Initial Results from the Study of the Open Source Sector in Belgium

Robert VISEUR
UMONS Faculty of Engineering
Rue de Houdain, 9
B-7000 Mons
robert.viseur@umons.ac.be

CETIC
Rue des Frères Wright, 29/3
B-6041 Charleroi
robert.viseur@cetic.be

ABSTRACT

The economy of FLOSS (Free and open source software) has been the subject of numerous studies and publications, particularly on the issue of business models. However, there are fewer studies on the local networks of FLOSS providers. This research focuses on the ecosystem of Belgian FLOSS providers and, more specifically, their geographical distribution, the activities, technologies and software they support, their business models, their economic performance and the relationships between companies. The research is based on a directory containing nearly 150 companies. This directory led to the creation of a specialized search engine that helped to improve annotation. The research also uses financial data provided by the Belgian Central Balance Sheet Office. The initial results of this study show a concentration in major economic areas. The businesses are more active in the services and are heavily involved activities such as infrastructure software and Web development, activities which were common in the early years of free software development. Services for the support of business software is also common. A first analysis of the graph of relationships between providers' websites highlights the role that is played by the multinational IT companies, by FLOSS editors, by commercial FLOSS associations and especially by the Walloon centers of competence that offer vast training catalogs that are dedicated to FLOSS. This research opens up many perspectives for improving the automation of the company directory updates, the analysis of the relationship between enterprises, and the automation of the financial analysis of companies.

Categories and Subject Descriptors

K.1 [**Computing Milieux**]: The Computer Industry - *markets, standards, statistics, suppliers*.

General Terms

Economics.

Keywords

free software, open source, business model, market analysis, financial analysis, graph analysis, web.

1. CONTEXT

The economy of free and open source software (FLOSS) is the subject of numerous studies. The publications have highlighted the increased participation of companies in the development of FLOSS [8], studied the practices of industry in terms of business models [5, 6, 7], detailed some valuation methods [13] and developed famous companies as case studies [4, 11]. However, the studies on local FLOSS fabrics are uncommon. The cases that were identified include the study published in 2013 by the CNLL (National Council of Free Software) for France and the Vasquez Bronfman and Miralles' research for Spain [3, 14].

This research presents the preliminary results of a study on the Belgian sector of FLOSS providers. The term "providers" includes mainly private companies and some parastatals with a service activity in the field of FLOSS.

2. RESEARCH QUESTIONS

This preliminary research aims to provide an initial response to the following five questions:

(1) What is the geographical distribution of Belgian FLOSS providers?

(2) What activities, technology and software are being supported by Belgian FLOSS providers?

(3) What are the business models followed by Belgian FLOSS providers?

(4) Are Belgian FLOSS companies economically successful?

(5) What are the relationships between the Belgian FLOSS companies?

3. METHODOLOGY

This study required a list of providers and their domain names so a directory was created. This directory can be seen at www.logiciellibre.be (the equivalent for the French market is available at www.logiciellibre.com/fr/). Only companies that showed significant use of free software in their marketing communications were included in the directory. The creation of the directory was organized in three stages. First, a list of providers was compiled based on market knowledge, on existing partial lists and on Google searches based on a list of keywords. Then the information about providers was encoded. For each company, a record was created that included information such as the VAT (Value Added Tax) number, the company name, the address, the contact person and a set of keywords which summarize the activity of the company based on the content of their Web communications. A specialized search engine was then created. The Web crawling was fed by the companies' domain names. The tool helped to identify more specific skills in the already known companies and thus helped to complete the

OpenSym '14 , Aug 27-29 2014, Berlin, Germany
ACM 978-1-4503-3016-9/14/08.
http://dx.doi.org/10.1145/2641580.2641581

annotation. Finally, the results of the research were promoted at events dedicated to free software, in order to collect requests for additions and corrections, and to facilitate the updates.

The collection of Web pages was conducted several times in 2012 and 2013. GNU Wget (www.gnu.org/s/wget/) open source software was used for recursive Web crawling. This software recovered a total of 35,933 files (3.3 GB). These items were classified in a directory. Each folder at the root directory contains the files of a website. Before running Wget the URLs of domain names were analyzed to detect dead links and redirections (based on HTTP headers). The specialized search engine was performed with the Zend Search open source indexation tool.

The geographical distribution of FLOSS providers was determined automatically by the geolocation of their addresses. This operation was carried out based on the Lambert coordinates (Lambert 72) of the city where the provider is located. Lambert is a flat coordinates system used in Belgium by the National Geographic Institute (www.ngi.be). Based on these coordinates the business locations were plotted on a map of Belgium, which was found on Wikipedia and that shows the outlines of the Belgian provinces.

