LINKED DATA FOR OPEN VOCABULARIES AND *HIVE*'S GLOBAL FRAMEWORK

Eva Méndez and Jane Greenberg

Eva Méndez is an associate professor at the *Library and Information Science Department* in the *University Carlos III of Madrid*, where she directs the *Master in Digital Libraries* program. She holds a PhD in information science and for the last fifteen years has been teaching and researching mainly in the following areas: metadata, semantic web, digital libraries, open access, information policies and social web. She is a member of the *DCMI Advisory Board* and the co-chair of the *DCMI Social Tagging Community*, and independent expert for the *European Commission* in the areas of digital libraries and open access to research data.

Universidad Carlos III de Madrid
Facultad de Humanidades, Comunicación y Documentación
C/ Madrid, 128 (Dcho. 14.2.17) – 28903 Getafe (Madrid), España
emendez@bib.uc3m.es

Jane Greenberg is currently chair of excellence in the *University Carlos III of Madrid* at the *Library and Information Science Department*. Dr. Greenberg is professor in the *School of Information and Library Science* (*SILS*) at the *University of North Carolina* at Chapel Hill, and director of the *SILS Metadata Research Center*. She earned her PhD from the *School of Information Sciences* at the *University of Pittsburgh* and received a MSc with a specialization in bibliographic control from the *School of Library Service* at *Columbia University*. She is the recipient of the *Margaret Mann Citation* for outstanding professional achievement in cataloging or classification.

School of Information and Library Science
University of North Carolina at Chapel Hill
205 Manning Hall, CB # 3360 – Chapel Hill, NC 27599-3360
janeg@email.unc.edu

Abstract

This paper summarizes new trends and advances in Knowledge Organization from the perspective of linked open data (LOD). Although this is particularly important to galleries, libraries, archives and museums, the so-called GLAM community, it is of more general relevance, and part of the value of LOD lies in its adoption beyond that community. LOD includes descriptive metadata and vocabulary encoding schemes that are being "skosified" (encoded in the SKOS format) or rendered in OWL (the web ontology language) and made available not only "on" the web, but "for" the semantic web. The paper highlights a few exemplary initiatives in the field. The paper also introduces the *HIVE* (*Helping Interdisciplinary Vocabularies Engineering*) framework and discusses the *HIVE-ES* (España) extension for Spanish language vocabularies, leading to a more global approach for linked open vocabularies (LOV).

Título: Datos enlazados para vocabularios abiertos: marco global de *HIVE*

Resumen

Se presentan brevemente las nuevas tendencias y avances en la organización del conocimiento desde la perspectiva de *linked open data* (LOD). Aunque esto es particularmente importante para galerías, bibliotecas, archivos y museos –la llamada comunidad GLAM-, es de importancia más general, y parte del valor de LOD se encuentra en su adopción más allá de esa comunidad. LOD incluye esquemas de metadatos descriptivos y de codificación de vocabularios que están siendo "skosificados" (codificados en el formato SKOS) o transformados en OWL (el lenguaje de ontologías web) y puestos a disposición no sólo "en" la Web, sino también "para" la web semántica. Se destacan algunas iniciativas ejemplares en este campo y se presenta el marco *HIVE* (*Ayuda a la Ingeniería de Vocabularios Interdisciplinares*) y se analiza la extensión *HIVE-ES* (España) para los vocabularios en español, dando lugar a un enfoque más global a los vocabularios abiertos enlazados (LOV).

1. Introduction

Over the last few years, increased attention has been directed toward linked data, linked open data, and sharing vocabularies in an open environment. Linked open data, or just LOD, has become a buzzword appearing in nearly every initiative focusing on digital information organization. LOD is an approach to fostering and advancing the semantic web and the web of data -an idea that initially gained prominence in the late 90's through the work of **Tim Berners-Lee**, the inventor of the World Wide Web. In many respects, linked data has evolved as the *de facto* means for the publication of structured data on the Web, involving a broad range of different data communities. The use of LOD is growing in terms of the number of disciplines committed to LD best practices and technologies for exposing and interlinking datasets for seamless access to and sharing, integration and reuse of those data. The progress has been tremendous, with close to 3,500 LOD datasets available according to *thedatahub.org* (May 2012).

> Vocabularies, formalized as knowledge organization systems (KOS), help address problems with digital information overload and aid information discovery

Vocabularies are the base of linked data. Vocabularies, formalized as knowledge organization systems (KOS), based on the languages of a domain, discipline, or community, help address problems with digital information overload and aid information discovery. It is clear that vocabularies matter, not only in the traditional library world but also for many different digital information stakeholders. A recent survey by *Semantic Web Company* found that 85.4% of the 158 participants, coming from IT, science, public sector and education, use controlled vocabularies in their organizations. A very large proportion (88%) of these KOS users also agree that most organizations could benefit from linked data. 48.7% of respondents stated that standards like SKOS (simple knowledge organization system) are "very important" with a further 29.1% rating this approach as "relevant" (**Kondert**; **Schandl**; **Blumauer**, 2011).

Vocabulary users in the new LD landscape include different communities using complementary approaches to share and take advantage of vocabularies:

– Web developers embedding structured data describing web content into their html pages using encoding standards such as microformats, microdata and RDFa (instead of building up Sparql endpoints), described with a shared markup vocabulary such as *Schema.org*.

