Ruiz-Calleja, A., Vega-Gorgojo, G., Asensio-Pérez, J. I., Gómez-Sánchez, E., Bote-Lorenzo, M. L., & Alario-Hoyos, C. (2014). SEEK-AT-WD: A Social-Semantic Infrastructure to Sustain Educational ICT Tool Descriptions in the Web of Data. *Educational Technology & Society, 17* (2), 321–332.

# SEEK-AT-WD: A Social-Semantic Infrastructure to Sustain Educational ICT Tool Descriptions in the Web of Data

**Adolfo Ruiz-Calleja*, Guillermo Vega-Gorgojo, Juan I. Asensio-Pérez, Eduardo Gómez-Sánchez, Miguel L. Bote-Lorenzo and Carlos Alario-Hoyos**

School of Telecommunication Engineering, University of Valladolid, Paseo Belén 15, Valladolid, Spain // adolfo@gsic.uva.es // guiveg@tel.uva.es // juaase@tel.uva.es // edugom@tel.uva.es // migbot@tel.uva.es // calario@it.uc3m.es

*Corresponding author

(Submitted February 28, 2013; Revised May 14, 2013; Accepted July 3, 2013)

**ABSTRACT**

There are several Information and Communication Technology (ICT) tool registries that support educators when searching ICT tools for their classrooms. A common problem in these registries is how their data is sustained, since educational descriptions of ICT tools are hard to create and maintain updated. This paper proposes SEEK-AT-WD, an infrastructure that aims at sustaining an educational ICT tool registry in the Web of Data following a social-semantic approach. Its key idea is to take advantage of the data already published in the Web to sustain a collection of ICT tool descriptions, as well as to enable the community of educators to enrich this collection sharing their experience using ICT tools. Following this approach, 6760 descriptions of educational ICT tools have been retrieved from the Web of Data to build an initial dataset. Moreover, the descriptions obtained from the Web are automatically updated without human intervention while more than a hundred tool descriptions have been enriched by educators. Finally, a search system and an annotation tool are presented to illustrate that educational applications can take advantage of SEEK-AT-WD.

**Keywords**

Educational ICT tools, infrastructure, Web of Data, Linked Data, Web 2.0

## Introduction

Information and Communication Technologies (ICTs) are massively employed in learning situations (i.e. scenarios designed for students to learn through the realization of learning tasks (Osuna-Gómez, 1999)). As learning situations may differ in many aspects, there are several characteristics of ICT tools that determine whether they are suitable for each specific learning situation (Kurti, Spikol, Milrad, & Flensburg, 2006). On the one hand, the tool functionality should support the tasks established for the educational situation (Vega-Gorgojo et al., 2008); e.g., a Wiki can be adequate to support students when writing and sharing documents. On the other hand, the tool should be compatible with the rest of the technology employed in the situation, as well as appropriate for the people who is going to use it (Gómez-Sánchez et al., 2009); e.g., a suitable tool for higher education may not fit in primary school. This diversity of purposes and situations makes the selection of ICT tools a critical step when designing learning situations (Gómez-Sánchez et al., 2009; Vega-Gorgojo et al., 2010; Vignollet Ferraris, Martel, & Burgos, 2008).

This technology selection requires educators to be aware of the capabilities of different software tools that can potentially be employed for educational purposes (Vega-Gorgojo et al., 2008) (this paper calls such tools "educational ICT tools"). In order to inform educators several educational organizations maintain ICT tool registries. Some examples are Sisoft (Universidad Complutense de Madrid, 2013) or Ontoolsearch (Vega-Gorgojo et al., 2010). These registries contain data about ICT tools and some of them structure it using semantic technologies, thus providing higher precision and recall in their results when educators search for tools (Vega-Gorgojo et al., 2010). A common issue to these registries is how to create and sustain the data they contain, which is a well-known problem in the educational domain (Bateman, Brooks, & McCalla, 2006). Some of these registries, such as Ontoolsearch, follow a traditional approach where an administrator is responsible for sustaining the dataset. However, this approach requires the organization to assume the whole effort of creating and updating the data. Others adopt the Web 2.0 principles (O'Reilly, 2005) to sustain their respective datasets, involving users in the creation and maintenance of the content. For example, tool providers publish descriptions of their ICT tools in the ROLE Widget Store (Govaerts et al., 2011). Another example is (CoolToolsForSchools, 2013), where educators not only retrieve tool descriptions, but they also publish and update them. Nonetheless, these specific search facilities are still limited because their communities are isolated: the data published in a registry cannot be retrieved from another. In addition, they suffer

from the *cold start problem* (Maltz & Ehrlich, 1995): at the beginning the utility of the registry is limited since it contains very few tool descriptions; hence, educators are not motivated to enrich the registry since they do not perceive its utility.

In our research, we address this data sustainability problem by taking advantage of the open data already published in the Web of Data (Heath & Bizer, 2011). The Web of Data is a recent proposal that foresees a Web-scale federation of datasets that follow a common methodology for publishing their information: the Linked Data principles (Berners-Lee, 2006). Currently, thousands of data providers, including some educational institutions as (University of Southampton, 2013), publish their information on the Web of Data, which is growing very quickly (Heath & Bizer, 2011, chap. 3). Our key idea is to automatically import ICT tool descriptions from third-party updated repositories of the Web of Data and relate them to a vocabulary understandable by educators; thus, the effort of creating and updating the dataset would be significantly reduced. Our previous work (Ruiz-Calleja, Vega-Gorgojo, Asensio-Pérez, et al., 2012; Ruiz-Calleja, Vega-Gorgojo, Gómez-Sánchez, et al., 2012) shows that retrieving descriptions of educational ICT tools from the Web of Data is feasible, since several thousands of them were obtained, and can well be employed to satisfy educators' information needs; nonetheless, the educational information that can be inferred from the Web of Data is limited and its support to educational applications can be improved (Ruiz-Calleja, Vega-Gorgojo, Asensio-Pérez, et al., 2012). Our approach to increase the quality of this data is to combine the Web 2.0 and the Linked Data principles in a social-semantic approach (Mikroyannidis, 2007). Thus, an initial dataset of educational ICT tools is generated and periodically updated from the Web of Data, while the community of educators can enrich it by publishing their experience using tools in the classroom. Also note that this dataset will count on a big collection of tool descriptions from the very beginning and thus the cold-start problem can be overcome.