The fields of activity and the software that are supported by the companies were calculated by using the keywords that were added to the company records in the directory during the annotation process. For example, the companies that are active in open source voice over IP software were detected by keywords such as *voip*, *sip*, *asterisk*, *openser* or *ekiga* that represent various open source VoIP technologies. The dashboards were generated automatically by applying this method.

The business models were determined by using the content of the websites of a group of 31 companies during 2010. Five types of activity were identified: selling hardware, editing software, supplying services, selling products and hosting applications (SaaS, ASP and Web hosting). The service activity is associated with an assessment of the degree of specialization. The scale has three levels: general (= weak), focusing on a family of technologies (= medium), focusing on a particular technology (= strong). Whether or not companies contributed to FLOSS projects was also investigated.

The issue of economic performance was dealt with in this preliminary research. In practice, based on a file that contained VAT numbers, statistics were ordered from the Central Belgian Central Balance Sheet Office (www.nbb.be) that the companies must feed with their balance sheets. The data that was collected for these companies included their name, their legal status, the model of their annual account (full or abbreviated), their turnover (if published), their operating income, their net income for the year, their debts, their equity, the number of full-time employees and the financial ratios calculated by the Belgian Central Balance Sheet Office [1].

The relationships between the companies were studied based on the networking between the companies' websites that were in the directory. The pages that were collected with Wget were analyzed with a PHP script that extracts URLs and calculates a set of links between domains. The latter is then converted to DOT file. DOT is a graph description language described in text format to represent and shape the edges of a graph. It is based on nodes (e.g. "www.beeznest.com" → "twitter.com"). In this case, the nodes are websites and the edges are the links between websites. This file was then analyzed with the Gephi (gephi.org) open source software and was converted into a graph with the DOT converter on Ubuntu GNU/Linux. The approach used here was based on Ortega and Aguillo's study on the Nordic academic Web, and operated with the measures defined by Hansen *et al.* [9, 12].

Partial results have been the subject of presentations to professionals in Belgium and France, in order to get feedback on the results of the study [15, 16].

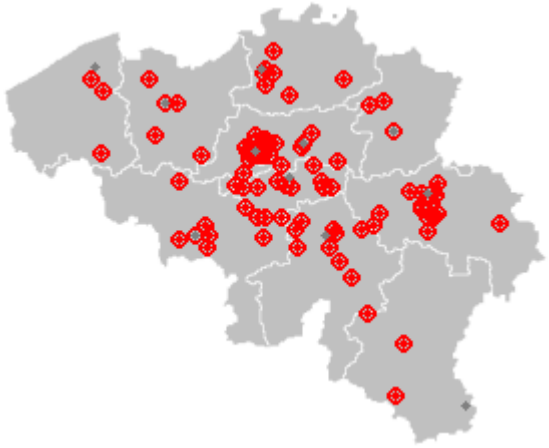
4. RESULTS

4.1 Geographical Distribution

A map of Belgium that shows the outlines of the provinces was used and the province capitals were highlighted. The digitized map, seen in Figure 1, is pale gray, the capital city of each province appears in dark gray and the companies in red.

Unsurprisingly, the major Belgian economic centers have the largest number of FLOSS providers (see Figure 1) with a significant concentration around the city of Brussels, the city of Antwerp and the former Walloon industrial basin (Mons-Charleroi-Liege). The southern province of Luxembourg is the least equipped.

Figure 1. Geographic distribution of Belgian FLOSS providers.



4.2 Supported Activities, Technologies and Software

Over one third of the providers are involved in activities that are related to IT infrastructure and Web development (see Table 1). There are domains where free software is historically well developed: Apache Web server, operating systems for GNU/Linux or BSD-type server, Samba file server, PHP Web development language, content management systems such as Drupal, Plone or SPIP, etc. The fields of Electronic Document Management (EDM), Voice over IP (VoIP) and business management software (ERP / CRM / BI) are also well covered. Almost one business in five offers training. The coverage is lower for the domains of embedded systems, Geographic Information Systems (GIS) and legal advice.

Table 1. Dashboard showing activities supported by Belgian FLOSS providers (total: 127 suppliers).

Activities	#	%
Infrastructure	48	37.80%

Activities	#	%
VoIP	14	11.02%
Embedded	5	3.94%
Web development and CMS	42	33.07%
LMS	2	1.57%
EDM	14	11.02%
ERP /CRM / BI	18	14.17%
Collaboration	5	3.94%
GIS	2	1.57%
Training	26	20.47%
Legal Advice	2	1.57%

Other statistics were generated regarding the support for Enterprise Resource Planning (ERP), Customer Relationship Management (CRM) and Business Intelligence (BI) (see Table 2), support for programming languages (see Table 3), support for Web frameworks and support for Content Management Systems (see Table 4).

