# *Evaluating the progress towards a more semantic web and the database use at University library Belgrade*

*Adam Sofronijevic, PhDc, department head, University library "Svetozar Markovic" Belgrade, sofronijevic@unilib.bg.ac.rs*

*The paper deals with two important issues of contemporary librarianship: evaluation of database use and creation of linked data. Evaluation of database use is presented in the broader context of the library management that seeks ways to provide for the rational distribution of limited resources. Experiences and results from University library Belgrade and other Serbian libraries are presented. Importance of linked data is presented along with basic concepts important for understanding semantic web creation. Advanced nano-publication concept is explained along with the everyday practice of metadata creation at University library Belgrade. Paper advocates implementation of various possibilities that linked data concept provides for libraries in general.*

**Keywords**: Evaluation of database use, Linked data, University library Belgrade, Nano-publication, Library management, Semantic web

## Introduction

Metadata or data about data are at the core of the librarianship and are still considered to be a prime product of librarianship practice (Coyle, 2005). In order to keep, preserve and organize data librarians need to describe them. Lately, financial pressures and expectations of stakeholders add a new role for metadata provided by libraries – visibility. In order to present results of academic research or promote national scientific and cultural heritage libraries need to provide metadata that allow for materials described to be effectively reachable by users and not buried somewhere in the depths of Internet. Linked data provide such visibility. Therefore the importance of creation of linked data in contemporary libraries and the need for libraries not to fell out of emerging semantic web. Various approaches are existing in contemporary libraries towards linked data. Extensible Markup Language (XML) based tools are dominant as well as metadata shemas providing compatibility with large international portals for digital documents.

In order to stay at top of user needs and expectation contemporary librarianship needs change and new service development. Evaluation of existing services should ground basis for effective and efficient changes needed in time of restricted financing. Especially important is the evaluation of use of costly resources such as databases. Database use evaluation results are usually a part of larger evaluation process that aims for more efficient distribution of limited resources across the library services. Therefore the need to evaluate use of databases in a broader context of holistic approach to library services.

## University library Belgrade and Kobson consortia

In order to present the evaluation of database use at University library Belgrade we will first describe the framework of acquisition of these databases that also define their use. University library Belgrade is the leading academic library in Serbia and founder and an important member of Serbian library consortia Kobson (Serbian Library Consortium for Coordinated Acquisition) that provides for joint acquisition of electronic journal databases for all academic libraries in Serbia. Since 2001 Kobson provides electronic scientific information for scientific community of Serbia, including electronic journals, proceedings and books. These materials are available for users of Academic Network of Serbia, a designated computer network for academic community of Serbia that spans all institutional facilities across the country. For university staff members access to materials acquired by Kobson is provided also from outside this network – remotely by EzProxy. Most of the use is generated in this way, by researchers searching and browsing databases of electronic journals and downloading papers from their homes or even while they are on the go or at a holiday, away from their usual work environment. Since EzProxy is used for most of the off-site traffic and general proxy for accessing the databases from Academic network allowing for uniform representation of Serbian users from publisher side, it is very difficult, if not impossible to provide use statistics of separate institutions. Having in mind a more broad picture of library services in which use is directly related to promotional and educational activities, which are also provided by joint effort of consortia, we will present result of general use of Kobson resources with some additional selective statistics illustrating specific situations of use.

**Database use**

Use of database materials in Serbian libraries, including University library Belgrade covers the downloading of files, mostly PDFs, containing single journal articles or book chapters. In 2012 users had access to databases that comprise 35 000 titles of e-journals, and more than 60.000 of e-books. Thomson Reuters ranked journals from the database World of Science are pretty good covered, ranging from as high as 85% in humanities to as low as 65% in agricultural science (Timotijevic & Kosanovic, 2013). One important statistics is number of unique visitors to Kobson Internet site, which is a starting point in a researchers quest for relevant scientific materials for researchers in Serbia. This number has been around five thousand per day, with variations throughout the year, depending on the month – for there is less use during summer months, with constant slow increase in last three years. Explanation is that in the early years of consortia electronic resources were new to everyone in scientific community of Serbia and after several years of sharp rise in use, the rise slowed down to almost a constant because most of those interested in these resources are already using them and new use is generated by young researchers entering their respective fields. In Table 1. we will provide some statistics based on publicly available data at Kobson Internet site (http://kobson.nb.rs/statistike/statistike.126.html) on aggregate downloads both on site – in libraries and other university facilities and off site. One can notice that majority of downloads are from one service and it has to be kept in mind when evaluating the price for this service. Also it can be noticed that

monthly use is not equal throughout the year, with less downloads during summer month.

