
- Coordinate Reference System
- Topology
- Time
- Dynamic feature
- Coverage (including geographic images)
- Unit of measure
- Directions
- Observations
- Map presentation styling rules

Understanding the use of “Feature”, “Geometry”, and “CRS” is critical in the use and development of any GML based encoding.

OGC work on the GML standard began in 1998. GML was first formally approved as an OGC standard in 2001. GML became an ISO standard in 2007. GML 3.2.16 is the most current revision of the joint OGC-ISO standard. Versions 3.2.2 and 4.0 are currently in progress.

KML Encoding Standard (KML)

The OGC KML Encoding Standard is an XML language focused on geographic visualization, including annotation of maps and images. Geographic visualization includes not only the presentation of graphical data on the globe, but also the control of the user’s navigation in the sense of where to go and where to look. (http://www.opengeospatial.org/standards/KML).

In 2006, Google submitted KML (formerly Keyhole Markup Language) to the OGC for consideration as a standard. KML was the first instance of a de-facto standard being submitted into the OGC standards process and as such the OGC modified our standards approval process to accommodate standards that have been developed externally from the OGC and then submitted into the OGC process. There were four objectives for this standards work:

That there be one international standard language for expressing geographic annotation and visualization on existing or future web-based online maps (2d) and earth browsers (3d). That KML be aligned with international best practices and standards, thereby enabling greater uptake and interoperability of earth browser implementations. That the OGC and KML community will work collaboratively to insure that the KML implementer community is properly engaged in the process and that the KML community is kept informed of progress and issues. That the OGC process will be used to insure proper life-cycle management of the KML candidate specification, including such issues as backwards compatibility.

KML was approved as an OGC standard in 2008 because the Membership believed that having KML as an OGC standard would encourage broader implementation and greater interoperability and sharing of earth browser content and context.

KML is complementary to most of the existing OGC specifications including key standards such as GML (Geography Markup Language), WFS (Web Feature Service) and WMS (Web Map Service). Currently, KML (v2.1) utilizes certain geometry elements derived from GML (version 2.1.2). These elements include point, line-string, linear-ring, and polygon.

GeoRSS

GeoRSS is an extension to the common RSS (Really Simple Syndication) used on websites to notify readers of new articles or updates. GeoRSS adds geographic coordinates and features to RSS and Atom items (Turner 2006). There are currently three encodings of GeoRSS:

- Simple (lightweight format that developers and users can quickly and easily add to their existing feeds with little effort),
- GML (supports a greater range of features, notably coordinate reference systems other than WGS-84 latitude/longitude),and W3C Geo (RDF vocabulary for representing the latitude and longitude of spatially-located objects).

GeoJSON

GeoJSON (Geometry JavaScript Object Notation) is a recently developed format based on JavaScript Object Notation (JSON). GeoJSON is geared toward consumption by Ajax oriented applications (such as OpenLayers) because its output notation is in JavaScript format. JSON is the standard object representation in JavaScript data structures. GeoJSON extends JSON by defining a format for geometry storage within the JSON format. One important extension, called TopoJSON, brings topology support and some other advanced features to GeoJSON. Rather than representing geometries discretely, geometries in TopoJSON files are stitched together from shared line segments called arcs. In addition, TopoJSON facilitates applications that use topology, such as topology-preserving shape simplification, automatic map colouring, and cartograms.

Free/Libre and open source tools for open geodata publishing

Free and open source software development is emerging as an alternative approach for developing large software systems. New types and new kinds of software processes are emerging within FOSS projects, as well as new characteristics for development project success, when compared to those found in traditional industrial software projects and those portrayed in software engineering textbooks. As a result, FOSS offers new types and new kinds of processes to research, understand, improve, and practice. The main advantages of FOSS software are:

- The availability of the source code and the right to modify and use the software in any way;
- Not tied to a single vendor;
- Big community to support;
- Good security, reliability & stability;
- Very good standard compliancy;
- Lower implementation cost.

The open source geospatial space includes products to fill every level of the spatial data infrastructure stack. Open source software can provide a complete alternative to proprietary software in most system designs.

Tools for open geodata publishing through web services/APIs

PostgreSQL is a powerful open source object-relational database system, with an emphasis on extensibility and standards-compliance. It has a development history of almost 20 years and therefore it has become widely used by individuals and companies due its reliability, data integrity and correctness. It runs on major operating systems: Linux, UNIX and Windows.