Support for business management software (ERP / CRM / BI) is characterized by a very strong support for OpenERP. The latter is produced by a local company that is based in Wavre and that has experienced strong growth for several years.

Table 2. Dashboard showing business management software supported by Belgian FLOSS providers. (total: 16 suppliers).

Activities	#	%
Compiere	1	6.25%
Dolibarr	1	6.25%
ERP5	0	0.00%
Jasper BI	2	12.50%
Openbravo	1	6.25%
OpenERP	6	37.50%
Pentaho	3	18.75%
SugarCRM	3	18.75%
Tryton	1	6.25%

Support for programming language is a very strong for PHP, which is linked to the strong representation of Web activities, for Python, which benefits from an active community in Belgium, and for Java, which is widely used in the industry. The results for Web frameworks (Django, Ruby on Rails, Symfony, Zend, etc) are not represented because the companies rarely indicate the tools that they use on their websites.

Table 3. Dashboard showing programming languages supported by Belgian FLOSS providers (total: 54 suppliers).

Activities	#	%
Java	12	22.22%
Lisp	1	1.85%
Perl	3	5.56%
PHP	30	55.56%
Python	18	33.33%
Ruby	4	7.41%
Smalltalk	1	1.85%

The Drupal, Plone, Jahia, Joomla and Typo3 content management systems benefit from a strong commercial support. Drupal is widely acclaimed and has also been launched by a developer from Antwerp. Drupal could be considered as a second Belgian success story. An active community is developed around Plone in Belgium, especially in the public sector with the CommunesPlone / IMIO (www.imio.be) project. Jahia was used in several projects that were carried out in the public sector, such as the Agoracités electronic sales point project, which has since been replaced by CommunesPlone.

Table 4. Dashboard showing content management software supported by Belgian FLOSS providers (total: 37 suppliers).

Activities	#	%
Drupal	15	37.84%
EZ Publish	2	5.41%
Jahia	5	13.51%
Joomla	6	16.22%
Liferay	3	8.11%
Plone	8	21.62%
Spip	3	8.11%
Typo3	6	16.22%

4.3 Business Models

Almost all of the companies that were looked at offer services (see Table 5). The offer is not specialized in almost half of the cases and very specialized in nearly one in three cases.

Table 5. Belgian FLOSS providers' business models. (total: 31 providers)

Activities	#	%
Hardware	5	16.1%
Edition	3	9.7%
Services:	31	100%
No specialization	1	3.2%
Weak specialization	15	48.4%
Medium specialization	5	16.1%
Strong specialization	10	32.3%
Other products	0	0.0%
Hosting	2	12.9%
Contributions	4	12.9%

In comparison to the French market, the activities related to edition and hosting appear less frequently. This fact could be explained by the smaller size of the Belgian market, the importance of small companies in the local economy. As a consequence there is some difficulty in reaching a critical mass to develop a service activity and transform it into a software edition activity that requires more investment and greater specialization.

The companies that claim to contribute to FLOSS projects represent 12.9% of the total. This low result may be explained by our criteria that is based on the promotion on the websites of clearly identified contributions to open source projects. Indeed, the participation of companies in open source projects is common; however, the developer involvement is not always apparent [2, 10].

4.4 Economic Performances

FLOSS companies have a low failure rate, with only five out of 140 firms failing between 2004 and 2011 according to the information provided by the Belgian Central Balance Sheet Office.

The solvency is very good (37.1%; median). The capitalization is good (129,239 euros; median). However, these values should be subject to further, more detailed, study as they may be biased because of the weight of very large companies in Belgium that are strongly encouraged to capitalize through tax incentives such as notional interests.

The median turnover is 3.25 million euros. However, this high value is biased because turnover is only calculated from the companies that provide their accounting information in full format [1], whereas small unlisted companies can use the abbreviated format. The return on assets is good (6.8%; median).

The share of wages in value added is very high (79.7%; median). This value reflects a labor-intensive activity and, presumably, the orientation of Belgian FLOSS sector to provide services.

These partial figures reveal a sector that is rather stable and that generates employment. A closer examination of the annual reports of 10 companies in the sub-sector of ERP providers confirms this impression, with an average of 10 employees per company. In addition, 45% of companies employ 5 or more people (compared to 25% on average in the Walloon ICT sector). This review also highlights the existence of less known firms that reach a significant size compared to the local economy, and the importance of the self-employed people and very small businesses.

