Polythematic Structured Subject Heading System & Creative Commons

KAMRÁDKOVÁ, Kateřina
National Technical Library
katerina.kamradkova@techlib.cz

KOŽUCHOVÁ, Kristýna
National Technical Library
kristyna.kozuchova@techlib.cz

MYNARZ, Jindřich
National Technical Library
jindrich.mynarz@techlib.cz

October 15, 2009

Abstract

This article presents how knowledge organisation systems in general and Polythematic Structured Subject Heading System in particular can be used in context of grey literature resources. Starting from improved accessibility and findability, employing knowledge organisation systems can benefit both authors of grey literature and users. It suggests some of the functions knowledge organisation systems can fulfill in the grey literature environment, such as metadata authoring, search query refinement, or alternative ways of navigation. Polythematic Structured Subject Heading System (PSH) is an example of knowledge organisation system that was published under the Creative Commons licencing terms in the form of *linked data*. The article discusses some of the practical applications of PSH with hands-on examples. Second part of this text focuses on Creative Commons (CC) licences which provide publishing protocol particularly suited for the field of grey literature. CC offers an easy way to publish grey literature in such manner that it is more accessible and findable without any help from commercial publishers.

PSH terms

knowledge systems, subject information languages, grey literature, intellectual property law, licences

The focus point of this article is Polythematic Structured Subject Heading System (PSH), a subject heading system produced by the National Technical Library of the Czech Republic. As an example of knowledge organisation system, it has many useful applications in the field of grey literature. Within this context it can help both authors (publishers) and users (consumers) of grey literature resources. PSH has been published under Creative Commons licencing terms and it conforms with the *linked data* principles. These features, which are described in detail in the following text, greatly extend the range of possible uses of PSH, most of which are applicable in the area of grey literature that frames this article's scope.

1 Context of grey literature

According to the generally accepted definition, grey literature is „*the information produced on all levels of government, academics, business and industry in electronic and print formats not controlled by commercial publishing i.e. where publishing is not the primary activity of the producing body*" (PERSPECTIVES, 1997).[1] If some resource is labeled "grey", it does not say anything about its content. „*Grey does not imply any qualification, it is merely a charac-*

---

[1] http://www.textrelease.com/textreleasehome/pressreleases.html

terisation of the distribution mode" (MACKENZIE OWEN, 1998). It is common though that the content of grey literature has some distinctive characteristics. It is by far the most immediate channel for new information stemming from the research, government or educational institutions. Also, it tends to be less permanent because it escapes the established means to preserve information (e.g. libraries).

Grey literature is an important source of information and it constitutes an alternative to the conventional white literature. It includes all types of literature that are not available through the traditional publishing channels. Especially the Internet has enabled dissemination of these publications that have not been formally published to a much wider audience.

As the definition of grey literature stands, it lacks "commercial control". Often it turns out that this feature also implies a lack of "bibliographic control". Frequently it is up to the creators of grey literature themselves to provide bibliographic information for their publication so that it can be indexed and found. *"And because there are no commercial issues at stake, improved referencing will always be dependent on the initiatives and willingness of producer or distributor organisations and a few committed professionals"* (SCHÖPFEL, 2006). From this issue arises the challenge of the quality of indexing and subject access. *"Self-posting is perhaps the most used method to disseminate information but it is not effective because of a lack of adequate indexing"* (DE BLAAIJ, 2005). Even the content of grey literature may *"represent topics that are too new or different to easily provide subject access for"* (CHILDRESS, 2003).

2 Role of knowledge organisation systems

Institutional repositories or digital libraries where grey literature is often kept can alleviate specific issues of grey literature by the use of a knowledge organisation system (KOS), such as controlled vocabulary, thesaurus or subject classification. Employing knowledge organisation system has many advantages both for authors (publishers) and users. They act as metadata standards which ensure consistency of subject indexing that is often lacking in subject organisation based on free keywords. Knowledge organisation system *"attempts to increase consistency and improve access to digital collections and web navigation systems via vocabulary control"* (TUDHOPE, 2006). The general purpose of vocabulary control is to reduce the ambiguity of natural, uncontrolled language.