– The *W3C* semantic web community, who develop and use ontologies as formalized vocabularies. This includes using OWL to build vocabularies or ontologies, and SKOS for creating KOS in the process of data enrichment.
– Communities around metadata standards such as *Dublin Core* or particular metadata application profiles that include sets of descriptive elements and properties to constitute a vocabulary or metadata schema.
– Users of knowledge organization systems (KOS) that encompass value-space structures, such as thesauri, subject classifications and authority files, traditionally considered "controlled vocabularies".

A perceptive view of the LD landscape is found in a recent post on **Bernard Vatant**'s blog (2012): He presents an inspiring list of existing and candidate "vocabulary publishers" including standard-developing organizations (*W3C*, *DCMI*); institutional heritage curators (*Library of Congress*) and the global organizations federating the work of those curators (e.g. *Ifla*, *Europeana*, *Oclc*); media groups and associations (*BBC*, *The NYT*, *The Guardian*); governments and institutional data providers (*data.gov*, *UN*, *World Bank*), research centers (*Deri*, *Inria*); specific funded research projects; small and medium enterprises in the field (*Talis*, *Mondeca*); large internet companies through initiatives like *Schema.org* (*Google*, *Bing*, *Yahoo!* and *Yandex*); and domain-specific and individual initiatives.

Linked data uses vocabularies in two ways: as schemas, providing a set of properties that an object might have; and as vocabulary encoding schemes describing the range of values a given property might take, in formal terms. Linked vocabularies aid knowledge acquisition by tightly controlling and contextualizing data (concepts, objects, etc.). This approach enables familiar metadata processes, while the ability to link both vocabularies and data records themselves provides an infrastructure that allows more effective information discovery and use.

This article first presents a brief history of ideas around linked data, from its semantic web foundations to the evolution of definitions of linked data, linked open data, and linked open vocabularies. We introduce the *HIVE* framework and discuss how the *HIVE-ES* extension for Spanish language vocabularies leads to a more global approach for LOV. The conclusion summarizes this work and offers several thoughts that have emerged from this inquiry.

2. Linked open data in context

Linked open data (LOD) or simply linked data (LD) has become a standard topic of calls for papers in the information and library science, web and computing science communities. Attention to linked data is also visible in a range of sub-

ject domains, such as health, government, and education. In fact, nearly every discipline producing data has a cluster of advocates promoting open and sharable data. Perhaps less obvious, but significant, is the goal to move toward linked open vocabularies (LOV), a step that aids data sharing and reuse as well as linking knowledge. This trend echoes earlier arguments put forth for moving toward the semantic web; linked data is integral for the development of a semantic web. In exploring LOV, it is useful to gain a sense of how this development fits into the history of the web.

2.1. A brief history of linked data

The idea of linked data did not emerge overnight. Rather, like many developments in the field of information and library science and computing, the foundational ideas and goals predate the Web. Common examples take us back to the notion of universal bibliographic control, or **Vannevar Bush**'s (1945) presentation of the *Memex*, a hypothetical device using associative linking to sustain memory over time. As **Dan Brickley** (2012) stated recently, some current work in linked data and semantic web could benefit from understanding its place in a longer history going back to 1912 annual report from the Belgian *Institute of Bibliography*.

Tim Berners-Lee's original ideas from 1989 were not quite the Web we have today, but focused more toward linking concepts, ideas, facts –i.e. data– on a global scale. This initial idea of the Web has been promoted as the semantic web, "an extension of the current Web in which information is given well-defined meaning, better enabling computers and people to work in cooperation" (**Berners-Lee**, 1999). **Berners-Lee** has been promoting, and perhaps defending, the same idea since at least 1989: a powerful structure of interconnected knowledge that linked information, documents

and data; he called it first the Web (1989), then, the semantic web (1998) and then linked data (2006). As he envisaged in *Weaving the Web* (1999), html allowed the hypertextual web of documents, while RDF and semantic web technologies (OWL, SKOS, Sparql) will allow the web of data through linked datasets defined as RDF triples.

Figure 1 shows historical landmarks toward the semantic web and linked data, using as a guideline the *World Wide Web Consortium* (*W3C*) standards and initiatives which made it possible. A thoughtful compilation of those semantic web standards has been completed in Spanish by **Pastor** (2011).

The semantic web allows machines to derive meaning from structured data which can be processed when published as linked data. Linked data is a key implementation step toward the semantic web, where representing information entities via URIs makes them machine-processable. Linked data are therefore essential ingredients of the semantic web. But to achieve a real global semantic web, just linked data or just open data is not sufficient. What is necessary is data that is both linked and open. That is why **Berners-Lee**, two decades after inventing the web, and after a decade focused on the development of the semantic web, encouraged people in his 2009 TED speech to open the "raw data, now!", and express them as linked data, to enable getting them out of their silos.