PostGIS is a spatial database extender for PostgreSQL object-relational database, meaning it adds support for geographic objects. PostGIS follows the Simple Features for SQL specification from the Open Geospatial Consortium (OGC). Using PostGIS one can perform location queries in SQL.

QGIS Server provides web map and feature services (WMS & WFS) using the same libraries as the GIS (QGIS) desktop application. Maps and print templates created in QGIS desktop can be published as web maps simply by copying the QGIS project file into the server directory. The resulting web maps look exactly the same as in the desktop. QGIS Server is usually run as CGI/FastCGI module within the Apache Webserver.

GeoServer is a web server that allows you to serve maps and data from a variety of formats to standard clients such as web browsers and desktop GIS programs. Data is published via standards based interfaces, such as WMS, WFS, WCS, WPS, Tile Caching and more. GeoServer comes with a browser-based management interface and connects to multiple data sources at the backend.

MapServer is an Open Source geographic data rendering engine written in C. Beyond browsing GIS data, MapServer allows you create "geographic image maps", that is, maps that can direct users to web content. The same application serves as a "map engine" for other portions of the site, providing spatial context where needed. MapServer was originally developed by the University of Minnesota (UMN) ForNet project in cooperation with NASA, and the Minnesota Department of Natural Resources (MNDNR). Later it was hosted by the TerraSIP project, a NASA sponsored project between the UMN and a consortium of land management interests.

deegree is a standards-based Java framework for spatial data infrastructures and the geospatial web. It includes the OGC Web Map Service (WMS) reference implementation, a fully compliant Web Feature Service (WFS) as well as packages for Catalogue Service (CSW), Web Coverage Service (WCS), Web Processing Service (WPS) and Web Map Tile Service (WMTS).

Tools for open geodata discovery

GeoNetwork OpenSource is a catalog application for managing spatially referenced resources. It provides powerful metadata editing and search functions, an embedded interactive web map viewer, and is based upon Open Standards.

pycsw enables users to publish a catalogue of dataset descriptions (metadata) to the web, using standards based interfaces and formats, such that it can be queried and updated by catalogue clients. Pycsw offers support for the following international standards ISO 19115, ISO 19139 and FGDC CSDGM.

Web frameworks for open geodata visualization

Web frameworks support the development of dynamic web applications by providing libraries, packages and templates into a single connected piece of software and, thus, alleviating the tasks of dealing with low-level details. More specifically in the geospatial world, web frameworks allow to easily combine the visualization of geodata and to publish them online without having to deal with every single detail regarding projection and interactivity functions. For instance, basic functions to navigate the map, such as zooming and panning, are already built in.