Functions addressing this purpose are grouped under the term terminology services. These include *"resolving search terms to controlled vocabulary, disambiguation services, offering browsing access, offering mapping between vocabularies, query expansion, query reformulation, combined search and browsing"* (TUDHOPE, 2006). The reduction of vagueness of natural language can be achieved by knowledge organisation system that defines the scope of a term, provides a set of synonyms or restricts the scope of a term (TUDHOPE, 2006). Free keywords are not present with these characteristics and thus searching with them might not yield relevant results. As Tudhope puts it, *"when searching free text with uncontrolled terms, significant differences can stem from trivial variations in search statements and from differing conceptualisations of an information need"* (TUDHOPE, 2006). Using controlled vocabulary provides a way to eliminate these variations by adhering to the terms representing precisely defined concepts which increases both precision and recall of the search query. KOS performs as a common ground which *"provides potential pathways (for human and machine) that connect a searcher and indexer's choice of terminology"* (TUDHOPE, 2006).

Metadata drawn from KOS can increase findability of grey resources because they can be crawled and indexed by harversters or semantic search engines (e.g., Sindice[2]).

Grey literature is often characterised by its limited availability. Also, the question of preservation is frequently left unresolved. These issues can be alleviated by the use of metadata, which can *"provide information required for permanent access to the dig-*

---

[2]http://sindice.com/

ital object." (HOORN, 2006). This implies the need of metadata standardization in the areas such as *„persistent identifiers, representation of vocabularies, protocols for programmatic access "* (TUDHOPE, 2006). Interoperability of KOS depends on a global identifier mechanism for referring to vocabulary concepts. From the proposed identifiers currently *„only "http" URIs offer a simple, widely deployed dereferencing mechanism "* (TUDHOPE, 2006). In this way, the necessary infrastructure is created upon which further services can be built.

2.1 Assisting authors

Employing KOS can provide advantages both for authors of grey literature and users. From the authors' point knowledge organisation systems can facilitate metadata authoring. Because of the self-published nature of grey literature it is often the case that the document metadata is made by authors themselves rather than by library professionals. In order to achieve reasonable indexing quality tools for metadata authoring can be utilised. The barrier of complicated metadata formats can be lowered by tools generating *metadata snippets* (short fragments of metadata) that can be readily embedded in documents (web pages, PDFs).

The process of subject terms assignment is considered difficult. The first step to make this easier is to simplify the selection of suitable indexing terms. Indexing system should have an easy-to-use interface that allows intuitive navigation through the knowledge organisation system (e.g., using visualization techniques), preferably with some helper tools for spelling correction (cf. "grey literature" and "gray literature"[3]) or term suggestion.

Next step further is automatic indexing that facilitates automatic subject term assignment according to the information content of the indexed document. It involves techniques of natural language analysis, subject analysis, machine learning processes, or simple phrase matching. The example of automatic indexing is term extraction technique aiming to extract from the document the terms and phrases that

make the most accurate representation.

This approach is used in integrated digital library system CDS Invenio which will be used for the National Technical Library's *National Repository of Grey Literature project*[4]. CDS Invenio *BibClassify module*[5] accepts controlled vocabularies in Simple Knowledge Organisation System (SKOS) format. It analyzes the document provided by the user and then it produces a list of suggested terms from the supplied knowledge organisation system recommended to be used in the document metadata. However, it does not use any machine learning or artificial intelligence techniques and therefore it produces only approximate results.

For example, for this article the Bibclassify module extracted from PSH these terms (followed by the scores reached): *grey literature* (40), *science* (31), *www* (24), *licences* (21), *plant developmental stages* (20), *author* (19), *work* (18), *data theory* (17), *documents* (15), *copyright* (12), *resources* (11).

An example of a more advanced automated indexing application is *Maui Indexer*[6]. It also accepts knowledge organisation systems in SKOS format from which the terms are assigned, but unlike BibClassify module it uses machine learning algorithms. Given a sufficiently large set of documents with manually assigned indexing terms from which it can "learn" (i.e. build a model), its indexing performance can be very high[7]. It also provides other features than automatic term assignment (e.g., automatic tagging, topic indexing with Wikipedia).