2.2. Linked data, LOD dissected and defined

Explanations, discussion, and formal definitions for linked data generally reflect **Berners-Lee**'s (2006) foundations and recognize LD as a set of best practices for publishing and connecting structured data on the Web. Linking data is about using the Web to connect related data that were not previ-
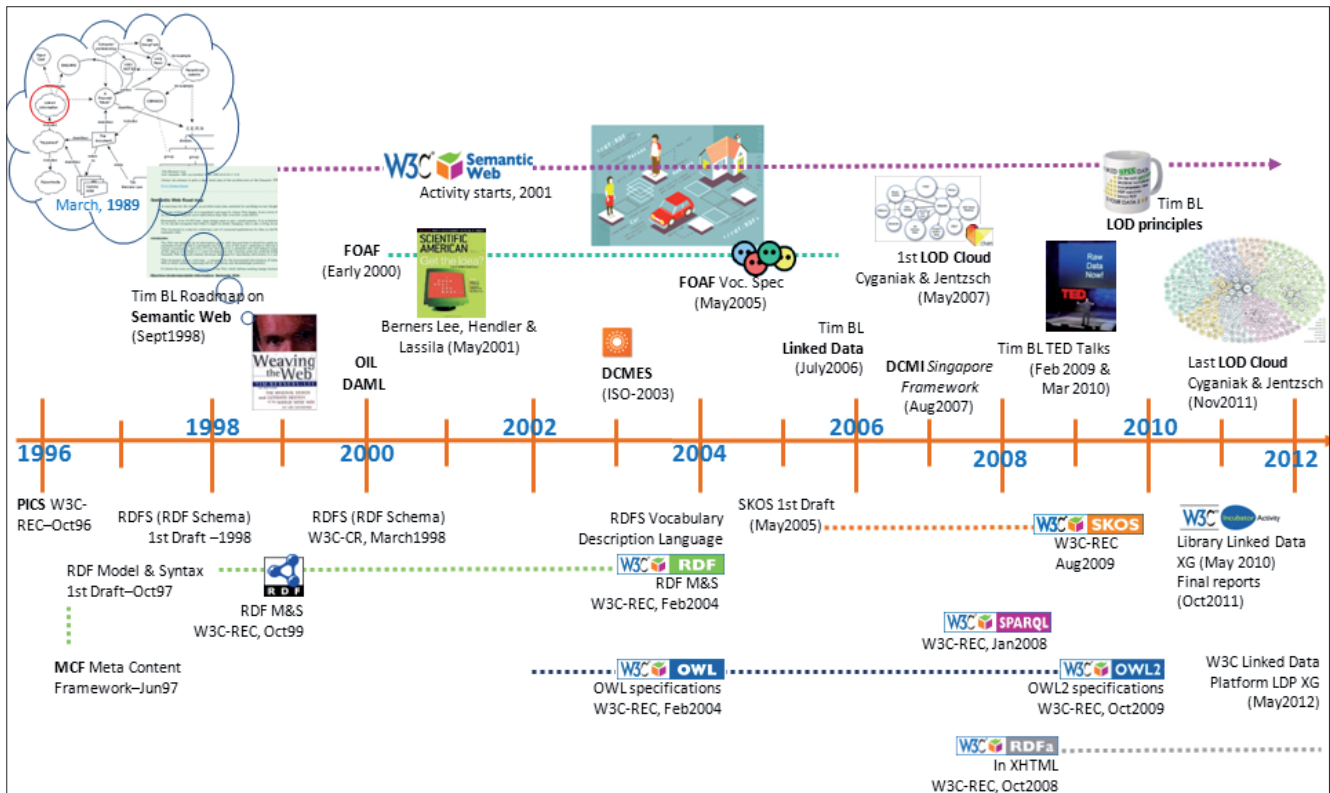

Figure 1. Evolution of the semantic web into linked data (image by E. Méndez)
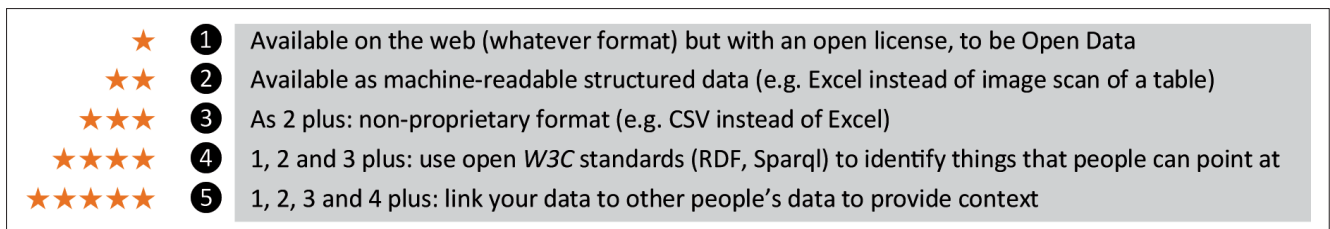
★	❶	Available on the web (whatever format) but with an open license, to be Open Data
★★	❷	Available as machine-readable structured data (e.g. Excel instead of image scan of a table)
★★★	❸	As 2 plus: non-proprietary format (e.g. CSV instead of Excel)
★★★★	❹	1, 2 and 3 plus: use open *W3C* standards (RDF, Sparql) to identify things that people can point at
★★★★★	❺	1, 2, 3 and 4 plus: link your data to other people's data to provide context

Figure 2. Openness principles added to the linked data paradigm (**Berners-Lee**, 2006 updated in 2010)

ously linked, or using the Web to lower the barriers to using data currently published using other methods (*Linkeddata.org*, 2012). Those foundations have basically four rules and five principles. The rules are pretty simple: 1) Use URIs as names for things (cool URIs for the semantic web are those that not change); 2) Use http URIs (dereferenciable URIs) so that people can look up those names; 3) When someone looks up a URI, provide useful information using the appropriate standards (RDF, Sparql); and 4) Include links to other URIs, so that people can discover more things. In 2010 he refined his conception of LD adding the openness philosophy in such a way that linked *open* data is linked data which is released under an open license, which does not impede its reuse for free, and he stated five levels or principles of LOD, like star ratings for hotels:

In general terms, linked data can be understood as an approach for encoding data at the most granular level. Data, in this context can be any atomic thing (including a concept), or an RDF statement or a set of statements, that has an identifier in the form of a URI. Data and objects that are part of linked data paradigm can come from standard vocabulary systems such as Dublin Core and from vocabulary encoding schemes in the form of controlled vocabularies, ontologies, taxonomies, name authority files, classificatory systems, and the like. Objects are selected from these vocabularies for property/value encoding, or refinements of this type of information. There is no limit to the types of vocabularies that can be transformed into linked data. One could publish the entire *Oxford English Dictionary* or the *Diccionario de la RAE* (Spanish Royal Academy Dictionary) as linked data.