In order to reach this aim, this paper proposes SEEK-AT-WD (Support for Educational External Knowledge About Tools in the Web of Data), a social-semantic infrastructure that collects educational descriptions of ICT tools. SEEK-AT-WD is used to sustain a dataset publicly available in the Web of Data, so its tool descriptions can be freely exported by third parties or directly used to build applications that leverage the benefits of SEEK-AT-WD. By means of an envisioned scenario and two prototype applications, this paper illustrates such benefits and how this infrastructure enables educators to interact with the information about ICT tools available in the Web of Data.

The rest of the paper is structured as follows: first, the state of the art of Linked Data in education is overviewed, with especial attention to the vocabularies and datasets of ICT tools. Next, SEEK-AT-WD is described, including its data model and its software architecture. Then, SEEK-AT-WD is evaluated analyzing the data obtained from the Web of Data and illustrating its consumption by educational applications. Finally, the conclusions of this research work are summed up.

## Related work: Educational linked data sources of ICT tools

During the last decade, significant efforts have been made to facilitate the interoperability of educational datasets. Specifically, several educational vocabularies (i.e. Learning Object Metadata (LOM) (Hodgins, 2002)) have been standardized and their use is now widespread in this domain. Additionally, the Semantic Web technologies (Berners-Lee, Hendler, & Lassila, 2001) were exploited as a way to reduce the interoperability effort of the educational data published in the Web (Devedžic, 2006, chap. 1). However, the impact of Linked Data on education is still limited (Dietze et al., 2013) although it has become the *de facto* standard to publish data on the Web. Current Linked Data efforts in the educational domain mainly focus on the publication of already-existing data on the Web. For example, (University of Southampton, 2013) and (Open University, 2013) are two pioneer projects carried out by educational institutions for the publication of Linked Open Data. However, most of the data currently published by educational institutions focuses on administrative information and not on information that can directly support educators in their classrooms. Other institutions that currently publish Linked Open Data are public libraries, such as the German National Library of Economics (Neubert, 2009) or the Open Library (Open Library, 2012), a social-semantic book registry in the Web of Data that contains over a million books.

Despite these efforts, there are very few proposals that consume Linked Open Data for educational purposes, although some examples can be seen at (Dietze et al., 2013) and (Open Library, 2012). In this respect, (Zablith d'Aquin, Brown, & Green-Hughes, 2011) detects a threefold benefit for educational applications: It reduces the

effort of sustaining data; it decreases the user effort to find relevant information; and it connects educational communities in a global data space. Our proposal looks for educational ICT tool registries to reach these benefits. Thus, current datasets of educational ICT tools available on the Web of Data are presented next.

### Educational ICT tool descriptions in the Web of Data

To the best of the authors' knowledge, two educational datasets of ICT tools can be found in the Web of Data: Ontoolsearch and ROLE Widget Store. Ontoolsearch is a specialized search facility that describes about 100 tools using an educational-specific vocabulary defined by Ontoolcole ontology (Vega-Gorgojo et al., 2008). This ontology considers a taxonomy of ICT tools based on the educational tasks they support. Each tool is described by stating its *tool type* (i.e., "drawing tool"), the *tasks* it supports (i.e., "collaborative edition of multimedia documents") and the *artifacts* it manipulates (i.e., "stores .mpg video documents"). Thus, using Ontoolcole it is possible to formally express complex descriptions, such as "whiteboard that allows the collaborative edition of images, as well as chatting." Nonetheless, Ontoolcole cannot express the educational experience using ICT tools in the classroom.

ROLE Widget Store (Govaerts et al., 2011) is a recommender system of widgets developed under the ROLE (Responsive Open Learning Environment) Project which trusts in a Web 2.0 approach to sustain its dataset. It recently moved part of its data to the Web of Data using common vocabularies (e.g., DublinCore), as well as the Role Vocabulary (Govaerts et al., 2011). The core of the Role Vocabulary is somehow similar to Ontoolcole since it defines some tool categories and some supported educational activities. However, the Role Vocabulary does not classify tool categories or educational activities in a taxonomy, nor it defines any relationship between different concepts. Additionally, it does not allow to formally describe the experience of using these tools in the classrooms.

There are some other educational registries of ICT tools in the Web of Data, but they only contain a small subset of the tools that are currently employed in the classrooms. Besides, other educational ICT tool descriptions can be found in cross-domain datasets (Heath & Bizer, 2011, chap. 3), although they are not specifically published for educational purposes. Among them, the most popular is DBpedia (Auer et al., 2007), which mirrors part of Wikipedia to the Web of Data. Others are Freebase (Google, 2011), OpenCyc (Cycorp Inc., 2013) or Factforge (Ontotext AD, 2012), which is especially interesting since it collects the data from the previous ones and offers it from a single endpoint. These cross-domain data sources provide descriptions of thousands of ICT tools potentially useful in educational scenarios. They describe tools specifying their *genre* (e.g., "Word processor"), several administrative data, a textual description and links to other sources where more information about the tool can be found. However, they are not related to an educational vocabulary, so further processing is required to make them useful for applications that manage educational abstractions.

## SEEK-AT-WD

This section presents SEEK-AT-WD, a social-semantic infrastructure that sustains a registry of educational ICT tools in the Web of Data to overcome the sustainability problems that have been previously detected. First, an envisioned usage scenario that SEEK-AT-WD should support is presented. This scenario will be used to collect the requirements that will guide the design and development of SEEK-AT-WD information model and its software architecture, as this section later presents.

### Envisioned usage scenario

Marie is a teacher in engineering who is designing a peer-review activity: as homework her students will write a document in pairs and, after that, each pair will review the documents written by three other pairs. Marie does not know which ICT tools could support this activity, so she uses an ICT tool search system to discover *tools that allow a group to write text, and that have been used by other to support peer-review activities in blended learning scenarios.* Several months before, someone published in Wikipedia a description of *Google Docs*, stating that it is a c*ollaborative text editor* and a *file hosting service*, and a description of *Wikispaces*, stating that it is a *proprietary wiki software*. Furthermore, some educators used an annotation tool to publish in the Web of Data their experiences using these tools, reporting that they used them satisfactorily to support a *peer-review* activity in *blended* scenarios.