Both these automatic indexing applications build upon well-structured SKOS format which is essential for their operation because it provides widely accepted formalization of KOS that serve as a standard. It is an example how semantic web technology can improve on traditional ways of KOS distribution. As Vlachidis puts it, *„today available semantic technologies promise to close the gap between*

---

[3]http://psh.ntkcz.cz/skos/PSH6454/html/en

[4]http://nusl.techlib.cz
[5]http://invenio-demo.cern.ch/help/admin/bibclassify-admin-guide
[6]http://code.google.com/p/maui-indexer/
[7]This can be illustrated by an AGROVOC test case – http://code.google.com/p/maui-indexer/wiki/Examples#Assignment_of_Agrovoc_terms_to_agricultural_documents

formal knowledge structures and textual representations enabling new access methods to information " (VLACHIDIS, 2009).

2.2 Enhancing user interface

Taking the perspective of the user, knowledge organisation systems enhance the information retrieval system's user interface. The improvements can be made to the search process by search query refinement. The interface may provide the user with suggested spelling corrections or offer auto-completion of entered search terms. The more sophisticated application of KOS is broadening the search using techniques of semantic query expansion.

Knowledge organisation system can be used as a tool for hierarchical navigation and other alternative ways of browsing. Building on the KOS's rich relationships between its terms user can be presented with contextual information, such as related documents. Some of the contemporary digital library software (e.g., JeromeDL[8]) makes it possible for users to assign social tags to documents based on underlying subject organisation. Tags constitute a custom-made access points to the documents in the digital library collection. If KOS is used, tags can be assigned in consistent manner which aligns with user's expectations.

2.3 Linked data

An emerging trend can be observed in the area of knowledge organisation systems as they get involved in building the so-called "semantic web". This leads to the *linked data* paradigm that represents the sum of best practices for exposing, sharing and connecting semantic web data.[9] Linked data is a way of publishing documents and data as mutually interlinked web resources. It utilizes semantic web technologies, mainly Resource Description Framework (RDF)[10] data model and Uniform Resource Identifiers (URIs)[11] as a primary way of resource identification. The way linked data is published is guided by four principles formulated by Tim Berners-Lee (BERNERS-LEE, 2006). They recommend using URIs as resource identifiers, particularly HTTP URIs, so people can look them up using a web browser. Linked data should also provide useful information about described things and links to other resources.

Knowledge organisation systems form an important part of web of linked data. They constitute a layer of interoperability between disparate datasets and can serve as hubs interlinking linked data resources. Rich hyperlinks add to the value of connected resources and increase their accessibility which is particularly important for grey literature documents. Tim Berners-Lee suggests that *„the value of your own information is very much a function of what it links to "* (BERNERS-LEE, 2006). Knowledge organisation systems are also used for annotating resources. This can take the form of semantic tagging, using established formats such as Dublin Core[12], Meaning-of-a-Tag (MOAT)[13] or Common-Tag[14]. Tagging is often viewed as putting the web resource into context (WILDE, 2008). Tags join related information and provide links users can follow to broaden the scope of their search for information.

2.4 Simple Knowledge Organisation System

In order to take advantage of the afore-mentioned applications of knowledge organisation systems they should be published as linked data as well. That means converting the KOS to some RDF format. There seems to be a growing consensus in adoption of Simple Knowledge Organisation System (SKOS)[15] which is specifically designed to represent traditional knowledge organisation systems. It provides means for expressing standard relationships that can be found in most thesauri or subject heading systems. SKOS is defined using RDF Schema and Web Ontology Language (OWL) and is expressed in RDF data model. In August 2009 it has gained

---

[8]http://www.jeromedl.org/
[9]http://linkeddata.org/
[10]http://www.w3.org/TR/rdf-syntax-grammar/
[11]http://www.w3.org/Provider/Style/URI

[12]http://dublincore.org/
[13]http://moat-project.org/
[14]http://www.commontag.org/Home
[15]http://www.w3.org/2004/02/skos/

the status of World Wide Web Consortium's (W3C) working recommendation and it is already widely used in many influential libraries (e.g., the Library of Congress). In June 2009 the list of SKOS implementations was joined by the National Technical Library (NTL[16]). Its product Polythematic Structured Subject Heading System (PSH) was published in SKOS format under Creative Commons (CC) licencing terms.