Linked data uses uniform resource identifiers (URIs) as globally unique identifiers for any kind of resource analogously to how we use identifiers in librarianship for authority control. **Eric Miller** underlined the importance of the identifiers and pointed out that the persistence of these identifiers is a crucial part of integrity within systems. Using http URIs as a way of unifying local primary keys inside databases charts out a universal data space, not only for library organizations, but for any organization that wants to share information with others. "Traditionally we have kept these local identifiers inside systems. Now we are exposing the local identifiers so that external information can be hooked to them" (**Miller**, 2011), so every single data/thing in the LD world could be linkable.

A number of information professionals and researchers are realizing the significant added value that galleries, archives and museums can acquire by opening and linking their cultural data, notably **Byrne** and **Goddard** (2011) in Canada, **Oomen** *et al.* in the Netherlands (2012), **Saorín** (2012) and **Peset** *et al.* (2011) in Spain, and **Jon Voss** (LOD-LAM, 2011)

in the U.S. But the most important contribution, seen from the perspective of memory organizations and people working in digital humanities are the reports created by *W3C*'s library linked data incubator group (W3C LLD XG, 2011a & b; **Baker**, 2012). These define linked data as data published in accordance with principles designed to facilitate linkages among datasets, element sets, and value vocabularies. Those vocabulary encoding schemes (value vocabularies) are integral to the *HIVE* approach discussed here below.

3. From LOD to LOV: linked open vocabularies as a part of the new knowledge organization ecosystem

Research has shown that the "subject" or conceptual search based on a topic is the most common type of search in the context of the Web (**Yu**; **Young**, 2004; **Savolainen**; **Kari**, 2006). That is, people look for information on a topic such as a place of travel, a health condition, or a historical event, more frequently than they are looking for a specific organization, the official movie trailer for a new release, or a specific article. Even when a person is in search of a "known entity" such as a person or place, he may use conceptual terms. It is the *aboutness* (the subject) that we seek, and it is a semantic search that we need.

> It is the *aboutness* (the subject) what we seek, and it is a semantic search that we need

For example a person interested in information about the history of the *Parque del Retiro* (Retiro park, in Madrid) may initiate a web query by searching with the concepts "Spain" and "parks," and perhaps the name "Retiro". The query in this case is not specific to the geographic coordinates, rather the searcher may be equally interested to learn of the park's development in 1632 as a retreat outside the city wall for King Philip IV and Royal family, or how it was used during the Spanish civil war. The *aboutness* or topical content can be represented by concepts from skosified vocabularies. With linked open vocabularies, there is a tremendous potential to bring this benefit to the larger global Web by linking information in an open environment.

3.1. Opening and linking vocabularies to build up a new KOS ecosystem

For vocabularies to be reused in this new ecosystem they need to be published as LOD. There are many vocabularies available on the web although their encoding may differ. Some time ago we foresaw the potential for XML/RDF to encode and express vocabularies at intranet or corporate level

(**Méndez**, 2000). The development of a standards-based semantic web (fig. 1) offers the possibility of global semantic interoperability through the web of LD. SKOS, a *W3C* recommendation since 2009, is a step forward to make thesauri more interoperable, making them sharable and accessible for project integration and linking content. This approach fosters the semantic web environment where vocabularies are explicitly created for the web. In this context, we present linked open vocabularies.

The technical implementations that link open vocabularies are based on semantic web technologies, standards and best practices, including RDF/SKOS, OWL, Sparql endpoints and triple stores. The use of semantic web technology offers the potential for better dissemination and integration of data into third-party applications and environments (**Vatant**, 2012). Using semantic web standards we not only put vocabularies "on" the web, for people to read about, but also publish them "for" the semantic web, enabling machines to use them directly. The *Datahub* is a community-run catalogue of datasets that also includes vocabulary encoding schemes, among a huge amount of "things" currently published as linked data. LOV might be understood as a clarified subset of the messy LOD cloud, those vocabularies that can be considered KOS (knowledge organization systems) or subject-oriented vocabularies. **Bernard Vatant** (2012) in his thorough provoking posts of *LOV stories* says that "there are as many data looking for good vocabularies as vocabularies looking for data". He also thinks that the activities to create a new ecosystem for linked open vocabularies should be funded in the philosophy of the commons and *co-opetition*, as shared resources in which each stakeholder has an equal interest leading them to establish a cooperative competition.