These two tools can support Marie's students in the peer-review activity and their descriptions are publicly available on the Web of Data; for this reason, when Marie submits the abovementioned request, she obtains *Wikispaces* and *Google Docs*. Marie reads these tool descriptions and she chooses *Google Docs* to support her activity.

Since Marie's experience using *Google Docs* to support her *peer-review* activity is very positive, she thinks that it may be interesting to use this same tool to support written debates: she publishes some documents and she writes a question in each one, expecting her students to discuss on-line about this question. Once the activity is finished she employs an annotation tool to describe her experience with *Google Docs*, and share it so that it can help others in selecting the right technology for their educational activities.

This scenario could be achieved by a search facility that allows educators to make queries using educational abstractions (Vega-Gorgojo et al., 2008) while relying on an infrastructure that gathers descriptions of educational ICT tools from different sources of the Web of Data (functionality F1 in Figure 1). As the information retrieved from the Web of Data does not describe tools using educational abstractions, the descriptions obtained should be converted (F2) relating them to an educational vocabulary. In order to allow the data consumption by external applications, the infrastructure should publish its information on the Web of Data with an open license (F3). Further, the infrastructure should also allow educational applications to submit data (F4), thus enabling the community of educators to enrich its dataset out of their educational experience using ICT tools. Finally, as educators are supposed to search and annotate tools through interactive applications, the response time of the infrastructure should not be too high. Figure 1 graphically sums up the data flow of this scenario.

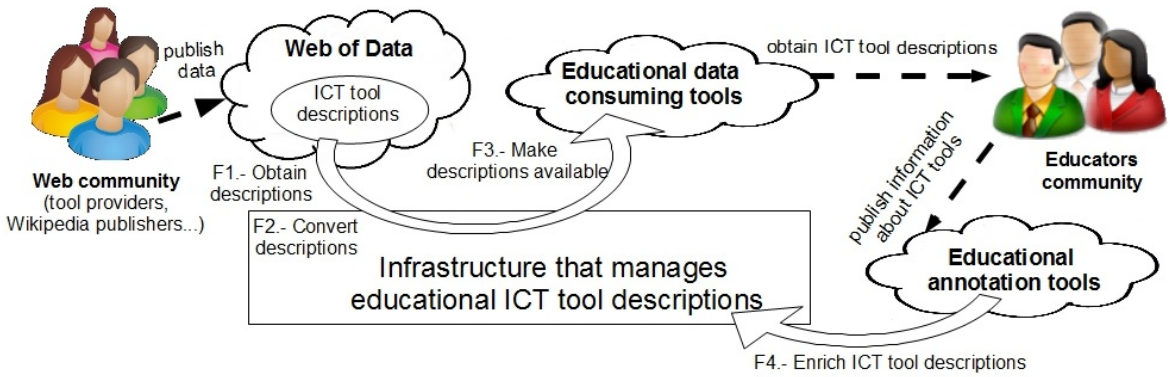


*Figure 1.* Educational consumption and publication of information about ICT in the Web of Data

Note that this approach has several benefits regarding the sustainability of the data. First, the cold-start problem would be solved since the Web of Data already contains information that can be used to support educators when discovering educational ICT tools, as evaluated in a previous research work by the authors (Ruiz-Calleja, Vega-Gorgojo, Asensio-Pérez, et al., 2012). Second, this approach takes advantage of the data already published by the Web community, reusing it for educational purposes. Finally, the educators' community can share more information obtained from their experience using ICT tools. The infrastructure that supports this scenario is called SEEK-AT-WD. The rest of this section describes its data model and its software architecture.

**SEEK-AT-WD data model**

In order to support the scenario described in the previous subsection, SEEK-AT-WD needs an information model whose objective is twofold: to describe educational ICT tools and to allow the community of educators to review them including educational information of their use. As previously seen, ICT tools can be described by stating their tool type, the tasks they support and the artifacts they manage. In addition, several authors (Jorrín-Abellán, & Stake, 2009; Kurti, Spikol, & Millard, 2008) agree that the educational context where technology is employed has a huge influence on its selection by educators. For this reason, the contexts where tools have already been used are important when educators discover and select tools (Gómez-Sánchez et al., 2009; Kurti, Spikol, & Millard, 2008).

SEEK-AT-WD information model is defined by an ontology called *SEEK Ontology*, which is designed and implemented following the methodology On-To-Knowledge (Staab, Studer, Schnurr, & Sure, 2001) and the best practices to publish data on the Web (Heath & Bizer, 2011, chap. 3-4). These methodologies and best practices highly recommend reusing vocabularies already employed by third parties since it reduces the development effort and facilitates the federation of datasets. As Ontoolcole is the only ontology that defines a taxonomy of tool types, educational tasks and artifacts, these taxonomies are taken for the description of ICT tools in SEEK Ontology (see (Ruiz-Calleja, Vega-Gorgojo, Alowisheq, Asensio-Pérez, & Tiropanis, 2012) for more details). On the other hand, the Review Vocabulary (Heath & Motta, 2007) is used as a base to describe reviews since it is commonly employed for this purpose in the Web of Data. The Review Vocabulary is a very simple ontology that defines reviews related to any reviewed entity and to the person who published the review. Nonetheless, SEEK Ontology also relates these reviews to the educational contexts where ICT tools were employed, thus collecting the domain-specific characteristics of the use of ICT tools. Figure 2 represents the relationships of the main concepts defined by SEEK Ontology.

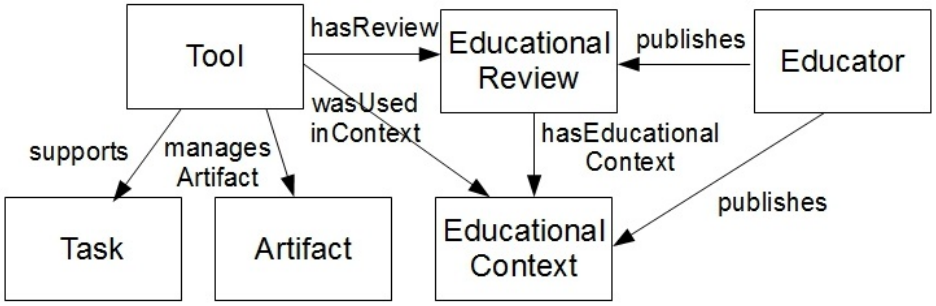


*Figure 2.* Main concepts of SEEK Ontology and their relationships

SEEK Ontology describes the concepts reused from other vocabularies with the same parameters as they are described in Ontoolcole and Review Vocabulary. However, no ontology that defines the characteristics of an educational context was found. Therefore, several information sources were analyzed, including educational literature (e.g., Kurti, Spikol, & Millard, 2008), other related ontologies (e.g., LOM), other educational applications currently in use (e.g., ROLE Project) and the analysis of educational context descriptions written by educators in real situations. After this analysis, the concepts that were shared among different source were included in SEEK ontology. Table 1 shows and exemplifies the most important concepts defined by SEEK Ontology to describe *Google Docs*, and Marie's review and educational context.