4.5 Relationship between Companies

A first analysis of the graph of relationships between websites (see Figure 2) identifies four types of key players: the Walloon centers of competence, the multinational IT companies, the FLOSS editors and the FLOSS business associations.

The Walloon centers of competence are centers for training, competitive intelligence and awareness. They often offer courses dedicated to FLOSS and focus the attention of the FLOSS industry. Several international IT companies have been investing in FLOSS software for several years (e.g. operating systems for servers, office suites, Web servers, application servers, etc.). In addition, they sell complementary products, proprietary software or hardware (e.g. servers), and have a well established network of partners. FLOSS editors are less common in Belgium but appear to be popular (see OpenERP).

Based on a Pagerank score that was calculated for all the nodes (i.e. websites) among the 50 most popular websites, the following websites stand out:

- the websites of the centers of competence and the research centers with training activity (Multitel, TechnofuturTIC, Technocité),

- the websites of Web 2.0 services (Delicious, Facebook, LinkedIn, Twitter, Youtube, Google Maps, Wikipedia),

- the websites of Web utilities (Adobe, Macromedia, W3C Validator) and Web references (IETF, W3C, Wikipedia),

- the websites of multinational ICT companies (Bull, HP, IBM, Microsoft, Novell),

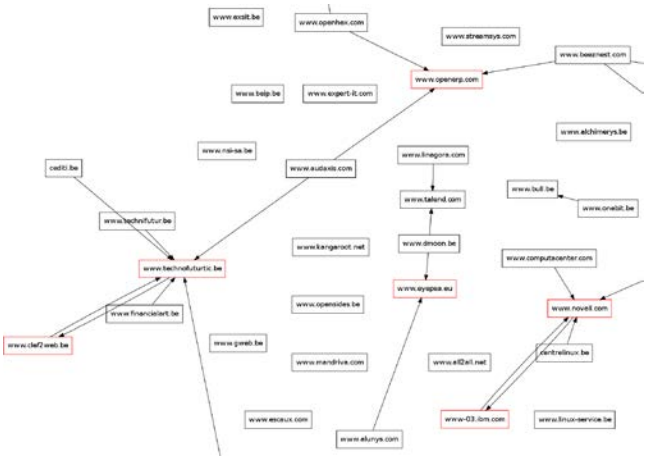
- the websites of popular FLOSS software (Drupal, JBoss, Joomla, Nagios, Worpress),

- the websites of public initiatives that support the Belgian ICT sector (eTIC, Infopole).

Based on a calculation of the Degree (i.e. DegreeIn and DegreeOut, that is to say inbound links and outbound links) calculated for all the nodes (i.e. websites), specialized companies and the Zea Partners association of FLOSS providers stand out. The centers of competence remain highly classified.

Based on a calculation of the Betweenness Centrality (centrality of a node in a network) which was calculated for all the nodes (i.e. websites), reference organizations, that is to say FLOSS associations, centers of competence, multinational ICT firms and FLOSS editors, all stand out.

Figure 2. Partial graph of relationships between Belgian FLOSS companies.



5. DISCUSSION AND PERSPECTIVES

These initial results on the study of Belgian FLOSS providers show a sector that is focused in the major economic centers of the country. The businesses are more active in the services and are heavily involved activities such as infrastructure software and Web development, activities which were common in the early years of free software development. The products on offer for business management software are, however, well developed.

The activities in open source geographic information systems are uncommon. However the strong development of OpenStreetMap (www.openstreetmap.org) and the recent public initiatives about open data in Wallonia -see for example the ICT Master Plan by Creative Wallonia (www.creativewallonia.be) or the Hackathon

eGov Wallonia platform (hackathonegovwallonia.net)- could stimulate this area of activity. The weak activities in the field of free software for embedded systems could be explained by the high number of self-employed specialists with poor commercial visibility.

The analysis of relationships between FLOSS companies revealed the importance of centers of competence. Their central role was not foreseen at the beginning of this study.

These preliminary results can be improved in different ways.

Firstly, work on the quantity and quality of data used for the analysis of the network of hyperlinks should be considered. On the one hand, the depth of the Web crawling should be increased. Indeed, Web crawling is now done with a depth level of 2. This involves the optimization of the software for links extraction, which is a time-consuming process. On the other hand, the standardization of hyperlinks should be improved in order to aggregate the links corresponding to a single firm or organization (e.g. linguistic or geographical sub-domains).