> Using semantic web standards we not only put vocabularies *on* the web, for people to read about, but also publish them *for* the semantic web, enabling machines to use them directly

3.2. Drawing the new knowledge organization system's landscape. Linked open vocabularies initiatives

There are a number of collective registries providing access to the open vocabularies this way in a linked data/semantic web environment, reconfiguring a new landscape for network knowledge organization systems. Four approaches are noted here:

Linked Open Vocabularies project (*LOV*). The LOV registry created by **Bernard Vatant** and **Pierre-Yves Vandenbussche** and published by *Mondeca* labs as part of the *Datalift* project, applied in March 2012 to be hosted by the *Open Knowledge Foundation* (*OKFN*), like the *Datahub*. As they stated in this application, *LOV* project aims to provide easy access to vocabularies, in particular the ways they link to each other, and by providing metrics on how they are used in the linked data cloud to help improve their comprehension, visibility, usability, and overall quality.

This is one of the most innovative projects in this area; and as already noted the ideas of *LOV* emphasize the ideas underlying this paper. The *LOV* project presents a "growing ecosystem of linked open vocabularies (RDFS or OWL ontologies) used in the LOD cloud". The project catalog gives access (by search or navigation) to 262 vocabularies so far, rendered in RDF, OWL and SKOS that are classified by vocabulary spaces, interlinked using the dedicated vocabulary *VOAF* (*vocabulary of a friend*) to describe ontologies that are part of the linked data cloud. It works as a registry, but they have taken an extra step to classify the vocabularies in a broad approach.

Open Metadata Registry (*OMR*). The initiative "began its development as the *NSDL Registry*, attempting to address the big question: What should these registries do and how can they operate in an open services environment?" (**Phipps**; **Hillman**, 2011) and is today becoming the most crucial registry for the library community. The *OMR* extends well beyond the library, archival, and museum community, given the original seed in scientific education, and includes a range of vocabularies (element sets, ontologies and controlled vocabularies). An innovative aspect of this registry is the *sandbox* component where individuals can experiment, play, share, and learn to how to participate in the LOV environment. Work is ongoing to enable the *OMR* to support multilingual vocabularies and separate but associated language versions. One of the chief creators, **Diane Hillman**, is a founder and community leader of the *DCMI-Vocabulary Management Community* (*DCMI-VMC*), a forum addressing best practices for vocabularies in the semantic web.

Amalgame (*Amsterdam alignment generation metatool*) is an interactive alignment server under development at the *VU University Amsterdam* in the context of the *PrestoPrime* and *EuropeanaConnect* projects. This tool realizes the second step of a specific workflow to skosify vocabularies and convert collection metadata to the *Europeana data model* (*EDM*). *Amalgame* aims to find, evaluate and manage vocabulary alignment within the context of the *Ontology alignment evaluation initiative* (*OAEI*), in which different alignment methods can be combined using a workflow setup. The main difference from the *LOV* project is that *Amalgame* includes a limited number of vocabularies, related mainly to cultural heritage, and looks for alignment among them. The main difference from *OMR* is that *Amalgame* only includes vocabulary encoding schemes, for an alignment among them based on axis vocabularies.

NCBO-BioPortal is a web-based application that provides access to 302 biomedical ontologies and vocabularies including thesauri in the field of biology and biomedicine. The ontology library specifically allows for browsing, searching, and downloading ontologies. There is also work enabling semantic mapping among vocabularies. In this case, ontologies registered are encoded in OWL; OBO, an ontology language for the bio domain; and increasingly SKOS renderings are being registered and provided. All these vocabularies or ontologies are used to support conceptual search over biomedical resources. It is more than a specialized vocabulary service. Once a search term is refined and aligned among different vocabularies, you can click and explore the re-
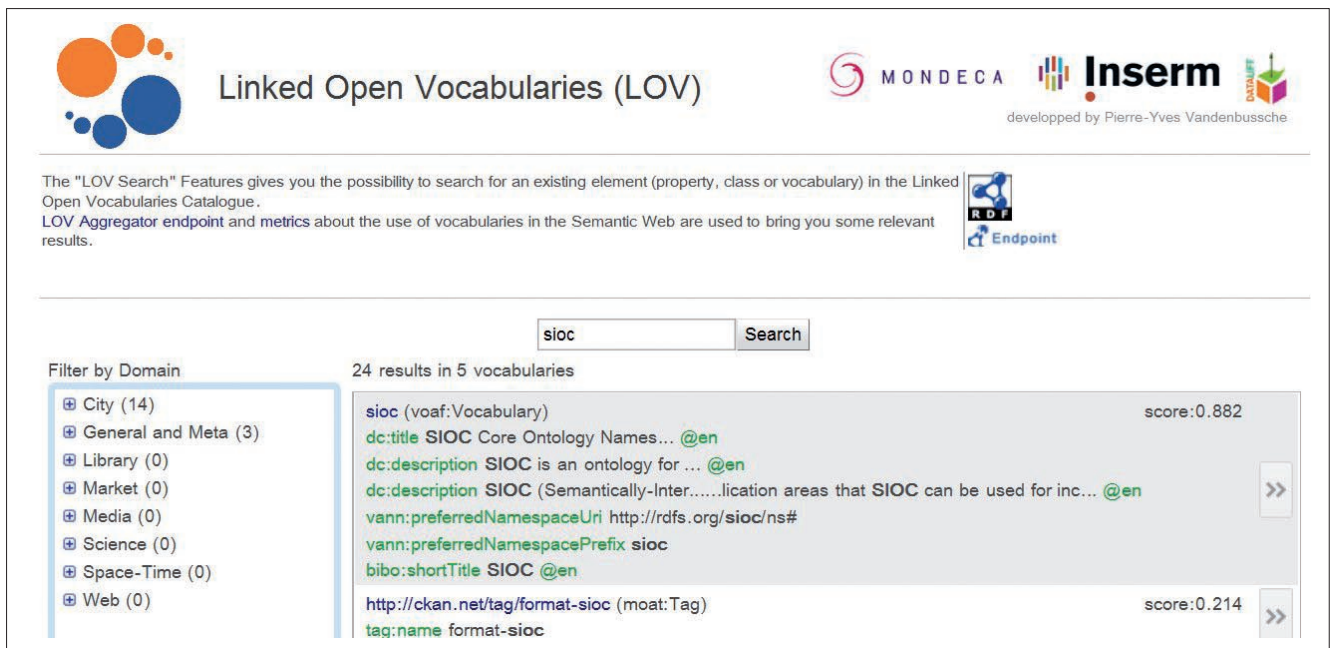
Figure 3. *LOV Project*, search about SIOC (Semantically-interlinked online communities) vocabulary

