

INFORMATION EXTRACTION AND DEPENDENCY ON OPEN GOVERNMENT DATA (OGD) FOR ENVIRONMENTAL MONITORING

Hussein Abdulmuttalib*

Dubai Municipality, GIS Department, Maktoom St., Dubai - hussenma@dm.gov.ae

ISPRS – UN-GGIM Forum

KEY WORDS: Open Government Data, Environmental Monitoring, GIS Data, spatial Analysis, Geo-design, Geo-knowledge

ABSTRACT:

Environmental monitoring practices support decision makers of different government / private institutions, besides environmentalists and planners among others. This support helps them act towards the sustainability of our environment, and also take efficient measures for protecting human beings in general, but it is difficult to explore useful information from 'OGD' and assure its quality for the purpose. On the other hand, Monitoring itself comprises detecting changes as happens, or within the mitigation period range, which means that any source of data, that is to be used for monitoring, should replicate the information related to the period of environmental monitoring, or otherwise it's considered almost useless or history.

In this paper the assessment of information extraction and structuring from Open Government Data 'OGD' , that can be useful to environmental monitoring is performed, looking into availability, usefulness to environmental monitoring of a certain type, checking its repetition period and dependences. The particular assessment is being performed on a small sample selected from OGD, bearing in mind the type of the environmental change monitored, such as the increase and concentrations of built up areas, and reduction of green areas, or monitoring the change of temperature in a specific area. The World Bank mentioned in its blog that Data is open if it satisfies both conditions of, being technically open, and legally open. The use of Open Data thus, is regulated by published terms of use, or an agreement which implies some conditions without violating the above mentioned two conditions. Within the scope of the paper I wish to share the experience of using some OGD for supporting an environmental monitoring work, that is performed to mitigate the production of carbon dioxide, by regulating energy consumption, and by properly designing the test area's landscapes, thus using Geodesign tactics, meanwhile wish to add to the results achieved by many efforts to make OGD useful In General and specifically for Environmental Monitoring purposes.

1. OGD OPEN GOVERNMENT DATA

1.1 OGD, Regulations and Geodesign

“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. (Open foundation, Open Data Handbook Documentation, Release 1.0.0)

Relating openness to Geodesign and Geoknowledge is mostly expressed by the capability of providing the needed information and data when mostly required, in the required usable format, showing an example of which, can be the Vital Landscapes Project in Hungary aimed to interpret landscape management as an interactive, integrative, conscious and open procedure. Geodesign approach supports the frames of the research workflow and study area activities of the Hungarian project partner, they have used a knowledge model for the design, “Knowledge base module is a platform that refers on all other modules or integrates parts of them. (Sándor JOMBACH& others,2012)

Hans-Georg SCHWARZ-v and others also used the openness in his / their phrase phrases, while explaining the ideas behind Geodesign, “the landscape designers call for a more open, process-oriented and flexible design process “,(Hans-Georg SCHWARZ-v, 2012)

Quoting from the World Bank's blog about Open Government Data “Data is open if it satisfies both conditions below:

- Technically open: available in a machine-readable standard format, which means it can be retrieved and meaningfully processed by a computer application
- Legally open: explicitly licensed in a way that permits commercial and non-commercial use and re-use without restrictions“.

The use of Open Data is regulated by published terms of use or an agreement which implies some conditions without violating the above mentioned two conditions.

1.2 Open Knowledge Organization, OGD catalogues

From the Open Knowledge organization an open government data working group was formalized, which is in fact open by itself to anyone interested in OGD to join and contribute, the

* Corresponding author

major purposes of this group fall under being central point of reference for interested parties in OGD, develop the principles openness of official government data, document the background and status of initiatives to make official information open in different countries, and finally provide the Support to develop OGD catalogues with ensuring interoperability of the different participating platforms.

The three major products of the working group are the open data handbook which is a general guide for open data, also formulating the principles of OGD in guidelines built upon the Open Knowledge Definition and being developed at: (<http://www.opendefinition.org/government/>), and the third and final major product is the Data Catalogues, where an introductory to the data is provided besides supporting setting up instances of CKAN www.ckan.net around the world, which is a tool playing the role between the groups of interest and the governments providing the row data.