| | Science Direct | JSTOR | EBSCO | Springer-link | Oxford Journals |
|---|---|---|---|---|---|
| January | 82213 | 10963 | 8569 | 15471 | 2669 |
| February | 88510 | 10016 | 9689 | 15835 | 2811 |
| March | 101271 | 10336 | 12664 | 17000 | 3169 |
| April | 84016 | 9307 | 11886 | 14235 | 2825 |
| May | 81774 | 9604 | 13078 | 14973 | 2772 |
| Jun | 74304 | 8557 | 9264 | 11976 | 2499 |
| July | 74560 | 7316 | 9064 | 10848 | 2494 |
| August | 61363 | 6492 | 7018 | 10444 | 2078 |
| September | 69658 | 7304 | 7713 | 11111 | 2597 |
| October | 89053 | 8257 | 10783 | 14595 | 3309 |
| November | 90073 | 10908 | 12007 | 13862 | 3150 |
| December | 85096 | 10507 | 12809 | 13811 | 2752 |
| 2012 | 981891 | 109567 | 124544 | 164161 | 33125 |

Table 1. Kobson use by month for 5 services in 2012

Another revealing statistics on Kobson databases use comes from the practice of University library Belgrade. Use of Kobson materials from the comfort of home has become norm for Serbian researchers. Due to administrative and technical problems PhD candidates and other students do not have this privilege and have to come to library or university buildings to acquire Kobson materials. Because of this University library launched "Kobson at home" a pilot service in 2010. This service provided PhD candidates at one of the Faculties of University of Belgrade with possibility to request a paper via email and receive a full text within hours. Authentication of users was a major concern and because of the procedure it required a lot of effort was put into the promotion of this service and educational sessions for this particular user group. The results of after three years are revealing. The number of PhD candidates over the period grew because of the reforms in higher education system, so the use also grow. But the growth of use was linear compared to increase in number of potential users. Over the three year period proportion of users remained approximately the same, around 10%. This might be surprising but one has to bear in mind the specifics of PhD studies. Also number of papers downloaded by one person per year proved to be approximately constant at around 50. Again one needs to bear in mind the humanistic field of studies where monographs, are still valuable and play a major part of literature besides journal papers.

If one is to consider comparison of database use in different countries based on public, freely available evaluation data the task turned out to be impossible. There are almost no such data available, one of the rare examples comes from Russia. Some use data can be found at www.neicon.ru/statistic but the language of the presentation and it's technical basis limits the effective use of this data for comparison purposes. General problem for availability of such data is certainly in the negotiating process between vendors - publishers and librarians or library consortia which involves these data as an important pricing tool. Also many national consortia similar to Serbian Kobson that do publish such data do it in their local languages only, since information on database use in this context is for local stakeholders and not for international audience. A useful tool that might substitute the lack of publicly available data is COUNTER (Counting Online Usage of Networked electronic Resources) codes of practice available at www.projectcounter.org. As stated at the website COUNTER is an international initiative to improve the reliability of online usage statistics. It is supported by the vendor, intermediary and librarian communities. COUNTER's objective is to ensure that vendor online usage reports are credible, compatible and consistent. This is achieved by the publication of Codes of Practice that specify the content, format, delivery mechanisms and data processing rules for a set of core usage reports that are easily implemented by vendors and easily understood by librarians.

A big task of evaluating database use rests with librarians and specifics that we briefly mentioned makes this task even more difficult in spite of different technological advantages available in digital era. Only by permanently educating themselves in various areas such as management, statistics, IT, etc. librarians may prove worthy of such a task.

**Linked data at University library Belgrade**

Overall situation in regards to metadata usage in Serbia today can be described as sketchy with different types of metadata used and a range of various purposes and ideas for metadata usage and even more ideas and initiatives for planning of future use of metadata.