3 Polythematic Structured Subject Heading System

Polythematic Structured Subject Heading System (PSH) is a set of subject headings which is used to describe and search the document by subject. PSH in its latest version 2.1 is bilingual (Czech–English) and the subject headings in both languages are interconnected. It contains over 13500 subject headings and it is divided into 44 thematic sections.

Subject headings are included in a hierarchy of six (or seven) levels according to their semantic content and specificity. Nearly none of mentioned thematic sections do not significantly exceed the others in number of its subject headings. The "depth" of each section is approximately the same.

The whole system is a tree structure and it represents various concepts from the most general to the more specific ones. There are hierarchical, associative ("see also") and equivalence ("see") relations in PSH. The set of subject headings has qualifiers appended. The qualifier has a two-letter-abbreviation form, which represents the belonging to one of the thematic sections.

In comparison with uncontrolled terms and keywords which are created freely, PSH is a controlled vocabulary which is based on binding principles similar to thesaurus construction. Despite this, PSH is a user friendly tool when indexing or searching for documents and information. It stems from the natural language and respects natural word order, and therefore it makes searching and indexing easier. Indeed, „controlled list of terms ensures consistency

in searching and indexing, helping to reduce problems arising from synonym and homograph mismateches" (TUDHOPE, 2006).

For indexers and searchers, PSH is a listing of terms authorized for use in an indexing system, together with relationships, variants and synonyms, and aids to hierarchical navigation through the subject heading system. It serves as a retrieval tool as well. The bilingualism of PSH guarantee its accessibility on international level. Thanks to the interconnection of each subject heading, PSH can be used to translate terms from Czech to English and vice versa.

Traditionally, the subject indexing of textual documents has been the responsibility of professional librarians and indexers. One of the disadvantages of expert indexing is that it is expensive and time consuming. That is the reason why some university libraries require documents to be indexed by authors (students). In the Czech Republic, subject indexing process of theses and dissertations is presently disunited. It is still mostly an intellectual process, where an author-based indexing is founded on the assumption that author´s interpretation of his/her own work is the authoritative view. To the contrary, the indexer must apply terms which appropriately identify the subject either by extracting words directly from the document or assigning words from a controlled vocabulary.

Among the benefits of using subject heading systems when indexing grey literature (e.g., theses and dissertations) belong the lucidity, easy hierarchical-structure orientation and manipulation, etc. This set of agreed-on terms represents specific concepts, and thereby reduces the number of possible wordings for meaning.

PSH has an intuitive browsing tool which is easy to handle for everyone. By the visualisation of PSH, displaying the hierarchical navigation and relations between terms, could be searching and indexing more interesting and easier. At the same time it is possible to display tags from PSH (metadata snippets - Dublin Core and CommonTag) that can be embedded in an HTML document its semantic description in a machine-readable way. Almost anyone authoring web content (e.g., bloggers) may found a use for it.

---

[16]http://www.techlib.cz/en

PSH can be used as a separate tool which authors refer to in deciding how to tag their documents while indexing. PSH can be used in authomatic indexing systems to generate a list of suggested terms as tags representing key concepts of documents. There is also a possibility of refining user´s queries for retrieval according to PSH.

The value of PSH in indexing and retrieval systems (particularly in the OPAC and other information and retrieval systems) is well recognized. It can serve for example to correct spelling errors and to provide:

◇ means by that the use of terms may be standardized,

◇ controlled hierarchies so that a search can be narrowed or broadened systematically,

◇ a guide for users of the system for choosing the correct term for a subject search,

◇ a navigation tool through the set of subject headings and their relations.

PSH is produced by the National Technical Library in Prague (NTL), which guarantee its quality. In the context of the PSH development process the Committee for Coordination of Polythematic Structured Subject Heading System was established. The Committee consists of external and NTL experts and group of representatives from libraries participating in usage of PSH. At the end of the year the Committee members gather for annual meeting to discuss and agree on proposals for improving PSH. Committee members can also suggest new ideas concerning recommendations focused on further development directions and possibilites of PSH. The last committee meeting took place on the 10th of Decemeber 2009. Among other things, there was a disccussion regarding feasibility of offering PSH under one of the CC licences, which was eventually realized.