1.3 The Data Handbook

The data handbook answers the questions of why, what, how and then what is the meaning, after downloading or opening government data, also highlights the major benefits of making government data open and give examples for the same, which are but not limited to:

1. Transparency and democratic control,
2. Participation,
3. Self-empowerment,
4. Improved or new private products and services,
5. Innovation,
6. Improved efficiency of government services,
7. Improved effectiveness of government services,
8. Impact measurement of policies,
9. New knowledge from combined data sources and patterns in large data volumes.

open data handbook is prepared and published on the website link: (<http://opendatahandbook.org/pdf/OpenDataHandbook.pdf>) to guide

1.4 Interoperability and OGD

An important measure which makes open data valuable and relate it to open knowledge, geoknowledge and geodesign is the interoperability, and for clarifying the importance of this fact to OGD some quotations are provided below from the open handbook that are selected to pin out the important issue.

” Interoperability is important because it allows for different components to work together. This ability to componentized and to ‘plug together’ components is essential to building large, complex systems”. ”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”, “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.”,(Open foundation, Open Data Handbook Documentation, Release 1.0.0).

In the discussion of this paper for using OGD to address the aspect of information extraction and structuring that is of implantation importance to environmental monitoring and sustainability, the major three key rules recommend by the working group for opening up data are assessed in this respect:

1.5 Rules of OGD & Relations to Environmental Monitoring

1.5.1 Keep it simple

Open Data Recommendations	Relations to Environmental Monitoring
Start out small, simple and fast	Fast can relates to data availability and temporal accuracy Small can relates to lack of data Simple is not always good for accuracy
There is no requirement that every dataset must be made open right now	Encouraging good for the long run, But decencies and requires commitment
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 good for getting more data and more reliable data and information

Table 1. Rules of OGD working group

1.5.2 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, This will ensure that the next iteration of your service is as relevant as it can better.	Accessibility to data and improvement Coverage and temporal accuracy Improvements towards reliability and completeness
It is essential to bear in mind that much of the data will not reach ultimate users directly, but rather via 'info-me diaries'. These are the people who take the data and transform or remix it to be presented	<ul style="list-style-type: none"> - Readable formats - Interoperability - May cause a delay - May cause data errors

Table 2. More rules of OGD working group

1.5.3 Address common fears and misunderstandings

Identify and address important fears	<ul style="list-style-type: none"> - Can add to metadata - May reduce availability and accessibility
--------------------------------------	--

Table 3. Other rules of OGD working group

1.6 Is Open Data Enough?

These elements are theoretical, but practically it is a subject of the type of environmental monitoring in hand and depends on the organization providing the data, and from here the question arises if only open data is enough or more than that is required, "Open Data is not enough, we need open tools and community to utilize open data (Rufus Pollock, 2011), Existing tools such as those created by different groups for example scarpwiki, <https://scrapwiki.com>, which are a sort of computer programs prepared to assist downloading Government information and make it more accessible to citizens.

They created websites like:

- TheyWorkForYou <http://theyworkforyou.com>, or
- Another tool such as Google Refine, <http://code.google.com/p/google-refine/wiki/Importers>.

That is to help in understanding a variety of data file formats. Currently, it tries to guess the format based on the file extension. For example, .xml files are of course in XML. By default, an unknown file extension is assumed to be either tab-separated value (TSV) or comma-separated value (CSV). Google Refine looks for a tab character; if one is found, it assumes a TSV format.

The formats currently supported (in version 2.0) include a bunch of formats which ease the process of searching and using open data, some of these formats include: TSV, CSV, Excel (.xls, .xlsx), XML, RDF as XML, JSON, Google Spreadsheets and RDF N3 triples among others.

- Also another tools are CKAN and the data hub tool can be found under, <http://ckan.org/>, <http://datahub.io/>,

CKAN is a powerful data management system that makes data accessible – by providing tools to streamline publishing, sharing, finding and using data. CKAN is aimed at data publishers (national and regional governments, companies and organizations) wanting to make their data open and available

Coming back to the answer to the question if only open data is enough or more than that is required and to the second part of the answer, "Open Data is not enough, we need open tools and community to utilize open data (Rufus Pollock, 2011), Building up user communities is the third element in the main set for OGD, such as:

- The community Sunlight found under the link: <http://sunlightfoundation.com/>, where they introduce themselves as:

"The Sunlight Foundation is a nonprofit, nonpartisan organization that uses the power of the Internet to catalyze greater government openness and transparency, and provides new tools and resources for media and citizens, alike. We are committed to improving access to government information by making it available online, indeed redefining "public" information as meaning "online," and by creating new tools and websites to enable individuals and communities to better access that information and put it to use"(quoted from <http://sunlightfoundation.com/>), Some of the important roles these communities play in benefiting the usage of OGD which can also ad to the practices of Environmental Monitoring are:

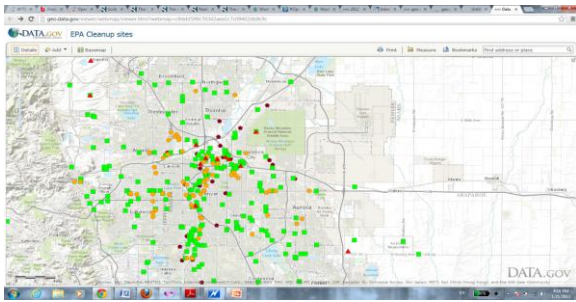
- catalyze greater government transparency
- Demanding policies that will enable to hold government accountable.
- Develops and encourages new government policies to make it more open and transparent.
- Make data searchable, sorted and machine readable databases,
- Builds tools and websites to enable easy access to information.
- Fosters distributed research projects as a community building tool.
- Require that government make data available in real time and trains users.

2. OGD EXAMPLES FOR ENVIRONMENTAL MONITORING

When coming to practical examples to OGD and especially to Geographic data that can be of value to environmental monitoring, examples selected as follows:

- From Geo Data Gov: <http://geo.data.gov/geoportal/catalog/search/resource/details.page?uuid=%7B4216352E-146C-4134-8575-8869F50DEFC9%7D>

La Parguera, Puerto Rico Water Quality Monitoring Data 2003 was chosen for illustrating environmental monitoring open data type.

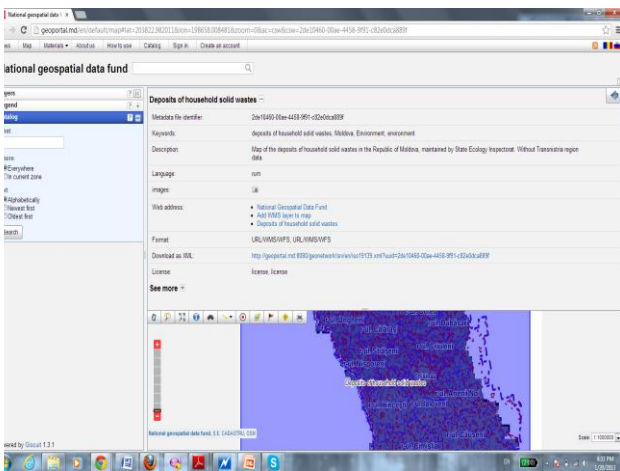


Clip 1. Example of OGD for Environmental Monitoring

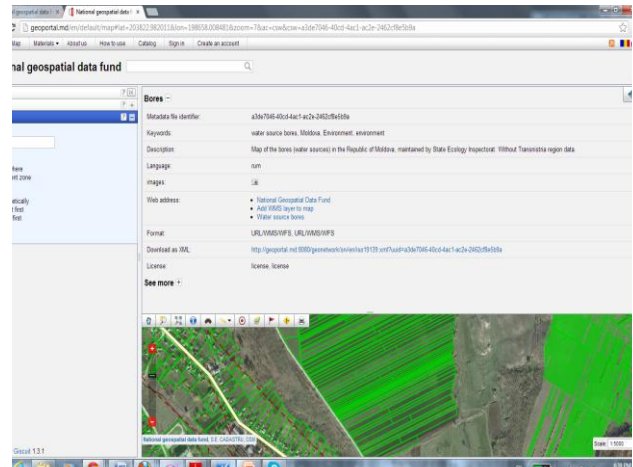
“These water quality data are one of many studies being done to assess and monitor coral reef ecosystems. The intent of this work is three fold:

- (1) to spatially characterize and monitor the distribution, abundance, and size of both reef fishes and mega-invertebrates (conch, lobster, Diadema);
- (2) To relate this information to in-situ data collected on water quality and associated habitat parameters;
- (3) to use this information to establish the knowledge base necessary for enacting management decisions in a spatial setting and to establish the efficacy of those management decisions” (NOAA’s Ocean Service, National Centers for Coastal Ocean Science (NCCOS)