*Table 1.* Part of the description of a tool, an educational review and an educational context with SEEK Ontology

| Tool: Google Docs | | Review: 0000 | |
|---|---|---|---|
| Label | Google Docs | Publisher | Marie |
| Description | "Google Docs is a free office suite…" | Text | "A very useful tool that…" |
| Type | Asynchronous text editor | Rating | 4 |
| Supports task | Writing | Has educational context | Educational context: ComAr |
| Manages artifact | Text document | Educational context: ComAr | |
| Operating System | Web application | Subject | Computers Architecture |
| License | Proprietary software | Area of knowledge | Engineering |
| Developer | Google | Learning goal | Requirement analysis |
| Has review | Review: 0000 | Teaching technique | Peer review |
| Was used in context | Educational context: ComAr | Delivery mode | Blended |

**SEEK-AT-WD software architecture and prototype implementation**

The aim of SEEK-AT-WD software architecture is to sustain a registry of educational ICT tools (F3) retrieving information from the Web of Data (F1) and educational annotation tools (F4). In order to design such infrastructure, the crawling pattern (Hartig & Langegger, 2010) is followed. This pattern is recommended for applications that integrate data from different sources while allowing complex queries with low response time (Heath & Bizer, 2011, chap. 6). It considers a crawler that traverse links in the Web of Data to obtain useful information, which is

automatically cleaned and stored in a data cache. Thus, the tasks of retrieving the data (F1) and aligning it to a common vocabulary (F2) are separated from offering this data (F3). As a disadvantage, the cache may contain stale data, so the Web of Data needs to be periodically crawled.

Figure 3 shows SEEK-AT-WD architecture. It consists of a *crawler* and a *data repository* (SEEK-KB) with its *corresponding interfaces to add and retrieve data*. The crawler plays a central role since it is responsible for gathering ICT tool descriptions from cross-domain linked data sources (F1) and aligning them to SEEK Ontology (F2) using techniques of ontology mapping (Choi, Song & Han, 2006). The current version of the crawler retrieves data from DBpedia and Factforge repositories, as well as any other dataset linked by these two. However, its design is extensible to retrieve information from other linked datasets if needed.

The ontology mapping techniques require relating several *DBpedia genre* concepts to SEEK Ontology. For example, the DBpedia category "*Collaborative real-time editor*" is related to the tool categories "S*ynchronous Text Editor*" and "*Text Viewer*" of SEEK Ontology, as well as the tasks of "*Writing*" and "*Synchronous Communication*" and the artifact "*Text Document.*" This way, the crawler infers the relationships to SEEK Ontology for any tool description classified by DBpedia as "*Collaborative real-time editor*", such as *Google Docs.* For further details see (Ruiz-Calleja, Vega-Gorgojo, Gómez-Sánchez et al., 2012).

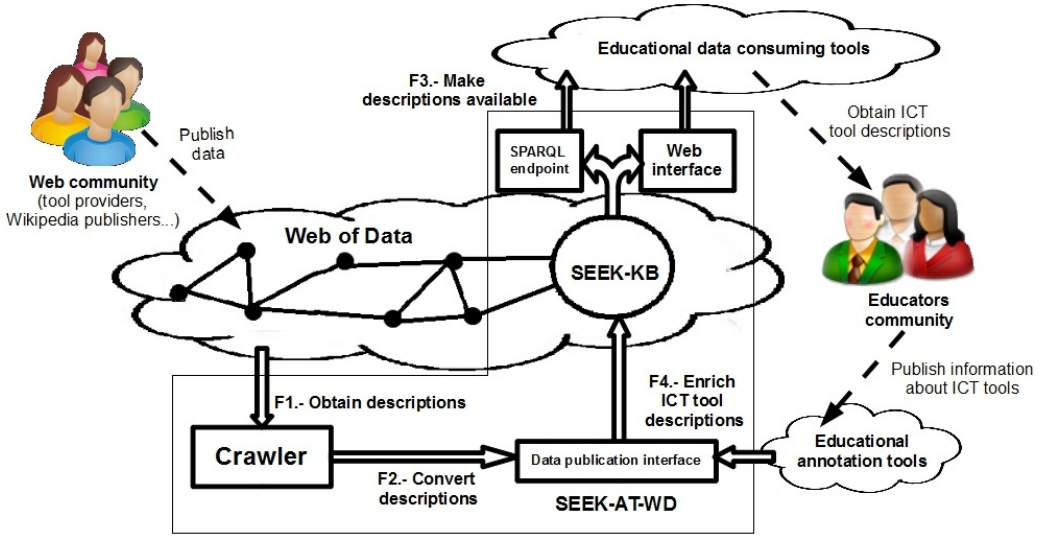


*Figure 3.* SEEK-AT-WD architecture

The formalization of these mappings requires several iterations and the definition of 114 relationships between DBpedia and SEEK Ontology concepts. However, once the mappings are formalized, SEEK Ontology is related to the ICT tools that are published in DBpedia and those that will be published in the future, either by DBpedia or by any other Linked Data source that uses the same vocabulary. In addition, this approach allows crawling the Web of Data: Once DBpedia mappings are defined, the relationships between SEEK Ontology and other linked vocabularies can be automatically inferred following the links that are published on the Web of Data. In fact, 599 additional tool categories that are also used by DBpedia and Factforge were automatically retrieved from different sources of the Web of Data. These additional categories are enough to formalize the relationships between DBpedia and Factforge, and consequently, between SEEK Ontology and Factforge. It can be seen that this architecture is extensible and can easily import data from other interesting sources – including educational-specific ones - that may appear in the future.

