THE OPEN DATA CONCEPT

MARIUS MIHĂIȚĂ CORCIOVĂ[1], PATRICIA NEAGA[2]

Abstract: Open data is information that is provided by public entities to be accessed and reused. Publishing open data is an excellent way to improve an organization's transparency and provide insight into the value of the organization. "Open Data" refers to data collected and shared with others to use as they wish, without restrictions on copyright or usage. Traditional examples of open data include government-collected data (e.g., weather reports, crime incident reports) as well as some academic sources (e.g., open-access journals, raw polling and survey data). Businesses are also realizing the benefits of sharing data and making it available for use. Getting started with open data involves identifying key open data resources such as data.gov, identifying library information that would be beneficial to publish, and creating programs that provide digital literacy training and create opportunities for patrons to engage with open data in new and creative ways [1].

Keywords: Open Data, Government data, Open Source Software, Open portal

1. INTRODUCTION

Open data and content can be freely used, modified, and shared by anyone for any purpose. An open work must satisfy the following requirements in its distribution:

- Open license - must allow free use of the licensed work; must allow redistribution of the licensed work, including sale, whether on its own or as part of a collection made from works from different sources;
- Open Access - the work must be available as a whole and at no more than a reasonable one-time reproduction cost, preferably downloadable via the Internet without charge;
- Open Data Format - the work must be provided in a convenient and modifiable form such that there are no unnecessary technological obstacles to the performance of the licensed rights.

[1] *Ars Industrial SRL, Romania*
[2] *Innova Global Consulting SRL, Romani, office@innovacons.ro.*

Whether a Data license is open or not, can be best checked via the two standard authorities for Open Data:

- Creative Commons - a nonprofit organization;
- Open Knowledge Foundation. [2]

Figure 1 shows that CC0, CC-BY (Attribution) and CC-BY-SA (Attribution-ShareAlike) are licenses which can be considered as licenses supporting Free Cultural Works, while the least open license is the NC-ND. CC-BY-SA can be considered equivalent to the Open Data base license (ODbL), which is maintained by the Open Knowledge Foundation [2].

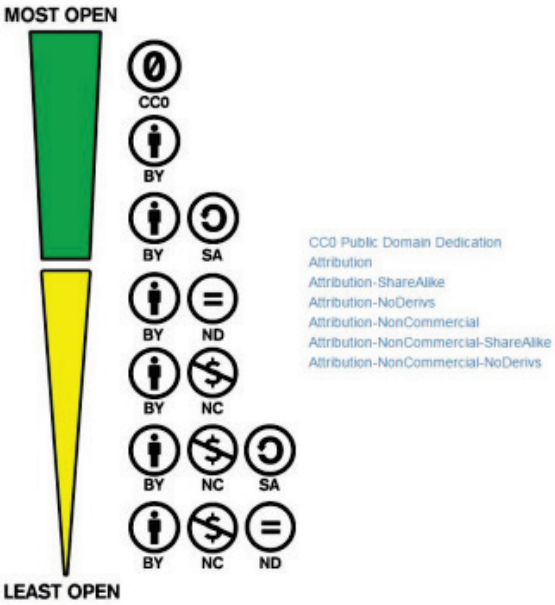

Fig.1. CC0, CC-BY (Attribution) and CC-BY-SA (Attribution-ShareAlike) licenses [2]

Open Data is important not only at a level of organization or country, it can also be considered a basis of global growth and democratization of society. An opposite of Open Data is copyrighted data and, in extreme cases, monopolized data. The big scientific publishers for decades have been collecting copyrights for publications written, edited and reviewed by scholars. [2]

2. OPEN DATA INITIATIVES

Below is a detailed description and analysis of Open Data initiatives of the United Kingdom, the United States of America and the World Bank. [3]

- UK datasets: The United Kingdom has achieved a significant success in developing its open data portal, http://data.gov.uk/. It features with over 5.600 datasets from all central government organizations and a number of

other public bodies and local authorities, which are published in various machine-readable formats, such as XLS, CSV, RDF and XML. The portal also allows downloading the metadata from the appropriate dataset in JSON format. Data can be filtered on several criteria, such as publishers or institutions, nation (England, Wales and Scotland), tags, types and it can also be sorted by relevance, title, ranking.