2.1 Other OGD Clip Examples Related to Environmental Monitoring



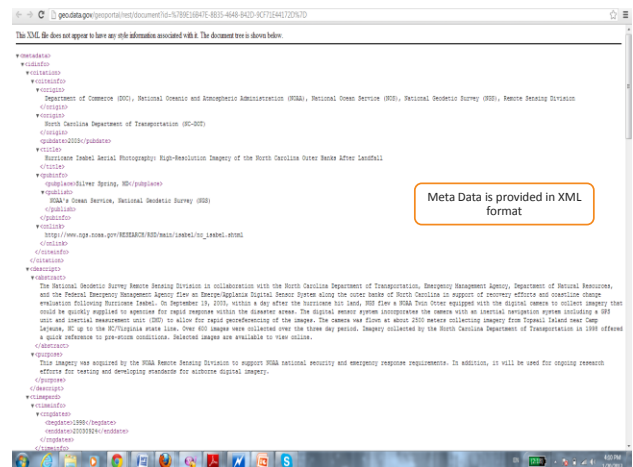
Clip 2. Example of OGD for Environmental Monitoring



Clip 3. Example of OGD for Environmental Monitoring



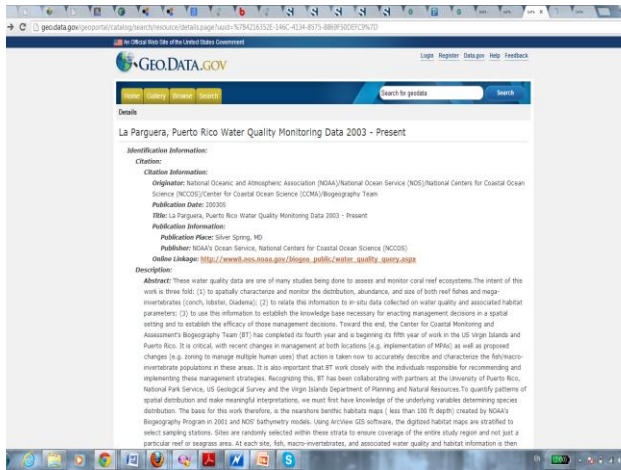
Clip 4. Example of OGD for Environmental Monitoring



Meta Data is provided in XML format

Clip 5. Example of OGD for Environmental Monitoring

Meta Data



Clip 6. Example of OGD for Environmental Monitoring
Water Quality Monitoring

3. CONCLUSION

It appears that huge efforts have been performed in the area of OGD, but still a lot to come and expected, the usage of open data for monitoring Environmental variations is possible in certain areas for specific tasks, but yet more efforts are expected towards formulizing this aspect of OGD in conjunction with Geoknowledge and Geodesign, nevertheless many Governments have released data to the public a matter which is expanding as formalized groups are pushing towards openness in a regulated manner.

REFERENCES

Sándor JOMBACH, László KOLLÁNYI, József László MOLNÁR, Áron SZABÓ and Tádé Dániel TÓTH, (2012). GeoDesign Approach in Vital Landscapes Project. DLA Conference, DLA Proceedings 2012 GeoDesign, 3D Modeling and Visualization

Daniel Dietrich, Jonathan Gray, Tim McNamara, Antti Poikola, Rufus Pollock, Julian Tait, Ton Zijlstra. Open foundation, Open Data Handbook Documentation, Release 1.0.0

Hans-Georg SCHWARZ-v.RAUMER and Antje STOKMAN, (2012). GeoDesign – Approximations of a Catchphrase. DLA Conference, DLA Proceedings 2012 GeoDesign, 3D Modeling and Visualization

Alek Tarkowski, Rufus Pollock, Carl-Christian Buhr, Neelie Kroes, Ellen Miller and David Eaves, (2011). Growing the ecosystems, data tools, community. Open Government Data Camp First Day Keynotes

Lisa M. Wedding; Barbara A. Gibson; William J. Walsh; Tim A. Battista, (2011). Integrating remote sensing products and GIS tools to support marine spatial management in West Hawai'i.

Journal of Conservation Planning Vol 7 (2011) 60–73. NOAA's Ocean Service, National Centers for Coastal Ocean Science (NCCOS)

HANIF RAHEMTULLA, (2011). The Journey of Open Government and Open Data Moldova. Information and Communications for Development, Promoting Information and Communications for Development (IC4D), world bank's blog

Hussein M. Abdulmuttalib, (Technical Specifications of Environmental Monitoring Project), Dubai Municipality 2003