The DBpedia and Factforge mappings, as well as all data extracted by the crawlers, are published at *http://www.gsic.uva.es/seek/dataset/*. This data has also been stored in SEEK-KB, which can be reached at: *http://seek.rkbexplorer.com/*. SEEK-KB offers a SPARQL endpoint (World Wide Web Consortium, 2008) and a Linked Data interface, as shown in Figure 4; thus, the data can be browsed and queried in a standard way. Moreover, it provides a data publication interface that is used by the crawler and other applications to enrich or update the data it contains; thus, the community of educators can publish more information about ICT tools following a social approach.

As an example of use, Marie's information request "*tools that allow a group to write text, and that have been used by other to support peer-review activities in blended learning scenarios*" can be coded with the SPARQL query shown on Figure 4 (left); "*Google Docs*" and "*Wikispaces*" are obtained if this query is launched.

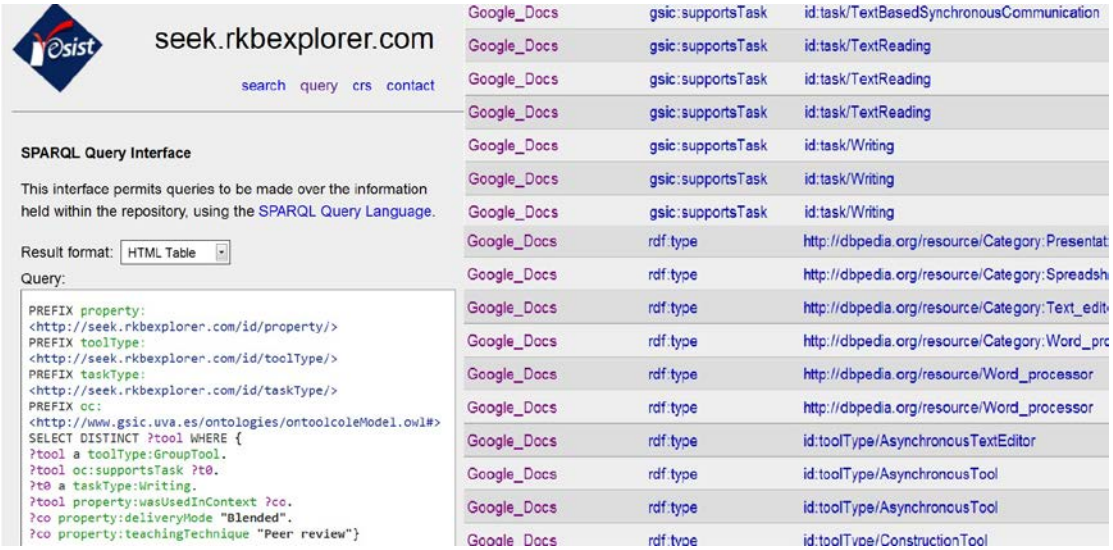


*Figure 4.* Snapshots of SEEK-AT-WD interfaces: SPARQL interface (on the left) and Linked-Data interface (on the right)

## SEEK-AT-WD evaluation

Once SEEK-AT-WD infrastructure is developed, it is needed to evaluate whether it provides the functionality required to support the envisioned usage scenario. This evaluation first focuses on SEEK-AT-WD crawler, analyzing the tool descriptions it gathers from the Web of Data. Their quantity, their relationship with SEEK Ontology concepts and their similarity to the descriptions that a human would create are taken into account. Then, SEEK-AT-WD data interfaces are assessed by means of two example applications that illustrate how end-user applications interact with the data stored by SEEK-AT-WD. This analysis will allow discussing -before making extensive use of it with a collection of educational applications- if SEEK-AT-WD supports the envisioned usage scenario overcoming the cold-start problem that other datasets suffer.

### Obtaining educational ICT tool descriptions from the Web of Data

The first evaluation task analyzes the tool descriptions inferred by SEEK-AT-WD crawler out of the information retrieved from the Web of Data. First, the descriptions obtained are quantified regarding their distribution in relation to SEEK Ontology concepts. After that, the quality of these descriptions is discussed analyzing how an expert on SEEK Ontology (i.e. someone who deeply understands the ontology) would classify 100 of the descriptions inferred by the crawler.

As on February 2013, 6760 tool descriptions were retrieved and almost all the range of tool types was covered. Specifically, SEEK Ontology defines 46 tool types, 11 of which obtained more than 350 results; 12 categories obtained between 130 and 350 tool descriptions, while another 12 categories obtained between 27 and 130 tool descriptions. Finally, 11 categories were related to less than 27 tool descriptions. It is also important to mention that three concepts of SEEK Ontology, such as *Questionnaire Manager*, could not be related to any concept of an external linked dataset vocabulary, since no synonyms were found in other vocabularies used in the Web of Data; hence, no tool descriptions were imported for these concepts. Table 2 shows some of the tool types defined by SEEK Ontology and the number of tool descriptions extracted by the crawlers related to them.

*Table 2.* Number of descriptions retrieved from the Web related to some categories of SEEK Ontology

| Tool type | #tools | Tool type | #tools | Tool type | #tools |
|---|---|---|---|---|---|
| Representation tool | 3565 | Information management tool | 1904 | Construction tool | 1479 |
| Audio player | 181 | Blog | 229 | Text editor | 473 |
| Video player | 154 | Document repository | 776 | Wiki server | 314 |
| Group tool | 1956 | Processing tool | 1268 | Electronic calendar | 2 |
| Chat | 128 | Compiler | 1005 | Concept map tool | 54 |
| Audio conference tool | 27 | Simulator | 237 | Slide composer | 44 |

As a general rule, the more specific a category is, the less amount of tool descriptions are obtained from the Web of Data; for example, the number of concept map tools retrieved is much lower than the number of document repositories. It is also important to note that not all of the data imported refers to educational ICT tools. For example, out of the 314 descriptions that the crawler relates to *Wiki server* there are many that do not refer to ICT tools, but to other concepts related to Wiki servers. Despite this noisy data, which is a well-known problem when reusing data from the Web (Heath & Bizer, 2011, chap. 6), our previous work showed that these descriptions can well be employed to support educators when discovering ICT tools (Ruiz-Calleja, Vega-Gorgojo, Asensio-Pérez, et al., 2012). Additionally, this noise can be expected to reduce over time since the community of educators can clean SEEK-AT-WD dataset.