source in its original site. The main difference with this registry from those previously described is the domain specific approach.

> *HIVE* and *HIVE-ES* provide a solid foundation for linking and opening interdisciplinary vocabularies to the multilingual environment

4. *HIVE-ES* framework: skosifying, opening, and linking vocabularies

HIVE (*Helping Interdisciplinary Vocabulary Engineering*), launched with support from the *U.S. Institute of Museum and Library Studies*, has been pursued as a demonstration project. *HIVE* presents a model using open, linked vocabularies for dynamically creating subject metadata at the time of indexing, drawing from multiple vocabularies. This process enables a selection of the best concepts for representing the content of the resource that is not limited to a single vocabulary. In the metaphor that **Greenberg** (principal investigator) and members of the *HIVE* team have used to explain it, a bee goes out to flowers seeking pollen, and brings nuggets back to the hive; in this case, relevant concepts. A similar idea, conveyed by **Bernard Vatant** (2012) through his "LOV gardens" metaphor where vocabularies like *Schema. org* are trees or flowers in the garden, complements this idea.

The *HIVE-ES* (España) initiative is building on this work to address known KOS challenges, and provide a means for simultaneous search and metadata generation drawing from multiple Spanish language vocabularies. Specifically, *HIVE* and *HIVE-ES* have been initiated to address vocabulary problems relating to cost, interoperability, and usability limitations (**Greenberg** *et al.*, 2011):

– Controlled vocabularies are expensive to create and maintain.
– Adherence to *ISO*, *Ansi/Niso*, *W3C*, *IETF*, and other standards does not confirm knowledge organization systems (KOS) interoperability.
– KOS design does not always support easy access and use.

The *HIVE-ES* initiative extends the initial *HIVE* activity to address these challenges in Spanish speaking countries. *HIVE-ES* allows for simultaneous search and metadata generation drawing from multiple Spanish language vocabularies. The project has been launched at the *Tecnodoc research group* (*Tecnologías Aplicadas a la Información y la Documentación*) of the *Department of Library and Information Science*, *Universidad Carlos III de Madrid*, the *National Library of Spain* (*BNE*), and *Metadata Research Center*, *School of Information and Library Science*, *University of North Carolina* at Chapel Hill (*Sils-MRC*).

HIVE-ES is making vocabularies and providing a demonstration project supporting interdisciplinary indexing. The emphasis here is on Spanish language vocabularies. Essentially, vocabularies integrated into the *HIVE-ES* are composed as linked data using SKOS encoding language, to represent their classes and properties.

HIVE-ES vocabulary server currently has three vocabularies: *The United Nations Food and Agriculture vocabulary* (*Agrovoc*) (the Spanish language section), *LEM* (*Lista de encabezamientos de materia*) and *EMBNE* (*Encabezamientos de materia-BNE*, the *Spanish National Library*'s equivalent of the *Library of Congress Subject Headings*). The integration of *EMBNE* in *HIVE-ES* was the most complex undertaking, given the need to convert Marc21 format for authority data into SKOS. The processing involved 1) conceptually mapping Marc21 fields (e.g., 1XX, 4XX and 5XX) to appropriate SKOS labels (e.g., 'skos:prefLabel' and 'skos:altLabel'); 2) reading and parsing the *EMBNE* file in Marc21 format; and 3) relabeling the individual contents of the *EMBNE* to the SKOS

Figure 4. *HIVE-ES* project wiki

equivalents. Conversion is not supported by a single tool, although the *Marc4J Java library*, which defines an application programming interface (API) for parsing large Marc files, was a tremendous help.

The framework and infrastructure of *HIVE-ES* follows that of the original *HIVE* and includes:

– *HIVE* Core for automatic metadata extraction, topic detection and concept retrieval. The extraction uses KEA (key-phrase extraction algorithm) and concept retrieval is supported via *Lucene*. These are the system's main functions. Additionally, the core includes RDF storage and management using *Elmo* to store objects and properties in an RDF repository supported by *Sesame*.
– *HIVE* vocabulary service includes *HIVE*'s web accessible interface so that users can browse and search through vocabularies in the demo system. This was developed using the *Google web toolkit*.
– *HIVE* REST (*representational state transfer*) aspect provides an API based on web services to facilitate integration for third party software.

