ASPECTS OF DATA CONSUMPTION
FROM MULTIPLE SOURCES

Adrian Nicolae MATEIA

Abstract: Today, the internet is available to anyone, anywhere. The benefit of Linked Open Data is to put information and data into a context that creates new knowledge. These data are two important viewpoints: publishing and consuming. Thus data management has become a important factor for business success and innovation. Open public data are those data produced or held by the public authorities and at the citizens' disposal freely, are free to access, reuse and redistribute.
Keywords: data, linked open data, open government data, reuse, pulishing, consuming.

INTRODUCTION

Data can be defined as a representation of information or records stored electronically and they can contain: documents, databases, videos or sounds records etc.

The concept of *Open Data (OD)* originates in the concept of transparent and participatory government. Thus, the open data must take the form of a bidirectional transfer between the administration and the citizens.

The RDF *(Resource Description Frame)* metadata model is based on the idea of making declarations about the resources in the form of expressions *subject, predicate, object* named triples. The subject denotes the resource, the predicate expresses a relationship between the subject and the object. RDF is an abstract model with several serialization formats and the way a triple is encoded varies from format to format. A triple can be described as 3 URI.

A collection of RDF statements represents an oriented and labeled pseudo-graph. As such, an RDF-based data model is more naturally suited to certain kinds of knowledge representation than the relational model. However, in practice, RDF data is often stored in relational database representations also called *Triple Stores*.

In the IT environment, *URI (Uniform Resource Identifier)* is a series of characters used to identify the name of a web resource. Such an identification

allows the interaction with representations of web resources through a network, usually World Wide Web, using specific protocols.

Wiki is a software managed on a server that allows the users to create, change and interconnect easily web pages. This type of software uses the collaborative power of the users to centralize the information.

Semantic Web is a collection of information linked between them in such a manner that it can be easily used by machines, at global scale. We can consider it as an efficient way of representing data on the World Wide Web or as a database globally linked.

Modeled after SQL, SPARQL query language seems to become a de-facto RDF language. SPARQL can be used to make queries on different data sources no matter if they are stored as RDF or seen as RDF through intermediary softwares. SPARQL has the ability to query graph patterns with their conjunctions and disjunctions. SPARQL supports tests on extensible values or restrictions on the queries on RDF graphs. The SPARQL results can be sets of RDF results or graphs.

Linked Open Government Data (LOGD)

Open data and *open government data (OGD),* represents an important resource for different entities that collect a great number of data types in order to reach different objectives. Among these entities, the government administration holds a significant place in this context, due to the quantity and the centralization of the data, as well as due to the fact that the majority of government data are public by law, and thus they can be open, becoming available to the public.

Open public data are those data produced or held by the public authorities and at the citizens' disposal freely, are free to access, reuse and redistribute. They are characterized mainly by the following characteristics:

- can be processed by automatic means;
- are supplies in an open format, on which no entity has exclusive control and, also, have an open license, which allows the free use of data, inclusively on commercial purpose, without being limited by rights of intellectual property.

Interconnected data (linked data) are a standard method of data representation on different subjects. The publication of interconnected data offers the developers the possibility to put in contact information from different sources and thus to realize new and innovative applications.

The main idea of an open government consists in linking the political class, the public administration, the industry and the citizens, having as a goal the transparency, the democracy, the participation and collaboration of these entities.

From the definitions found in the specialty literature two main ideas can be deducted:

- e-government aims at the operationalisation of a new model of government where it is emphasized the need to reconfigure the relations developed with the different categories of interested parts - internal (employees) and external (citizens, organizations, other public institutions etc.);
- on the other side, e-government is the result of the transformations imposed by the adoption and use of the information and communication technology in the public systems.

In the context of those mentioned above, the following clarifications are necessary:

- the citizens'participation at the process of e-government implies two distinct levels: on the one side the citizens are consumers of public goods and services and, on the other side, are an essential part of the democratic system;
- the association of the e-government with the ICT operationalisation is the object of several controversies in the specialty literature. The assimilation process of the information and communication technology in the public system is not relatively new.

Romania is situated in the stage of minor transformations, which implies: the presence in the on-line environment, in order to improve the process of informing the interested parties; an average level of interaction. From the perspective of the evolution model of the e-government, Romania is situated in the interaction stage, without being completely integrated (64%).

The key element of the e-government concept is to improve the relation and the exchange of information between the public sector, on the one side, and the citizens and the business environment, on the other side. E-government means to offer public services in electronic format for the citizens and business environment, a more efficient and cheaper alternative, which would allow the Government to be closer to the citizens and adjust its services according to their needs.

In order to implement the "e-government" concept a country must have adequate infrastructures of data, institutional and technological systems, a compatible legal framework, human resources adequately trained and, last but not the least, a strategic thinking and coordination.