Regarding the quality of the data, 100 descriptions were randomly selected and their classification was completed by an expert on SEEK Ontology. Thus, the descriptions automatically inferred by SEEK-AT-WD are compared to the ones that a human would publish. Table 3 quantifies how these 100 descriptions were related to the three taxonomies defined by SEEK Ontology: tool types, educational tasks supported and artifacts managed.

*Table 3.* Classification of 100 tool descriptions retrieved from the Web of Data by SEEK-AT-WD

| Characteristic | #descriptions |
|---|---|
| Exact categorization | 66 |
| Lack of some tool types | 23 |
| Lack of some tasks supported | 26 |
| Lack of some artifacts managed | 14 |
| Erroneous data | 0 |

66 out of 100 descriptions retrieved by the crawlers are classified in the same categories as the expert would. Typically, these are descriptions of tools that only support the generic tasks defined by the mappings (e.g. *Google Mail*, which allows to *communicate* and to *send e-mail messages*). On the other hand, the crawler could not infer all the tasks supported by 26 tools. An example is *Windows Live Messenger*, which is described as a *chat client*, but the crawlers did not infer some functionalities that were described by the expert, such as *video communication* or *collaborative drawing* because this tool is not classified by DBpedia as a videoconference tool nor as an image editor. Further, 23 descriptions include all the educational tasks that the tools support, although some information about the artifacts they manage is missing. An example is *FreeMind*, which is properly classified as a concept map tool that manages models, but the crawler could not infer that it also manages images. Finally, it is very positive that the crawlers did not infer any erroneous data in the 100 tool descriptions analyzed. Again, these results are very satisfactory, but the tool descriptions retrieved from the Web of Data can be improved by the community of educators if a social annotation tool is available.

**Current applications that interact with SEEK-AT-WD data**

The data gathered by SEEK-AT-WD can be retrieved and enriched through its data access and data publication interfaces. However, as previously discussed, educators cannot be expected to manipulate these interfaces. Instead, end-user applications are required to support educators the interaction with the data contained in SEEK-AT-WD. In order to evaluate the feasibility of such applications, U-Seek and We-Share are presented as two example applications that allow educators to retrieve and publish information about educational ICT tools in SEEK-AT-WD.

U-Seek (Vega-Gorgojo, Ruiz-Calleja, Asensio-Pérez, & Jorrín-Abellán, 2012) is a web application that provides a graphical user interface to obtain information from SEEK-KB. It supports the creation of queries adding restrictions

about the functionality of the tools obtained and about the contexts where they have been used. When an educator launches a query, U-Seek creates a SPARQL query accordingly, submits it to SEEK-AT-WD and presents the results obtained. As an example, Figure 5 shows the formulation of the query that Marie would submit in the envisioned usage scenario. U-Seek is now available at: *http://www.gsic.uva.es/seek/socialuseek/*

U-Seek is an interactive application that is directly used by educators. For this reason, the response time of SEEK-AT-WD to submit the results of a query should be lower than a few seconds. In this regard, 1637 queries were launched with U-Seek before February 2013; the arithmetic mean of SEEK-KB response time when answering these queries was 608 milliseconds, while its standard deviation is 518 milliseconds. It can be seen that SEEK-KB response time is low enough as to satisfy the requirements of interactive applications that take advantage of its data. In fact, none of its users complained about it when being asked.

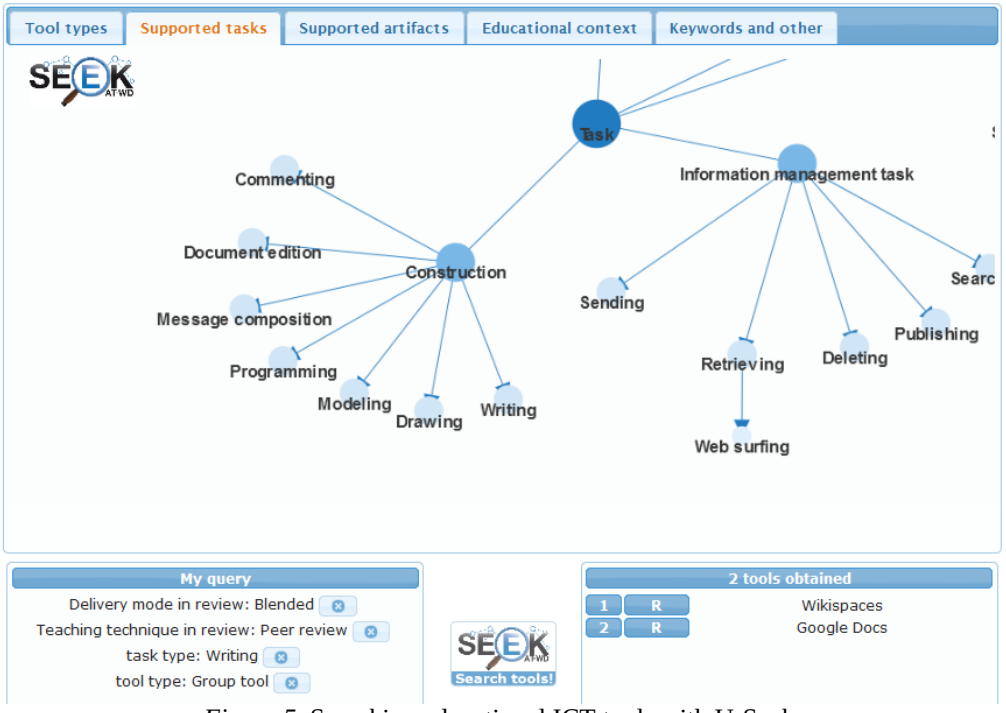


*Figure 5.* Searching educational ICT tools with U-Seek

We-Share is a social annotation application. It is a web application that is currently employed by a community of educators to submit data to SEEK-AT-WD. It provides an interface based on forms that can be easily manipulated by educators to publish descriptions of ICT tool, educational contexts or educational reviews. When an educator adds or modifies data using We-Share, this application automatically relates that data to SEEK Ontology and submits it to SEEK-AT-WD. We-Share is available at http://seek.cloud.gsic.tel.uva.es/weshare/.

Between January and February 2013, We-Share users published 27 educational contexts and 94 educational reviews; moreover, they published or updated 116 ICT tool descriptions. All the data created by We-Share is reachable at http://seek.cloud.gsic.tel.uva.es/upload/ and is currently published by SEEK-AT-WD (and hence queried by U-Seek) in conjunction with the data retrieved from the Web of Data. For example, the Google Docs description presented in Figure 4 (http://seek.rkbexplorer.com/id/tool/Google_Docs) contains data obtained by SEEK-AT-WD crawler, as well as data created with We-Share. It can be seen that all the information provided is coherently combined.