Important trend in creation of metadata describing digitized materials in Serbia is usage of internationally accepted metadata models driven by institutions' participation in international digitization and aggregation projects. Most of contemporary digitization initiatives in Europe are in relations with Europeana portal. Almost all of European Union funded multinational digitization initiatives are regarding this portal. Hence the importance of metadata model used in Europeana for all European libraries (Antonic, Mitrovic & Sofronijevic, 2011). In Serbia in particular projects that have the aim of providing metadata for Europeana are especially potent drivers for creation of metadata that adhere to ESE – Europeana semantic elements and EDM – Europeana data model, metadata models used by Euorpeana (Europeana Semantic Elements Specification, 2011), (Europeana Data Model primer, 2010). During 2011 and 2012 University library in Belgrade was a partner in Europeana libraries project that brought digital materials from European academic libraries to Europeana. Thanks

to this project University library got experience in creation of metadata in ESE metadata model and aggregation scheme of Europeana. Also, in January 2012 new metadata model was implemented – EDM (Europeana Data Model) (Sofronijevic, 2011). As of May 2013 University library in Belgrade is a partner in Europeana Newspapers project that aims to add newspaper materials to Europeana. In addition, the project addresses challenges particularly linked with digitized newspapers: use of refinement methods for OCR, OLR/article segmentation, and named entity recognition (NER), and page class recognition to enhance search and presentation functionalities for Europeana customers; quality evaluation for automatic refinement technologies; transformation of local metadata to the Europeana Data Model (EDM); metadata standardization in close collaboration with stakeholders from the public and private sector. Being a hub of digitization activities in Europe Europeana is dynamicly developing and along with it its metadata model is changing in order to facilitate more rich structure and to be more user oriented. Also semantic elements are important for enriching user experience of Europeana and because of this EDM metadata model is primarily being introduced (Doerr, Gradmann, Hennicke, Isaac, Meghini & Herbert van de Sompel, 2010). Its development and implementation is proved to be hard for some librarians since many ideas and concept contains have originated in computer sciences (Definition of the Europeana Data Model elements, 2011). Because of this it is important for libraries in Serbia to be in front rows of EDM implementation and that is what only international projects can provide for them.

Emergence of the semantic web is perceived as an opportunity and a challenge in Serbian libraries. Implementation of technologies such as Resource Description Framework (RDF), Web Ontology Language (OWL), Simple Knowledge Organization System (SKOS) and Protocol and RDF Query Language (SPARQL) is difficult with scares resources and librarians need new skills and additional education in order to implement them. Because of this debate is open between management of big Serbian libraries on means to gain necessary resources and educate stuff in order for Serbian libraries to hop on a fast moving train of semantic web. Plans for employment of a more ICT oriented stuff have been made and new curricula in LIS education and long life learning programs for employees are discussed. In parallel monitoring of new developments in related areas is intensive and workshops, lectures and experience exchange events are organized with participants coming from foreign libraries that are more advanced in this field. As of May 2013 collaborative effort is planed to lunch a working group consisting of metadata, ICT and library management experts that will propose project ideas, guidelines and other support for implementation of RDF, OWL, SKOS, SPARQL and other tools and technologies that ground basis for efficient creation of metadata compliant with semantic web. Exiting experiences in usage of Extensible Markup Language (XML) for creation of metadata is an important starting point and therefore Serbian libraries participating in EU related projects that have experience with XML based metadata schema such as EDM will lead this initiative. First results in the form of proposals and national guidelines are expected by the end of 2013.

In order to prepare organizationally for this change University library Belgrade put in motion a plan of organizational structure transformation. Existing cataloguing department has been downsized and librarians from it transferred "on the loan" to the department for digitization in order to work closely with more ICT oriented colleagues on creation of metadata for digital objects. The plan is to expand this subdivision of digitization department and to finally form a full department for digital metadata creation that will deal with most of the perceived cataloguing work in the future that will encompass creation of metadata for digital objects coming from retrospective scanning, new born digital objects, Internet archiving materials, etc. Building expertise and skill in presenting metadata on these objects based on technologies and with the tools that allow for linked data creation University library Belgrade hope to add a small but important part to the emerging semantic web.