- US datasets: The US open data portal, http://www.data.gov/, is a leading portal that shows the way for democratization of data and introduces innovation. The data available on the portal is gathered from several locations within the government. The portal has over 305.800 datasets which are published in various machine-readable formats, such as XML, CSV/TXT and RDF. The datasets can be queried by keywords, by categories such as finance, education, or by public institutions. The datasets include: review of deposits; investment of companies and class information; information about government projects; index of producer prices; annual data for electricity; business employment dynamics; projections of employees; industry program for funding; resources needed for the budget.
- World Bank datasets: The World Bank has an open data portal as well. It has two APIs for providing access to different datasets: one for indicators – time series data – and one for projects – data in relation to the operations of the World Bank. Both APIs implement RESTful interfaces which enable users to perform queries over the available data, using selection parameters. This technology allows for use of the services from both applications and web browsers. The indicators and projects from the datasets are available in XML, JSON and ATOM formats. The indicator API provides access to more than 3.000 indicators, some even spanning 50 years in the past. The API for projects provides access to all of the World Bank projects, including both active and closed projects. [3]

2.1. Open Data and the OpenStreetMap example

An Open Data oriented organization stimulates companies to build up on the basis of data and software products with minimum administrative requirements (which under the CC-NC is not possible). One of the best examples of such organization is the OpenStreetMap foundation. [2]

Fig.2. OpenStreetMap [2]

Commercial applications of OpenStreetMap includes:

- Web-mapping applications used by Apple, Flickr, MapQuest, MapBox and similar;
- Navigation and vehicle tracking software (offline);
- Geographic analysis and spatial planning. [2]

3. BIG DATA AND OPEN DATA

Data generated by people and sensors and later processed and stored by computers have reached incredible levels and not only are assuming a crucial role inside the information systems of organizations but also represent an incredible source of opportunities for even the individual citizen. Data have gone from relatively small-scale databases relegated inside information systems to superabundant data made available from many different data sources (Big Data), with this heterogeneity mainly constituted by data coming from structured institutional and enterprise datasets, unstructured social media, sensors and people themselves. All these new perspectives of data management bring new, relevant benefits but also significant headaches, especially for public administrators and IT personnel who are on the forefront of this data deluge and who have to provide professional, prompt and efficient responses to these challenges. [4]

Big Data and the related area of Big Data Analytics have such large applications and datasets to manage (from terabytes to exabytes) with a very high level of variety and complexity (from sensors to social media data) that they require new advanced data storage and management tools and new analysis and data visualization technologies. [4]

Some of this data must be given back to the original owners, the citizens, because that data have a public origin and public destination; the Open (Government) Data is another idea of managing data and is here to stay. Open Data is particularly engaging to many public administrations of the world, under the strong pressure of stakeholders who consider (open) data a form of democracy and an opportunity for business. [4]

Engagement with Open Data should meet the following:

★ Be demand driven: choices about the released data should be based on community needs and demands. This also includes answering the data requests of users.

★★ Consider putting data into context: clear metadata must be provided, including a description of the data quality, the frequency of updates, the data formats, the lineage and links to any additional document that describes how to address the data (manuals, publications, reports, deliverables, analysis of the data, etc.).

★★★ Support conversations pertaining to the data: users should be able to comment on datasets or create structured conversations around data with other users. Data publishers, metadata catalogue maintainers and individual "data owners" should be able to be contacted directly if needed. [3]

★★★★ Build capacity, skills and networks: links to tools for people who would like to work with the published datasets should be provided, along with a "How to Guidance" and Open Data analysis tools, so that people can increase their capacity and skills for interpreting and using data in the manner they need to. Additionally, skill-building sessions on using data in particular ways and sponsoring or engaging with capacity building to help the community work with open data should be put in place.

★★★★★Stimulate collaboration on data as a common resource: feedback loops to improve the quality of datasets and collaborations with communities to create new data resources, lend support to build and sustain useful tools and services that work with the published data, and work with other organizations to connect published data sources should be adopted. [4]

5. OPEN DATA PORTAL - CKAN

An open data portal is any online platform which supports users in accessing collections of open data. Typical open data portals present the data of the organization which hosts the portal. [5]

Government organizations sometimes host open data portals as a way of meeting their regional freedom of information legal requirements. Another common use case is open data portals for sharing data in some field of research for the benefit of other researchers. The simplest open data portal is list of datasets with instructions for how anyone can access and use that data.

Characteristics of good open data portals include the use of open standards, access to data without human intervention, and analytics about what data people use.

Open data portals contain information of interest to citizens, business owners, nonprofit administrators, researchers, and journalists. [5]

CKAN is the world's leading open-source data portal platform. CKAN makes it easy to publish, share and work with data. It's a data management system that provides a powerful platform for cataloging, storing and accessing datasets with a rich front-end, full API (for both data and catalog), visualization tools and more. [6]

CKAN is a tool for making open data websites. (Think of a content management system like WordPress - but for data, instead of pages and blog posts.). It helps to manage and publish collections of data. It is used by national and local governments, research institutions, and other organizations who collect a lot of data.

Registration is needed for most publishing features and for personalization features, such as "following" datasets. It is not needed to search for and download data.