Functioning like a multi-channel communication system between the government, citizens, business environment and services suppliers, the e-government solutions would allow the administrative organisms to ensure a better access to information, to centralise and integrate data and applications, to make more efficient the decision-making process in public administration, to increase the incomes, to decrease the costs.

In order to successfully implement the "e-government" it is necessary to define a coherent architecture of applications and a predefined set of generic services and instruments for the development, implementation and future administration of the applications. This architecture must be updated in order to ensure the adjustment to the users' demands and the new technologies that appear.

LOGD emphasizes three stages:

- In the first stage (*open*), the government agencies play an important role in publishing government data in the on-line environment, in a reusable format and in keeping central the OGD catalogues which support the citizens with relevant and valid data.
- In the second stage (*link*), the communitary participants (industry and academic environment, for example), help to consolidate the quality of OGD data. The human intervention and the calculation power can be used to generate additional declarative links (as vocabulary standard, the maping concept and links to external data) and value-added services (as the automatic extraction of entity and resolution).
- In the third stage (*reuse*), the developers use the OGD data sets in order to build applications as good (useful) as possible. In the future, the data emerging markets will become mechanisms which transform the contributions of voluntary addition of data in a profitable business sector.

LOGD represents a new paradigm of data integration for a long lasting OGD growth and, as a consequence, it can be considered a new enterprise as approach. In the first place, it is opened the application domain of data integration in environments traditionally closed by the enterprise, such as data deposits for the entire web. The users can access governmental data with crowdsourced data, private data, and many other types of nongovernmental data. In the second place, it makes possible a data-oriented architecture (DOA), which decouples complex objects of data in linked fine-grained reusable data from the Web. A service-oriented architecture (SOA) decouples the services used by the applications to make reusable other applications and systems; a DOA, on the contrary, decouples data to make them reusable. The application of this DOA principle on Web means that any entity can bring its contribution to LOGD by the implementation of partial contributions, but interconnected.

Data consumption from multiple sources

After publishing a set of interconnected data, it is necessary to make sure that the RDF links from external links refer to URIs in the data set. This thing is useful in order to ensure the fact that data can be discovered by the RDF

browsers and crawlers, and it can be realised by the filling in of the existing data sets (held by the same source or by third parties), with RDF links which indicate resources in the new data set.

Third parties may need to be convinced about the value of connection to a new set of data. The determinant factors in this kind of situation can be the value of the new set of data, the value it adds to the set of existing data, if they are interconnected and the cost of creating high quality links.

A strategy is to create necessary RDF links and to ask to third parties to include these triples in their sets of data. This approach was successfully used by DBpedia, which integrates different sets of links generated by third parties which are linked by DBpedia resources with those of other 114 data.

The availability of the interconnected data is a relatively recent phenomenon, the majority of the present applications being at the first generation (some even prototypes). These probably will suffer changes due to the learnt lessons from the development and implementation stages.

In spite all these, the applications already have a vision of what could be in the future an architectural model, which is in development in the area of interconnected data.

The applications of the interconnected data can be classified into two categories:
- generic applications
- applications with specified domain

CONCLUSIONS

The explosion of web technologies is due in the first place to the computers'evolution which have an increasing power of calculation and to the network equipments which ensure a large belt width (an increase of the data transfer speed).

In this context, it is extremely important to put the data at the public disposal under an open license, because it allows from a juridical point of view to freely reuse all tea data bases, works and other elements put at the public disposal through technically adequate means, without the fear of a future process or a discrimination towards an actor on the market. On the contrary, the data can be changed freely from the technical point of view, but we may have an illegality from the juridical point of view.

Bibliography

Bauer F., Kaltenbock M., -*Linked Open Data: The essentials*. A Quick Start Guide for Decision Markets

Berners-Lee T, Hendler J, Lassila O, "*The Semantic Web*", Scientific American, 5, 2001

Chen G., - *Ontology alignemnt techiques for linked open data ontologies*, 2013

B. Daum, U. Merten, *System Architecture with XML*, Elsevier Science, 2003

Ding L., Peristeras V., Hausenblas M.,- *Linked Open Government Data*

Neamţu F., - *Comunicare şi interacţiune în serviciile publice din România prin implementarea programelor de eGuvernare*, Teză de doctorat, 2013

* * *, http://opendatahandbook.org/ro/why-open-data/index.html

* * *, http://www.management.ase.ro/reveconomia/2002/11.pdf

* * * http://ogp.gov.ro/guvernare-deschisa/date-deschise/

NOTES ON THE AUTHOR

ADRIAN NICOLAE MATEIA, PhD., "Dimitrie Cantemir" Christian University, Faculty of Management in Tourism and Commerce Timisoara, Romania.
E-mail: anmateia@yahoo.com