4   Creative Commons licences

Creative Commons licences are a product of Creative Commons Initiative, an American non-profit organization founded in 2001. CC licences origin is connected with the problem of protecting but at the same time using work. With the development of the Internet and particulary with the newest communication technologies commonly labeled as Web 2.0, we have been witnessing the rapid growth of "participatory media", whereby both proffesionals and amateur authors create and share digital content with greater ease than before (CHELIOTIS, 2007). CC licences are currently used especially for electronic resources accessible through Internet network. „The Creative Commons licenses is a perfect example of the sort of copyright changes the modern world needs to come to grips with in the digital age, information should be free to all." - John Harvey (Flickr photographer) (CASE, 2009b).

Creative Commons' main goal is, in particular, enable legal using and sharing of publicly avaliable author work. In response to a various requests by a broad community of authors CC has introduced a number of licence types to suit different needs (CHELIOTIS, 2007). CC defines the spectrum of possibilities between full copyright and public domain. Authors can choose one of the CC licences thereby provide opportunity to share work with others. If author choose one of those CC licences he will actually give up some rights and allow using his work within the scope of the chosen CC licence. CC licences are not an alternative to copyright, they work alongside it (CREATIVE, 2001). „It takes the concept of "all rights reserved" and splices it into manageable parts in order to create a "some rights reserved" system." (FILTER, 2009). However, CC licences must be understood in the context of the existing copyright regime (CHELIOTIS, 2007).

Creative Commons Initiative provides a set of licences composed of requirements and permissions represented by four basic licencing elements - requirement for attribution, sharing derived works under same or similar licence, prohibition of commercial use, and restriction against creating derivative works. In turn for adhering with these requirements users may be permitted to reproduce, distribute, share the licenced work and also make derivative works. Special conditions can be specified for use in developing countries.

The licencing terms are expressed using Creative

Commons Rights Expression Language (ccREL)[17] which is RDF-based and can be serialized to fit in various document formats (HTML, PDF). The most common is the use of RDFa (RDF in attributes)[18] form that can be easily included in a web page. Metadata describing CC licencing terms serves as an explicit connection between the document and the abbreviated version of licence. In this way, the document is clearly marked so that users know what they are allowed or restricted to do with it. Formalized metadata also enables indexing by search engines which in turn increases visibility of the licenced work.

Creative Commons presents publishing protocol that can greatly benefit grey literature authors. Their main motivation often is to share and improve access to their knowledge or results of their work, while not being conditioned by the expectation of a financial reward. Instead, they want others to use their work and cite it. „*The attribution requirement in the CC-Licence ensures that access to their work builds upon their reputation*" (HOORN, 2006). CC licences allow authors to skip the intermediaries (e.g., commercial publishers) which is what most grey literature authors want or have to do. This is mostly in accord with related *open access* publication model. „*In fact, a large portion of the available GL [grey literature] is OA [open access]*" (HAWKINS, 2008). For the afore-mentioned purposes, CC licences are an optimal tool for effective dissemination of grey literature.

The first set of CC licences was released in 2002 for free to the public. CC licences have grown at an exponencial rate around the world and expanded over American border. National CC licences versions are based on the generic licence, they are created in national language and adapted to specific system of law.

Czech version of Creative Commons licences was translated by civil association Iuridicum Remedium in cooperation with the Faculty of Law, Charles University and the National Library of the Czech Republic. Czech version of Creative Commons licences was published in April 2009.

Licencing PSH under CC is a result of the Committee for Coordination of PSH meeting. But there were more motivating factors behind the decision to use CC licence for PSH. The National Technical Library as a modern institution recognises that new approches to copyright management should be considered in order to exploit the enormous potential for knowledge distribution proffered by the Internet. The main goal was to make PSH avaliable under an open licence to promote the use and dissemination of PSH.

Key motivating factors to release PSH under CC licence include:

⋄ Encourage library community and also Internet users contribution of knowledge to participate in NTL projects including PSH.

⋄ Dissemination of PSH, where possible, in open formats and as open data.

⋄ Give others easy access and the right to use PSH without seeking NTL´s permission.

