

Challenges for open data Usage: Open Derivatives and Licensing

Miel Vander Sande¹, Marc Portier², Erik Mannens¹, and Rik Van de Walle¹

¹ Ghent University - IBBT - Multimedia Lab,
Gaston Crommenlaan 8 bus 201, B-9050 Ledeborg-Ghent, Belgium
{miel.vandersande,erik.mannens,rik.vandewalle}@ugent.be,

² Westtoer,
Koning Albert I-laan 120, B-8200 Sint-Michiels, Belgium,
marc.portier@westtoer.be

1 Introducing Westtoer and Multimedia Lab

IBBT, the Interdisciplinary institute for BroadBand Technology, is a research institute founded by the Flemish Government, focusing on information & communication technology (ICT) in general, and applications of broadband technology in particular. The IBBT was founded as a virtual research centre with over 800 researchers, based on research teams from existing knowledge centres.

Multimedia Lab (IBBT-MMLab) is a young research group within IBBT and part of the Ghent University (Faculty of Engineering). It accounts for about 25 researchers and it has a portfolio of basic research, applied research and contract-based research with industrial partners. IBBT-MMLab is very active within MPEG and W3C standardization via the submission of technical contributions, by chairing several ad-hoc groups, and through the editorship of several specifications. The main areas of expertise of IBBT-MMLab are: video coding and compression, image/video processing and analysis, multimedia content adaptation, meta-data technology and gaming technology. IBBT has gained a great amount of expertise and know-how on Semantic Web technologies and Linked Open Data in particular through its contributing role in several W3Cs Standardization groups, through participation in the Flemish research projects, a.o., PISA, GEISHA, MediaLoep, CUPID, and Archipel. Furthermore, IBBT-MMLab is core contributor to the freeyourmetadata.org movement and the Open Source DataTank framework.

Westtoer is an organization responsible for the promotion of Tourism (marketing, product and recreation development, support, consulting and knowledge gathering) in the province of West Flanders. The area covers all of the Belgian coast, Bruges and surroundings, the "Flanders Fields" region and the valley of the "Leie". Marc Portier works as a civil servant with Westtoer. Within the organization his role is to normalize, organize, educate and inspire on the proper usage of modern digital (social, local, mobile) tools, techniques and opportunities.

2 Motivation for attendance

At IBBT-MMLab we perform demand-driven research. Concerning Open Data projects, data consumption creates this demand. The more output is generated from Open Data, the more questions will rise. Many of these questions are perfect to apply our Semantic Web and Linked Data expertise to. This workshop offers a unique opportunity to get a broader view on how Open Data is used. Not only through the vision of peer organisations, but also from a consumer point of view. Together with Westtoer, we are currently looking into better information exchange between governmental organisations in Flanders. We believe that Semantic technologies applied to Open Data are the key for providing a stronger cooperation. Therefore, IBBT and Westtoer share a common interest in how Open Data can be used in order to find the right approach.

3 Consequences and Challenges of Open Data Usage

Today's Open Data movement demands governments to open their data based on some fundamental ideas. Government data is paid by the public and should therefore be accessible by the public. Also, government data is collected anyway, so why not let others benefit from this as well? Although these are well-founded arguments, they do not ensure the three potential benefits as cited by Neelie Kroes [1]. Governments should have the right and duty to release data using smart restrictions or extensions to ensure the underlying goals are reached. Therefore, a well-thought licensing model is more beneficial than blindly releasing data just because we can (or should). Only then can data usage be steered towards a guaranteed return-on-investment. Concerning Open Data usage, we believe two evolutions are crucial: moving toward data co-ownership and creating a feedback loop from Open Data output. To protect these prospects, a legal framework needs to be developed today.

3.1 Data co-ownership

The link between data owners and data consumer is most cases a one-way channel. When a consumer wants to provide the owner with feedback, a different communication channel is required (e.g., Email). This results in inaccurate data and non-efficient data management. Data exchange between owners and consumers needs to go from this heterogeneous dual-channelled system into a mono-channelled one. By changing data directly at the source through the way it was received, we can evolve to a data co-ownership: studying each others data becomes taking care of each others data. When applied on a data-field level, we move from a master-slave feeding system to a more granular structure. The right of ownership will then be based on the amount of effort one puts to maintain the correctness of a specific data-field, instead of the owner of the first version.

The concept of co-ownership requires clients to be conform with the one data exchange channel. This has some implications for data maintainers (actors storing the data) and data participators (actors providing feedback on the data).

Firstly, when working on a single two-way channel, a well-described communication protocol is needed, which needs to be supported by both parties. Secondly, data manipulating clients need to guarantee full feedback support, stating that they can provide all the feedback requested by the data maintainer. Finally, a clear identification of participators is mandatory, in order to assure trust and provenance.

3.2 Feedback loop of Open Data application

Usage of Open Data creates highly-valuable results on their part. However, the open character should somehow be reflected in the results, thus finding their way back into Open Data. Two principal outcomes can be discussed: usage statistics and Mashups/Aggregations.

Collected usage statistics are considered to be valued feedback to data owners, since it gives notion of relevance and enables segmentation. This allows companies to propose new related information to the user. Since the information market tends to develop monopolies (e.g., media, software), only the company in question would have a good view on these useful statistics. Regardless whether this is good or bad, in prospect of the goals of the Open Data movement it is certainly relevant to release these data as well. This requires clients able to provide the needed information to allow this monitoring.

In case of created mash-ups or aggregation on Open Data, we look for transparency in the conclusions drawn in these applications. Source referencing should be mandatory in order to validate Open Data outcomes and not only in Data Journalism. Anybody should be able to recalculate these outcomes at any time in order to avoid non-founded opinions. Also, produced output should partly be republished, preferably contributing to the originally used sources.

3.3 Supporting legal framework

To ensure that Open Data usage points governments in the right direction, a legal framework that covers all crucial aspects is essential. A first aspect is defining the domains of society where *openness* can be applied to. This definition is independent from the public, private or mixed nature of the organisations who try to cover these domains.

A second aspect includes some well-thought licenses that can enforce previously discussed restrictions. On other areas (e.g., web content, software), these viral approaches have already been well developed and well-used in *copy-left* licenses. In order to reuse this expertise, a modified or extended version of existing GPL³ or CC⁴ should meet our requirements.

A third aspect is facilitating government transparency. In this case, Open Data is used to recalculate any disclosed conclusion by the government. Legal constraints can encourage or oblige governments to invest in a platform enabling

³ GNU General Public License - <http://www.gnu.org/licenses/gpl.html>

⁴ Creative Commons - <http://creativecommons.org/>

this recalculation in an open and neutral manner. These need to be included in the legal framework if we want to grow towards true transparency.

References

1. Neelie Kroes. Digital agenda: Commission's open data strategy, questions & answers, 12 2011.