HIVE-ES is in the early stages of development and implementation, and plans are under way to add more Spanish language vocabularies. Current *HIVE* renderings that have been documented are monolingual. The current *HIVE* code doesn't support the integration of multilingual vocabularies, but the open nature of this initiative and sharing of code opens the door for other developers to contribute to *HIVE* in this manner. The *HIVE-ES* project pushes the edge here highlighting the need to look at vocabularies in languages other than English, moving toward a global context.

5. Conclusion

This paper provides insight into the evolving trends advancing knowledge organization and sharing vocabularies from the perspective of linked data and linked open data. Among several firm conclusions to be offered are:

– The technological infrastructure to support LD/LOD has become quite powerful through the use of networked technologies and the development and adoption of *W3C* standards (RDF, SKOS, Sparql, etc.).
– Vocabularies in their many forms (thesauri, taxonomy, ontology, and discipline, domain and community languages) can be leveraged and made more powerful via RDF/SKOS.
– Many communities are embracing this new potential, skosifying their vocabularies and making them open and interoperable.
– Registries are progressing to a new level with developments such as *LOV project*, *OMR*, *Amalgame*, and the *NC Bioportal*, and providing a sustainable means for sharing vocabularies and supporting semantic web operations.
– Publishing open vocabularies is not just about vocabularies developed for the web. **Tim Berners-Lee**'s "five stars" principles for LOD are generally applicable to enabling knowledge organization systems to augment the store of information available for effective use (open licensing).
– The community of people embracing LOD and LOV is growing swiftly and with enthusiasm, embracing both scientific domains and humanities-driven endeavors such as Europeana.

Proprietary thesauri and other knowledge organization systems have the potential to create a new landscape, and efforts are forging ahead beyond a limited display of textual or hypertextual content. These KOS are increasingly both "on" the web and being created "for" the semantic web. The examples provided in this paper, and additional progress with several aspects of the linked data infrastructure, are important steps in this new ecosystem. *HIVE* is also an exemplary approach in this new LOV landscape. The *HIVE-ES* extension of the original *HIVE* is for Spanish language vocabularies, leading to a more global approach for linked open vocabularies. These steps provide a solid foundation for linking and opening vocabularies to the multilingual environment and sharing value spaces vocabularies across languages.

As digital information initiatives aim to make their resources, or descriptions of their holdings, accessible via the global web, they strive to provide a venue for interdisciplinary and cross-disciplinary research. Linked open vocabularies are part of this growing trend; however, frameworks allowing for simultaneous searching and indexing via multiple vocabularies in this context require further work to reach a fully interoperable environment. The move to sharing in this way needs to be supported by a broader social and technological infrastructure that brings people, technologies, and vocabulary together. A sense of community and shared vision is integral to make this approach sustainable. The *DCMI Vocabulary Management Community* is gaining significant interest in this respect, with leadership, inspiration, and engagement from members of the *OMR* and *LOV* projects. It is the collection and coordination of developments and community that will allow this evolution to reach its full potential.

6. Acknowledgements

7. References

Baker, Thomas. "Libraries, languages of description, and linked data: a Dublin Core perspective". *Library hi tech*, 2012, v. 30, n. 1, pp. 116-133.
http://dx.doi.org/10.1108/07378831211213256

Berners-Lee, Tim. Information management: a proposal, 1989.
http://www.w3.org/History/1989/proposal.html

Berners-Lee, Tim. "Semantic web road map". *Design issues*, 1998, September.
http://www.w3.org/DesignIssues/Semantic.html

Berners-Lee, Tim. *Weaving the Web: The original design and ultimate destiny of the World Wide Web by its inventor.* San Francisco: Harper, 1999.

Berners-Lee, Tim. "Linked data". *Design issues*, 2006-07-27.
http://www.w3.org/DesignIssues/LinkedData.html

Berners-Lee, Tim. "The next Web of open, linked data". *TED*, February, 2009
https://www.ted.com/talks/tim_berners_lee_on_the_next_web.html

Brickley, Dan. "Semantic web, part 5". In: *Libraries, media & the semantic web*, hosted by the *BBC*, 28 March 2012, video.
http://www.youtube.com/watch?v=_-6mhdjE1XE&feature=player_embedded

Bush, Vannevar. "As we may think". *The Atlantic monthly*, 1945, July.
http://www.theatlantic.com/magazine/archive/1945/07/as-we-may-think/3881

Byrne, Gillian; **Goddard, Lisa**. "The strongest link: Libraries and linked data". *DLib magazine*, 2010, v. 16, n. 11-12.
http://www.dlib.org/dlib/november10/byrne/11byrne.html

Greenberg, Jane *et al. HIVE*: Helping Interdisciplinary Vocabulary Engineering. *Bulletin of the American Society for Information Science and Technology*, 2011, v. 37, n. 4.
http://www.asis.org/Bulletin/Apr-11/AprMay11_Greenberg_etAl.html

Kondert, Florian; **Schandl, Thomas**; **Blumauer, Andreas**. *Do controlled vocabularies matter? Survey results.* Semantic Web Company, Viena, 2011.
http://issuu.com/andreas_blumauer/docs/survey_do_controlled_vocabularies_matter_2011_june#download

Méndez, Eva. "Metadatos y tesauros: aplicación de XML/RDF a los sistemas de organización del conocimiento en intranets". In: *VII Jornadas españolas de documentación*, Bilbao (Spain), 2000, October, 19-20, pp. 211-219.
http://hdl.handle.net/10760/12698