⋄ Licencing under CC allowed us to make PSH avaliable to anyone who might be interested in controlled indexing and searching.

⋄ We want to take every opportunity to communicate new ideas broadly and seek feedback because great ideas do not develop in isolation.

Proposal how to improve PSH by publishing it under CC licence was approved immediately after publishing Czech CC licence version. PSH is licenced under the Creative Commons Attribution Non-Commercial Share Alike 3.0 (Czech).[19]

The National Technical Library allows any individual user to copy, distribute, display and perform the PSH. But they must give the original author credit, they may not use this work for commercial purposes and if they alter, transform, or build upon this work, they may distribute the resulting work only under an identical licence that governs PSH.

---

[17]http://www.w3.org/Submission/ccREL/
[18]http://www.w3.org/TR/xhtml-rdfa-primer/

[19]http://creativecommons.org/licenses/by-nc-sa/3.0/cz/

5 Conclusion

Many target groups may get benefit from PSH.

Our aim is to encourage use of PSH by the wider web audience outside of libraries. We will try to ensure PSH has a role in the present-day ever-changing web environment and will not become an obsolete tool. This involves staying up to speed with current trends and recents developments in knowledge organisation systems area which is most probably shifting towards the semantic web.

To librarians we want to communicate the message PSH is a progressive subject heading system that is there available free to use. In future, our effort is to increase consistency of indexing and searching and improve access to digital collection of grey literature by using PSH, e.g., in the National Repository of Grey Literature in the National Technical Library in Prague.

And because in the future publishing certainly will be no shortage of grey literature we will attempt to convince the ever-increasing number of authors and publishers operating in this field of benefits of using PSH. Knowledge organisation systems such as PSH are able to keep the knowledge these people are trying to share from fragmenting into unmanageable chaos of irretrievable and unaccessible bits and bytes.

Bibliography

BERNERS-LEE, Tim. *Linked data : design issues* [online]. Published 2006-07-27. Last change 2009-06-18 [cit. 2009-06-10]. Available from WWW: <http://www.w3.org/DesignIssue/LinkedData.html>.

BINDING, Carl; TUDHOPE, Douglas. KOS at your service : programmatic access to knowledge organisation systems. *Journal of Digital Information* [online]. 2004 [cit. 2009-09-30], vol. 4, no. 4. Available from WWW: <http://journals.tdl.org/jodi/article/view/110/109>. ISSN 0972-7272.

Case study : Blender Foundation [online]. 2009a. Last change 2009 [cit. 2009-10-02]. Avaliable from WWW: <http://wiki.creativecommons.org/Case_Studies/Blender_Foundation>.

Case study : Filter Magazine [online]. 2009b. Last change 2009 [cit. 2009-10-02]. Avaliable from WWW <http://wiki.creativecommons.org/Case_Studies/Filter_Magazine>.

Case study : Monkeyc.net [online]. 2009c. Last change 2009 [cit. 2009-10-02]. Avaliable from WWW <http://wiki.creativecommons.org/Case_Studies/Monkeyc.net>.

Creative Commons [online]. [2001] [cit. 2009-10-05]. Avaliable from WWW: <http://creativecommons.org/>.

CHELIOTIS, Giorgos [et al.]. Taking stock of the Creative Commons experiment monitoring the use of Creative Commons licenses and evaluating its implications for the future of Creative Commons and for copyright law. In *Proceeding of the 35th conference on research communication, Information and Internet Policy (TPRC)*. Virginia : George Mason University School of Law, 2007. Also avaliable from WWW: <http://wiki.creativecommons.org/images/7/71/Taking-stock-of-the-creative-commons-experiment_eng.pdf>.

CHILDRESS, Eric. Going grey : grey literature and metadata. *Journal of Internet Cataloging*. 2003, vol. 6, no. 3. ISSN 1091-1367 (print). DOI 10.1300/J141v06n03_02.

DE BLAAIJ, Cees. Grey literature from invisibility to visibility. *Publishing Research Quarterly*. 2004, vol. 20, no. 1, pp. 70–76. ISSN 1053-8801 (print). DOI 10.1007/BF02910862.