In the effort to implement linked data and new forms of scientific publishing plans have been made to implement nano-publication format at University library Belgrade and across faculties of University of Belgrade. Nano-publicaton is a very short declaration connecting two concepts by means of a third and providing metadata about this relation (conditions under which relation is viable, author, timestamp, etc.). Nano-publication is a concept in making that appeared as a response to growing inadequacy of various aspects of existing tools for scientific communication. Originating in life sciences nano-publications seem to be envisioned and increasingly shaped as a tool for efficient publishing of datasets. Nano-publication concept has potentials to successfully face the challenge to provide a novel method of evaluating datasets and scientific work based on them while at the same time preserving the values of traditional means of scientific communication (Mons et al., 2011). In order to define nano-publications clearly an example available at (Nanopublication Downloads section, 2012) is provided, with just most important parts presented for sake of clarity.

```
<nanopublication id="0">
<assertion>
<subject>NG_000007.3:g.70628G>A</subject>
<predicate>has variant frequency</predicate>
<object>0.25%</object>
</assertion>
<condition>Sardinian</condition>
<provenance>
<dateofcreation>March 24, 2011</dateofcreation>
…
<evidenceType>empirical</evidenceType>
<authorID>Giardine et. al.</authorID>
<curatorID>unresolved</curatorID>
…
</provenance>
</nanopublication>
```

Nano-publicaton is identified uniquely by its id. Assertion part of the nano-publication offers relation of concepts and is followed by the condition under which the assertion holds and by metadata describing the context of the relation. Nano-publication is presented by means of XML which enables machine readability.

Nano-publications can presented in more detail by a description of steps necessary to perform in order to create a nano-publication as described in (Mons & Velterop, 2009).

Nano-publications have the potential to be an important driver of scientific research in first world countries which are ample in resources that allow for production of datasets in areas like life sciences. Could nano-publications also play a significant role in fostering communication and usage of OA resources in scientific communities of developing and transitional countries? In general these countries lack the resources and the infrastructure for conducting research that produce large datasets. Consequently entire scientific disciplines in these countries are left behind or become obsolete. In order to prevent this scientific community of these countries should be focusing their efforts on usage of freely available scientifically relevant datasets. This was not so in the past because datasets were not freely exposed. All of this may change with the rise of nano-publications.

**Conclusions**

If libraries are to provide efficient services for their patrons they have to be up-to-date with evolving technologies, but more importantly to have a vision of user needs and expectations and provide efficient tools for fulfilling them. Evaluation of services, with focus on the expensive ones like databases and creation of linked data are the means to achieve this in general, but only by staying on top of the changes, technological and sociological ones among many, libraries can survive and thrive.

## Literature


1. Antonic, S., Mitrovic, J., & Sofronijevic, A. (2011). Fostering Open Access usage by creation of the library aggregator for Europeana: project Europeana libraries. INFORUM 2011: 17th Conference on Professional Information Resources. Prague, May 24-26. Retrieved from http://www.inforum.cz/pdf/2011/antonic-sanja.pdf.
2. Coyle, K., Baker, T. (2009). *Guidelines for Dublin Core application profiles*. Retrieved on May 5th 2013 from http://dublincore.org/documents/profile-guidelines/
3. Europeana Data Model primer. (2010). Retrieved from http://group.europeana.eu/c/document_library/get_file?uuid=718a3828-6468-4e94-a9e7-7945c55eec65&groupId=10605.
4. Europeana Semantic Elements Specification: version 3.4. 2011. (2011). Retrieved from http://version1.europeana.eu/c/document_library/get_file?uuid=77376831-67cf-4cff-a7a2-7718388eec1d&groupId=10128.
5. Mons, B. et al. (2011). The value of data, Nature Genetics, 43, 281-283. Retrieved February 15, 2012, from http://www.nature.com/ng/journal/v43/n4/full/ng0411-281.html

6. Mons, B., & Velterop, J. (2009). Nano-Publication in the e-Science Era, Workshop on Semantic Web Applications in Scientific Discourse (SWASD 2009)
7. Nanopublication Downloads section, Human Hemoglobin Genetic Variation. (2012). Retrieved April 10, 2012, from http://www.nanopub.org/XML/xml_ethnicity.txt
8. Sofronijevic, A. (2011). Biblioteke Europeane: 24 evropske biblioteke i Univerzitetska biblioteka "Svetozar Marković" stvaraju bibliotečkiagregator za Europeanu. Visokoškolske biblioteke, 1. Retrieved from http://www.zbus.rs/cir/index.php?a=visokoskolske&b=cirilica&c=osam&d=jedan#2.
9. Timotijevic, T. & Kosanovic, B. (2013). m.KoBSON: A step closer to the new technologies, in Proceedings of 19th Annual Conference on Professional Information Resources, May 21-22, Prague.