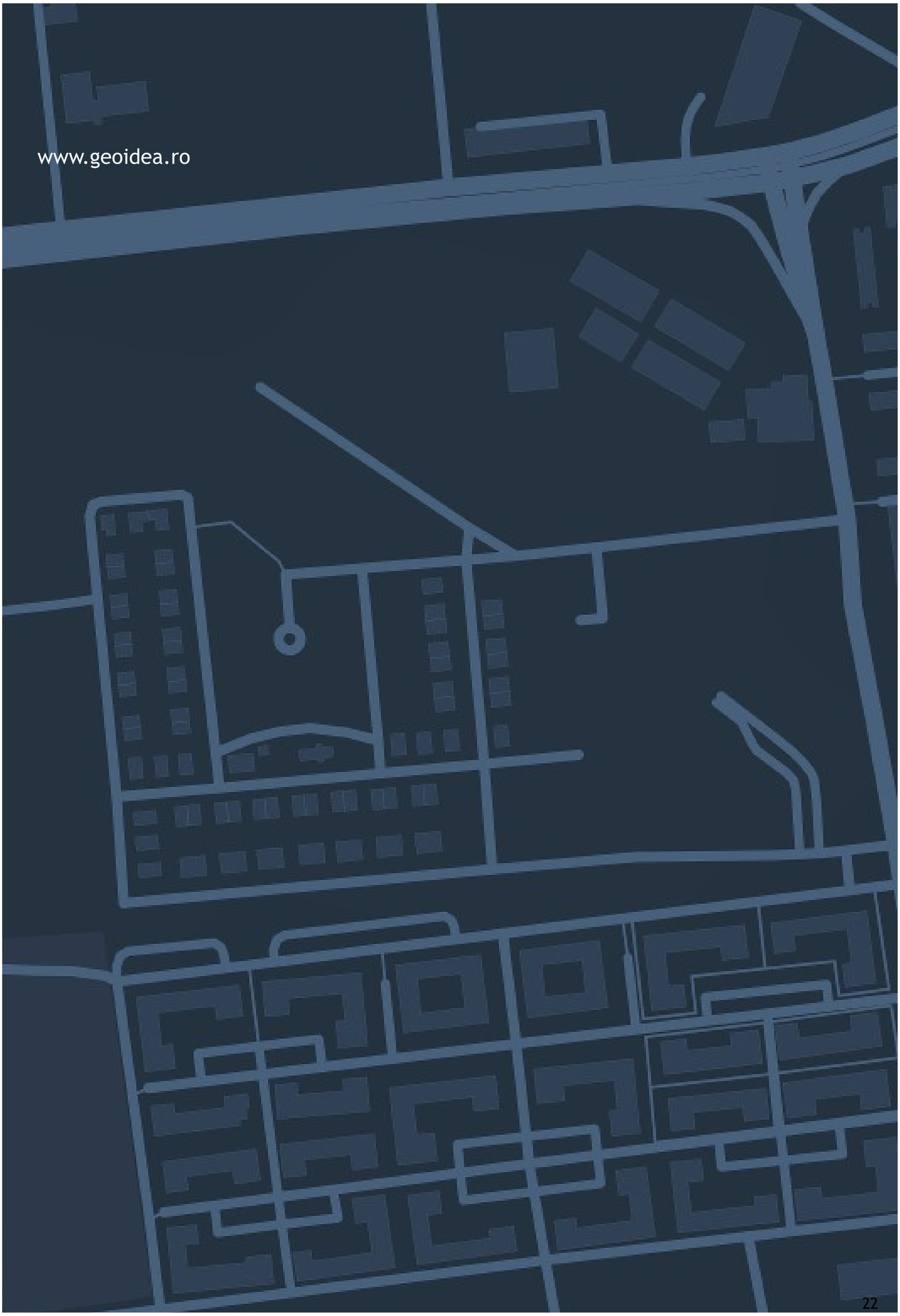
OpenLayers 3 is a free and open source JavaScript framework that is licensed under a BSD 2-clause license and is an Open Source Geospatial Foundation project. It offers the possibility to publish dynamic maps in any webpage. Moreover, it is supported by most modern browsers because it is pure JavaScript and that there are no server dependencies. OpenLayers 2 supports many different data formats, such as WMS, WMTS, GeoJSON, GML, KML, GeoRSS and WFS which enables this framework to be compatible with the major data formats that can be found in the geospatial web field. It offers spatial navigation tools and other built-in functions such as switch layer, show coordinates, pan, draw and select functions. The default projection system used is the Spherical Mercator, but there is support for reprojection, however, an additional library might be needed. On a side note, a WMS should be able to reproject automatically the layers when sending them to Open Layers. Version 3 offer is a comprehensive rewrite of the code of the popular OpenLayers 2 implementation and is integrated the WebGL library (for 3D visualization) and the Cesium library (3D globe), as well as target more precisely HTML5 and CSS3 (web standards for web page and styling).

Leaflet is a free and open source software and is very similar to OpenLayers. However, as said in their motto “simplicity, performance and usability”, it aims at being simpler and smaller in its core functions. Then, it is extended by a plugin system offering the more complex and specialized functions. It supports WMS, WMTS, GeoJSON, and vector data and allows for spatial navigation functions and layer control. Like OpenLayers, the important standards that are HTML5 and CSS3 are also integrated. Dealing with different projection systems also requires the use of an additional library.

jQuery is a cross-platform JavaScript library designed to simplify the client-side scripting of HTML. jQuery, free and open-source software licensed under the MIT License, is the most popular JavaScript library in use today. jQuery's will be used within CTEP Browser to make it easier to navigate a document, select DOM elements, create animations, handle events, and manage AJAX requests.

D3.js is a JavaScript library for producing dynamic, interactive data visualizations in web browsers. It makes use of the widely implemented SVG, HTML5, and CSS standards.

Bootstrap is a free and open source (MIT license) collection of tools for creating websites and web applications. It contains HTML and CSS based design templates for typography, forms, buttons, navigation and other interface components, as well as optional JavaScript extensions. It aims to ease the development of dynamic websites and web applications. Bootstrap will be heavily use for the GeoBrowser GUI.



www.geoidea.ro