DE BLAAIJ, Cees. Two worlds : about bars and stars in scientific information publishing, an analysis of open source ideology as a means of self-controlled publishing. *Grey Journal*. 2005, vol. 1, no. 1. ISSN 1574-1796 (print).

HAWKINS, Donald T. Gray literature : what's new on the information landscape. *Information Today.*

February 2008, vol. 25, no. 2, pp. 27–29. ISSN 8755-6286.

HOORN, Esther. *Creative Commons licences for cultural heritage institutions : a Dutch perspective* [online]. Amsterdam : University of Amsterdam. Faculty of Law. Institute for Information Law, 2006 [cit. 2009-09-30]. Available from WWW: `<http://www.ivir.nl/creativecommons/CC_for_cultural_heritage_institutions.pdf>`.

KRUK, Sebastian Ryszard; MCDANIEL, Bill (eds.). *Semantic digital libraries*. Berlin, Heidelberg : Springer, 2009. ISBN 978-3-540-85433-3.

MACKENZIE OWEN, John S. The expanding horizon of Grey Literature. In *Perspectives on the design and transfer of scientific and technical information : proceedings of the 3rd international conference on grey literature*. Amsterdam : TransAtlantic, 1998, pp. 9–13. Also available from WWW: `<http://cf.hum.uva.nl/bai/home/jmackenzie/pubs/glpaper.htm>`.

Perspectives on the design and transfer of scientific and technical information : 3rd International conference on grey literature, Luxembourg, 13–14 November 1997. Amsterdam : Transatlantic, 1997. ISBN 90-74854-17-7.

RABINA, Debbie L. Copyright licenses and legal deposit practices of grey multimedia materials. *Grey Journal*. 2009, vol. 5, no. 1. ISSN 1574-1796 (print).

SCHÖPFEL, Joachim. Observations on the future of grey literature. *Grey Journal*. 2006, vol. 2, no. 2. ISSN 1574-1796 (print).

SHIRI, Ali Asghar; REVIE, Crawford. Thesauri on the web : current developments and trends. *Online Information Review*. 2000, vol. 24, no. 4, pp. 273–279. Available from WWW: `<http://dlist.sir.arizona.edu/163/01/thesauri.pdf>`. ISSN 1468-4527.

SHIRKY, Clay. *Ontology is overrated : categories, links, and tags* [online]. 2005 [cit. 2009-09-30]. Available from WWW: `<http://www.shirky.com/writings/ontology_overrated.html>`.

SIEGEL, Gretta E. Capturing academic grey literature : starting at home. *Publishing Research Quarterly*. 2004, vol. 2, no. 1, s. 62–69. DOI 10.1007/BF02910861. ISSN 1053-8801 (print).

TUDHOPE, Douglas; KOCH, Traugott; HEERY, Rachel. Terminology services and technology : JISC state of the art review. *UKOLN Newsletter* [online]. 2006 [cit. 2009-09-30], iss. 17. Also available from WWW: `<http://www.jisc.ac.uk/media/documents/programmes/capital/terminology_services_and_technology_review_sep_06.pdf>`. ISSN 0963-7354.

VESELY, Martin [et al.]. CERN Document Server : document management system for grey literature in networked environment. In 5th *international conference on grey literature : grey matters in the world of networked information*. Amsterdam : TextRelease, 2004. GL-conference series, no. 5. Also available from `<http://www.greynet.org/images/GL5-Vesely.pdf>`. ISBN 90-77484-01-9.

VLACHIDIS, Andreas [et al.]. *Excavating grey literature : a case study on rich indexing of archaeological documents by the use of natural language processing techniques and knowledge based resources* [online]. 2009 [cit. 2009-09-30]. Preprint for ISKO UK 2009 : Content Architecture: Exploiting and Managing Diverse Resources. Available from WWW: `<http://www.iskouk.org/conf2009/papers/vlachidis_ISKOUK2009.pdf>`.

WILDE, Eric; LIU, Yiming. Lightweight Linked Data. In *Proceedings of the 2008 IEEE International Conference on Information Reuse and Integration*. New York (NY) : IEEE Systems, Man, and Cybernetics Society, 2008. Also available from WWW: `<http://dret.net/netdret/docs/wilde-iri2008-lightweight-linked-data.pdf>`.