To create a user ID, use the "Register" link at the top of any page. CKAN will ask for the following:

- Username – choose a username using only letters, numbers, - and _ characters. For example, "jbloggs" or "joe_bloggs93";
- Full name – to be displayed on your user profile;
- E-mail address – this will not be visible to other users;
- Password – enter the same password in both boxes. [7]

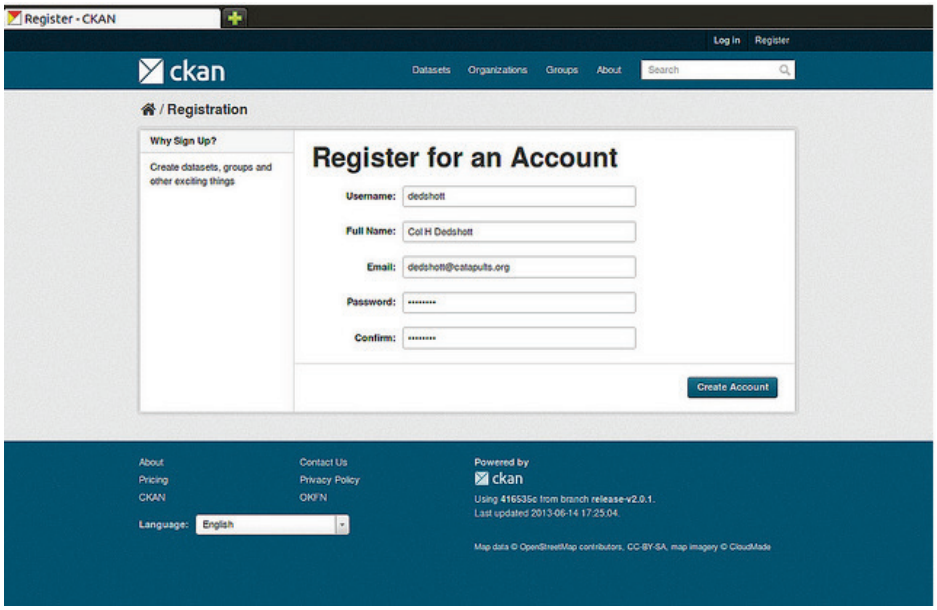

Fig.3. Registering and logging in [7]

6. CONCLUSIONS

Open data can bring benefits in various fields, such as health, food security, education, climate, intelligent transport systems, and smart cities - and is considered an essential resource for economic growth, job creation and societal progress, opens in a new window.

Open data publication and re-use brings a variety of benefits, including:

- Increasing the quality, efficiency, and transparency of public services;
- Greater efficiency in processes and delivery of public services;
- Cost saving. [8]

ACKNOWLEDGEMENTS

The project is founded by the EU under the Operational Program Competitivity (POC) from The European Regional Development Fund (FEDR), priority axis 2 – Information and Communication Technology (ICT) for a competitive digital economy, action 2.2.1. – Supporting the added value generated by the ICT sector and innovation in the field through cluster development.

REFERENCES

[1]. **Lori Bowen Ayre & Jim Craner.** (2017). *Open Data: What It Is and Why YouShould Care*. Public Library Quarterly, 36:2, 173-184, 2017. DOI: 10.1080/01616846.2017.1313045.

[2]. **Eisenmann, Thomas R. and Parker, Geoffrey and Van Alstyne, Marshall W.,** *Opening Platforms: How, When and Why?* (August 31, 2008). This paper has been published under the same title as Chapter 6 in Platforms, Markets & Innovation (ed. Gawer, 2009) pp 131-162, Harvard Business School Entrepreneurial Management Working Paper No. 09-030, Available at SSRN: https://ssrn.com/abstract=1264012 or http://dx.doi.org/10.2139/ssrn.1264012

[3]. **Hengl, T., Wheeler, I., & MacMillan, R. A.** (2018). *A brief introduction to Open Data, Open Source Software and Collective Intelligence for environmental data creators and users*. PeerJ Preprints, 6, e27127v2. https://doi.org/10.7287/peerj.preprints.27127v1.

[4]. **Kostovski, M., Jovanovik, M., & Trajanov, D.** (2012, September). Open data portal based on semantic web technologies. In Proceedings of the 7th South East European Doctoral Student Conference (pp. 1-13). University of Sheffield. 10.13140/RG.2.2.23588.88969.

[5]. **Molinari, A., Vincenzo, M., Lorenzino, V., Andrea, A., and Eleonora Bassi.** (2014). *Big data and open data for a smart city*. IEEE-TN Smart cities white papers, Trento, Italy. DOI:10.13140/RG.2.2.33170.56001

[6]. https://en.wikipedia.org/wiki/Open_data_portal

[7]. https://github.com/ckan/ckan

[8]. http://docs.ckan.org/en/2.9/user-guide.html#what-is-ckan

[9]. https://data.europa.eu/en/datastories/benefits-and-value-open-data

This page intentionally left blank