## Discussion

Evaluation results show that a big collection of ICT tool descriptions can be automatically obtained from the Web of Data and related to SEEK Ontology. Thus, a registry of educational ICT tools can be automatically created and kept updated, since the data sources where information is retrieved are continuously updated (Auer et al., 2007). As the

tool descriptions gathered from the Web of Data cover almost the whole range of concepts defined by SEEK Ontology, they can well be used to overcome the cold start problem that social registries suffer. Further, as tool descriptions are semantically structured, the advantages of semantic searches for this domain (Vega-Gorgojo et al., 2010) – such as more accurate results - are still present.

It was also shown that the tool descriptions inferred by SEEK-AT-WD crawler are classified in a similar way as a human expert would do, although a third of the descriptions analyzed lack of some relationships to SEEK Ontology concepts. In this regard, the social facet of SEEK-AT-WD is remarkable because it enables the educational community to complete these descriptions, to add new ones and to share their experience about the use of ICT tools in specific educational contexts. As these contexts are formalized with an ontology, semantic searches can use them to filter results. This way, Marie not only can ask for tools that allow a group to write, but also she can restrict the results to those that were previously employed in blended scenarios to support peer-review activities. Thus, she obtains less, but more relevant results.

Finally, by mean of two example applications it was shown that SEEK-AT-WD interfaces allow interactive end-user applications to consume and enrich the data it contains. Hence, the envisioned usage scenario is supported by SEEK-AT-WD: educational queries can be submitted to obtain descriptions of ICT tools that someone published in the Web of Data in conjunction with educational reviews of these same tools. Moreover, this same infrastructure allows educators to publish in the Web of Data information about their experience using ICT tools in the classroom. Finally, only a few tens of educational reviews have already been collected by We-Share yet, but currently there is a community of educators using it, so we expect more reviews to be published in the near future. Therefore, SEEK-AT-WD utility will increase since more educational-specific knowledge will be available.

## Conclusions and future work

This paper presented SEEK-AT-WD, a social-semantic infrastructure that sustains an educational ICT tool registry in the Web of Data. SEEK-AT-WD currently retrieves the educational ICT tool descriptions linked by DBpedia and Factforge, but it can be easily extended to obtain information from other datasets of the Web of Data. In February 2013 SEEK-AT-WD gathered 6760 updated descriptions of ICT tools potentially useful for educational purposes, covering almost all the range of concepts defined by SEEK Ontology. Moreover, as the Web of Data is periodically crawled, these tool descriptions are automatically updated and new ones can be discovered. Further, the quality of the descriptions obtained is satisfactory since two thirds of the tools were classified as an expert would, and the others did not contain erroneous data, although some of the tool characteristics were missing. This information retrieved from the Web of Data can well be used to overcome the cold-start problem that social registries suffer while it can be further enriched by the educational community, exploiting the social facet of SEEK-AT-WD.

Two educational applications illustrate the feasibility of allowing educators to interact with SEEK-AT-WD. They enable the educational community to publish and consume information about the use of ICT tools in the classroom in an interactive way. Moreover, as this information is published as Linked Open Data, third parties can make use of it to develop their own applications. Thus, the potential impact of SEEK-AT-WD data is much higher than other isolated data sources since other educational communities can take advantage of it.

Near future work will focus on an evaluation of the data created by We-Share. Specifically, it will be very interesting to discuss how the data created by a community of educators can complement the information about ICT tools automatically obtained from non-educational data sources of the Web of Data. This same evaluation will also be used to gather information about the use of SEEK Ontology by its end-users. Then, the efforts will focus on giving a better support to the educators through the development of a collection of applications that take advantage of SEEK-AT-WD data. For example, semantic technologies can be exploited to give multilingual support or to facilitate the evolution of SEEK Ontology through the use of folksonomies; on the other hand, social information can be further exploited to recommend educators combinations of tools that can support a given context.

## Acknowledgments

This work has been partially funded by the Spanish Ministry of Economy and Competitiveness (Project TIN2011-28308-C03-02), by the Government of Castilla y Leon, Spain (Project VA293A11-2) and the postdoctoral fellowship Alianza 4 Universidades.

## References

Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., & Ives, Z. (2007). DBpedia: A nucleus for a web of open data. *Proceedings of the 6th International Semantic Web Conference* (pp. 722–735). Busan, Korea: Springer.

Bateman, S., Brooks, C., & McCalla, G. (2006). Collaborative Tagging Approaches for Ontological Metadata in Adaptive E-Learning Systems. *Proceedings of the 4th International Workshop on Application of Semantic Web Technologies for E-Learning* (pp. 3–12). Retrieved from http://www.win.tue.nl/SW-EL/2006/proceedings/SWEL-2006-proceedings.pdf

Berners-Lee, T. (2006, July 27). Linked data - Design issues. Retrieved February 2013, from http://www.w3.org/DesignIssues/LinkedData.html

Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic web. *Scientific American, 284*(5), 34–43.

Choi, N., Song, I. Y., & Han, H. (2006). A survey on ontology mapping. *ACM Sigmod Record*, *35*(3), 34–41.

CoolToolsForSchools (2013). Retrieved February 2013, from http://cooltoolsforschools.wikispaces.com/

Cycorp Inc. (2013). Retrieved February 2013, from http://www.cyc.com/platform/opencyc

Devedžic, V. (2006). *Semantic web and education.* New Your, NY: Springer.

Dietze, S., Sanchez-Alonso, S., Ebner, H., Yu, H. Q., Giordano, D., Marenzi, I., & Pereira-Nunes, B. (2013). Interlinking educational resources and the Web of Data - A survey of challenges and approaches. *Electronic Library and Information Systems*, *47*(1).

Google (2011). Freebase. Retrieved February 2013, from http://www.freebase.com/

Gómez-Sánchez, E., Bote-Lorenzo, M. L., Jorrín-Abellán, I. M., Vega-Gorgojo, G., Asensio-Pérez, J. I., & Dimitriadis, Y. (2009). Conceptual framework for design, technological support and evaluation of collaborative learning. *International Journal of Engineering Education*, *25*(3), 557-568.