Miller, Eric. Linked data and libraries [recorded by Micheline Westfall]. *Serials librarian*, 2011, v. 60, n. 1-4, pp. 17-22.
http://dx.doi.org/10.1080/0361526X.2011.556427

Oomen, Johan; **Baltussen, Lotte-Belice**; **Van Erp, Marieke**. "Sharing cultural heritage the linked open data way: why you should sign up". In: *Museums and the Web 2012.* San Diego, April, 11-14.
http://www.museumsandtheweb.com/mw2012/papers/sharing_cultural_heritage_the_linked_open_data

Pastor, Juan-Antonio. *Tecnologías de la web semántica.* Colección EPI. Barcelona: UOC, 2011.
http://www.elprofesionaldelainformacion.com/libros/tecnologias-web-semantica.html

Peset, Fernanda; **Ferrer-Sapena, Antonia**; **Subirats-Coll, Imma**. "Open data y linked open data: su impacto en el área de bibliotecas y documentación". *El profesional de la información*, 2011, marzo-abril, v. 20, n. 2, pp. 165-173.
http://dx.doi.org/10.3145/epi.2011.mar.06

Phipps, Jon; **Hillman, Diane**. "The Open Metadata Registry: an update". *Bulletin of the American Society for Information Science and Technology*, 2011, v. 37, n. 4.
http://www.asis.org/Bulletin/Apr-11/AprMay11_Phipps_Hillmann.pdf

Saorín, Tomás. "Cómo linked open data impactará en las bibliotecas a través de la innovación abierta". *Anuario ThinkEPI*, 2012, v. 6, pp. 288-292.
http://hdl.handle.net/10760/16913

Savolainen, Reijo; **Kari, Jarkko**. "User-defined relevance criteria in web searching". *Journal of documentation*, 2006, v. 62, n. 6, pp. 685-707.
http://dx.doi.org/10.1108/00220410610714921

Vatant, Bernard. LOV stories, Part 2: Gardeners and gatekeepers. *The wheel and the hub: Tracks in the knowledge commons* [blog], 22-03-2012.
http://blog.hubjects.com/2012/03/lov-stories-part-2-gardeners-and.html

W3C LLD XG. *Library Linked Data Incubator Group Final Re-*

port, 25 October 2011. Tomas Baker *et al.*, eds., 2011a.
http://www.w3.org/2005/Incubator/lld/XGR-lld-20111025

W3C LLD XG. *Library Linked Data Incubator Group: Datasets, value vocabularies, and metadata element sets*, 25 October 2011, Antoine Isaac *et al.*, eds., 2011b.
http://www.w3.org/2005/Incubator/lld/XGR-lld-vocabda taset-20111025

Yu, Holly; **Young, Margo**. "The impact of web search engines on subject searching in opac". *Information technology and libraries*, 2004, v. 23, n. 4, pp. 168-180.
http://dx.doi.org/10.1234/12345678

8. Relevant cited websites

Amalgame (Amsterdam alignment generation metatool)
http://semanticweb.cs.vu.nl/amalgame

DC, DCMI (Dublin Core, Dublin Core metadata initiative)
http://dublincore.org

DCMI-VMC (DCMI Vocabulary Management Community)
http://dublincore.org/groups/vocabulary-management

DataHub
http://thedatahub.org

EDM (Europeana data model)
http://pro.europeana.eu/edm-documentation

FOAF (friend of a friend)
http://www.foaf-project.org

HIVE (Helping interdisciplinary vocabularies engineering)
http://ils.unc.edu/mrc/hive

HIVE Project Wiki
https://www.nescent.org/sites/hive/Main_Page

HIVE-ES (Helping interdisciplinary vocabularies engineering-España). Project wiki
http://klingon.uc3m.es/hive-es/wiki

HIVE-ES server
http://klingon.uc3m.es:8080/home.html

KEA (keyphrase extraction algorithm)
http://www.nzdl.org/Kea

Linked Data
http://linkeddata.org

LOV project (linked open vocabularies)
http://labs.mondeca.com/dataset/lov

Marc4J java library
http://marc4j.tigris.org

NCBO-BioPortal
http://bioportal.bioontology.org

OAEI (Ontology alignment evaluation initiative)
http://oaei.ontologymatching.org

OBO (The open biological and biomedical ontologies)
http://obofoundry.org

OMR (Open metadata registry)
http://metadataregistry.org

OKFN (Open Knowledge Foundation)
http://okfn.org

Schema.org
http://schema.org
http://blog.schema.org

Semanticweb.com
http://semanticweb.com

SIOC (Semantically-interlinked online communities)
http://sioc-project.org

VOAF (Vocabulary of a friend)
http://labs.mondeca.com/vocab/voaf

W3C (World Wide Web Consortium)

– Linked data
  http://www.w3.org/standards/semanticweb/data
– MCF (meta content framework)
  http://www.w3.org/TR/NOTE-MCF-XML-970606
– Sparql
  http://www.w3.org/TR/rdf-Sparql-query
– RDF (resource description framework)
  http://www.w3.org/standards/techs/rdf
– RDFa (RDF attributes)
  http://www.w3.org/TR/xhtml-rdfa-primer
  http://rdfa.info
– RDFS (RDF schema)
  http://www.w3.org/TR/rdf-schema
– SKOS (simple knowledge organization system)
  http://www.w3.org/standards/techs/skos
– Semantic web
  http://www.w3.org/standards/semanticweb
– Vocabularies
  http://www.w3.org/standards/semanticweb/ontology