Secondly, more advanced techniques for graph analysis could be considered. At this stage, common metrics (Degree, Pagerank, Betweenness Centrality) were used to identify important nodes (i.e. websites) in the network of hyperlinks and a first visualization helped to understand the relationships between organizations. However, other tools such as clustering would allow to partially automate this analysis.

Thirdly, the annotation of business files requires a lot of manual work for encoding and updating (because the company's business may change over time). Automating the annotation could therefore be considered. A first test, that consisted of the extraction of terminology and named entities from a set of websites in English, gave encouraging results [17].

Finally, the business records contain the VAT number for each company. For this research, data provided by the Belgian Central Balance Sheet Office was used as well as balance sheets in PDF format that were analyzed for companies belonging to the same sector (ERP). However, there is a more ambitious approach to automate the analysis of the balance sheets in XBRL format that are provided by the Belgian Central Balance Sheet Office. XBRL (xbrl.org) is an XML-based data standard for the communication of data from companies. The analysis of these files would make it possible to analyze the sector and, especially, to benchmark companies or groups of companies in more detail.

This research may well lead to the creation of a methodology and a tool for the analysis of industrial sectors on the basis of data from different sources. The process would be based on an initial directory of companies and, in particular, would include a Web crawling system, an extraction tool able to (semi-)automatically update the directory (annotation), the automated creation of dashboards, the automated retrieval and analysis of XBRL files, the extraction of hyperlinks from collected pages, the calculation of metrics on the graph of relations between company websites (hyperlinks) and the visualization of that network.

6. REFERENCES

[1] BNB 2000. Dossier d'entreprise - Méthodologie et mode d'utilisation. Centrale des Bilans, mai 2000.

[2] Capra, E., Francalanci, C., Merlo, F. & Lamastra, C.R. 2009. A survey on firms' participation in open source community projects. Open Source Systems, vol. 299, Springer, Boston, MA, USA, pp. 225-236.

[3] CNLL 2013. Panorama de l'Open Source en France. Conseil National du Logiciel Libre, 24 janvier 2013.

[4] Dahlander, L. and Magnusson, M. 2008. How do firms make use of open source communities?. Long Range Planning, 2008, vol. 41, no 6, pp. 629-649.

[5] Daffara, C. 2007. Business models in FLOSS-based companies. Workshop presentation at the 3rd Conference on Open Source Systems (OSS 2007), 2007.

[6] Elie, F. 2006. Économie du logiciel libre. Eyrolles.

[7] Feller, J., Finnegan, P. and Hayes, J. 2006. Open source networks: an exploration of business model and agility issues. Proceedings of the 14th European Conference on Information Systems.

[8] Fitzgerald, B. 2006. The transformation of open source software. Mis Quarterly, 2006, pp. 587-598.

[9] Hansen, D., Shneiderman, B., and Smith, M.A. 2010. Analyzing social media networks with NodeXL: Insights from a connected world. Morgan Kaufmann.

[10] Lundell, B., Lings, B. and Lindqvist, E. 2010. Open Source in Swedish Companies: Where are We?. Information Systems Journal, Vol. 20, No. 6, pp. 519-535.

[11] Munga, N., Fogwill, T., and Williams, Q. 2009. The adoption of open source software in business models: a Red Hat and IBM case study. Proceedings of the 2009 Annual Research Conference of the South African Institute of Computer Scientists and Information Technologists. ACM, 2009. p. 112-121.

[12] Ortega, J.L. and Aguillo, I.F. 2008. Visualization of the Nordic academic web: Link analysis using social network tools. Information Processing & Management, vol. 44, no 4, p. 1624-1633.

[13] Välimaki, M. 2003. Dual licensing in open source software industry. Systemes d'Information et Management, vol. 8, n° 1, p. 63-75.

[14] Vasquez Bronfman, S., Miralles, F. 2007. Business Models in Open Source Software: do they exist?. Actes du 12e Colloque de l'AIM, HEC Lausanne, 18-19 juin 2007.

[15] Viseur, R. 2011. Cartographie des marchés Open Source belges et français. Rencontres Mondiales du Logiciel Libre. Université de Strasbourg, Strasbourg, 11 juillet 2011.

[16] Viseur, R. 2013. Mapping of Belgian open source market. Rencontres Mondiales du Logiciel Libre. Université Libre de Bruxelles, Bruxelles, 11 juillet 2013.

[17] Viseur, R. 2014. Automating the Shaping of Metadata Extracted from a Company Website with Open Source Tools. International Journal of Advanced Computer Science and Applications (IJACSA), Special Issue on Natural Language Processing 2014, pp. 31-35.