Govaerts, S., Verbert, K., Dahrendorf, D., Ullrich, C., Schmidt, M., Werkle, M., … Law, E. L.-C. (2011). Towards responsive open learning environments: The ROLE interoperability framework. *Towards Ubiquitous Learning* (Vol. 6964, pp. 125–138). Heidelberg, Germany: Springer Berlin.

Hartig, O., & Langegger, A. (2010). A database perspective on consuming Linked data on the web. *Datebank-Spektrum*, *10*(2), 57-66.

Heath, T., & Bizer, C. (2011). *Linked data: Evolving the web into a Global Data Space*. doi:10.2200/S00334ED1V01Y201102WBE001

Heath, T., & Motta, E. (2007). Revyu. com: A reviewing and rating site for the web of data. *Lecture Notes in Computer Science*, *4825*, 895-902.

Hodgins, W. (2002). Draft standard for learning object metadata. Retrieved February 2013, from http://ltsc.ieee.org/wg12/files/LOM_1484_12_1_v1_Final_Draft.pdf

Jorrín-Abellán, I. M., & Stake, R. E. (2009). Does ubiquitous learning call for ubiquitous forms of formal evaluation?: An evaluand oriented responsive evaluation model. *Ubiquitous Learning: An International Journal*, 1(3), 71-82.

Kurti, A., Spikol, D., & Millard, M. (2008). Bridging outdoors and indoors educational activities in schools with the support of mobile and positioning technologies. *International Journal of Mobile Learning and Organisation*, *2*(2), 166–186.

Kurti, A., Spikol, D., Milrad, M., & Flensburg, P. (2006). Increasing the value of information: Putting content in context: Is that enough? *Proceedings of the 29th Information Systems Research Seminar in Scandinavia*. Retrieved from http://www.itu.dk/iris29/IRIS29/12-5.pdf

Maltz, D., & Ehrlich, K. (1995). Pointing the way: Active collaborative filtering. *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 202–209). San Jose, USA: ACM.

Mikroyannidis, A. (2007). Toward a social semantic web. *Computer*, *40*(11), 113–115.

Neubert, N. (2009). Bringing the "thesaurus for economics" on to the web of Linked Data. *Proceedings of the Workshop on Linked Data on the Web.* Retrieved from http://ceur-ws.org/Vol-538/ldow2009_paper7.pdf

Ontotext AD. (2012), FactForge - Fast track to the center of the data web. Retrieved February 2013, from http://www.ontotext.com/factforge

Open Library. (2012). Open library data APIs. Retrieved February 2013, from http://openlibrary.org/developers/api

Open University (2013). The open university datasets. Retrieved February 2013, from http://data.open.ac.uk/

O'Reilly, T. (2005, September 30). What is web 2.0. Design patterns and business models for the next generation of software. Retrieved February 2013, from http://oreilly.com/web2/archive/what-is-web-20.html

Osuna-Gómez, C. (1999). *DELFOS: An educational telematic Framework based on levels oriented to cooperative learning situations* (Doctoral dissertation). University of Valladolid, Spain.

Ruiz-Calleja, A., Vega-Gorgojo, G., Alowisheq, A., Asensio-Pérez, J. I., & Tiropanis, T. (2012). Supporting educators to discover and select ICT tools with SEEK-AT-WD. *Proceedings of the European Conference on Technology Enhacend Learning* (pp. 306–319). Saarbrücken, Germany: Springer.

Ruiz-Calleja, A., Vega-Gorgojo, G., Asensio-Pérez, J. I., Bote-Lorenzo, M. L., Gómez-Sánchez, E., & Alario-Hoyos, C. (2012). A linked data approach for the discovery of educational ICT tools in the web of data. *Computers & Education*, *59*(3), 952–962.

Ruiz-Calleja, A., Vega-Gorgojo, G., Gómez-Sánchez, E., Alario-Hoyos, C., Asensio-Pérez, J. I., & Bote-Lorenzo, M. L. (2012). Automatic retrieval of educational ICT tool descriptions from the web of data. *Proceedings of the 12th International Conference on Advanced Learning Technologies* (pp. 105–109). Rome, Italy: IEEE Press.

Staab, S., Studer, R., Schnurr, H. P., & Sure, Y. (2001). Knowledge processes and ontologies. *IEEE Intelligent Systems*, *16*(1), 26–34.

Universidad Complutense de Madrid. (2013). *Sisoft software library*. Retrieved February 2013, from http://ssii.ucm.es/catalogo/fichas/2-5-4.asp

University of Southampton. (2013). *University of southampton open data*. Retrieved February 2013, from http://data.southampton.ac.uk/

Vega-Gorgojo, G., Ruiz-Calleja, A., Asensio-Pérez, J. I., & Jorrín-Abellán, I. M. (2012). U-Seek: Searching educational tools in the web of data. *Proceedings of the European Conference on Technology Enhanced Learning (EC-TEL 2012)* (pp. 525–530). Saarbrücken, Germany: Springer.

Vega-Gorgojo, G., Bote-Lorenzo, M. L., Asensio-Pérez, J. I., Gómez-Sánchez, E., Dimitriadis, Y. A., & Jorrín-Abellán, I. M. (2010). Semantic search of tools for collaborative learning with the Ontoolsearch system. *Computers & Education*, *54*(4), 835–848.

Vega-Gorgojo, G., Bote-Lorenzo, M. L., Gómez-Sánchez, E., Asensio-Pérez, J. I., Dimitriadis, Y. A., & Jorrín-Abellán, I. M. (2008). Ontoolcole: Supporting educators in the semantic search of CSCL tools. *Journal of Universal Computer Science*, *14*(1), 27–58.

Vignollet, L., Ferraris, C., Martel, C., & Burgos, D. (2008). A transversal analysis of different learning design approaches. *Journal of Interactive Media in Education*, 1-11. Retrieved from http://jime.open.ac.uk/article/2008-26/373

World Wide Web Consortium. (2008). *SPARQL query language for RDF*. Retrieved February 2013, from http://www.w3.org/TR/rdf-sparql-query/

Zablith, F., d'Aquin, M., Brown, S., & Green-Hughes, L. (2011). Consuming Linked Data within a large educational organization. *Proceedings of the Consuming Linked Data Workshop (COLD11).* Retrieved from http://ceur-ws.org/Vol-782/ZablithEtAl_COLD2011.pdf