A Semantic Approach to Cross-Disciplinary Research Collaboration

Laurens De Vocht, Davy Van Deursen, Erik Mannens, Rik Van de Walle
Department of Electronics and Information Systems – Multimedia Lab
Ghent University - IBBT
Ghent, Belgium
{laurens.devocht,davy.vandeursen,erik.mannens,rik.vandewalle}@ugent.be

Abstract— Along with the latest developments in ICT, more specifically Social Media and Web 2.0 tools, there is a tendency towards the use of online services in research and education. This is often referred to as Research 2.0 and Technology Enhanced Learning. Web 2.0 tools are especially useful in cases where experts from different disciplines want to collaborate. There is need for an integrated method that embeds these services in research and learning processes, because it is a laborious task for researchers and learners to check and use all varying types of tools and services. In this paper we explain a flexible model that uses state-of-the-art semantic technologies to model both structured and unstructured research data. The research data is extracted from many online resources and Social Media. We implement learning objects as an abstraction of the semantically modeled research data. We propose an environment that improves the scientific research and learning process by allowing researchers to efficiently browse the information and concepts represented as learning objects.

Keywords- Technology Enhanced Learning; Research 2.0; Learning Objects; Semantic Technologies; Social Media; Linked Data

I. INTRODUCTION

It is now an age distinguished by a rapidly changing knowledge society [1]. Web 2.0 tools and Social Media allow sharing of the latest information as soon as it is made available online. Web 2.0 tools such as “Crossref” [2], a citation link network and the official “Document Object Identifier” (DOI) registration service, or “Bibsonomy” [3], a social bookmark and publication sharing system, facilitate reliable disclosure of research data and information. Social Media boost interaction and information exchange at a global scale. For example: Twitter’s ‘Mentions’ and ‘Retweets’, Facebook’s ‘Shares’ and ‘Likes’. The use of such Web 2.0 tools and principles in scientific research and learning is often referred to as “Research 2.0”. It is an application field of “Technology Enhanced Learning” (TEL) which covers the entirety of learning and research with use of new media.

Research data and information is usually made available via so-called digital libraries or institutional repositories, indices containing citations and references to articles such as PubMed, Web of Knowledge or search engines like Google Scholar [4]. Metadata has a crucial role in this; it describes the characteristics of the available data. A researcher can find information stored in institutional repositories using metadata. Learning objects (LOs) enable and facilitate the use of research and educational information. Internationally accepted specifications and standards such as IEEE “Learning Object Metadata” (LOM) or ADL “Sharable Content Object Reference Model” (SCORM) make LOs interoperable and reusable for applications and learning environments. The metadata that describes LOs facilitates searching and rendering them accessible. We define LOs as a group of research or educational (meta)data as entities with links to the people involved. The set of entities belonging to each LO are combined based on a central concept.

A. Problem

Learning object repositories are typically not providing a way to find other learning objects via related online resources. It can be very difficult to navigate, and difficult to integrate several objects into a coherent environment [5]. Cross-disciplinary collaboration depends on creating an environment that makes collaboration easier by meeting motivational needs and enabling user interactions that lead to a shared vision through the construction of a collective conceptual model [6]. Such an environment requires a lot of (meta)data from several information resources. The entirety of all the different (meta)data available online forms a heterogeneous and unstructured whole. In an effort to overcome heterogeneity, communities drive ontologies, annotations and feedback to describe online available resources. DBPedia [7] is a good example, it structures and uniformly describes Wikipedia articles. But it is still a laborious task for users to check all the services one by one as each provides access to a (small) part of the entire available information ‘cloud’. Most of the information provided by these services is not linked except perhaps by non-semantic (‘meaningless’) hyperlinks or references in documents. It is more and more impossible for any university or research institute to guarantee access to all or most of the available research information [8]. On the other hand research data such as those in journal publications has been made available for any researcher via the Web.

B. Proposed Solution

We introduce a model to align, transform and present various resources of research information to students, researchers and anyone involved with knowledge intensive tasks and processes. We investigate several methods to interlink different heterogeneous sources to get better
information disclosure. On top of the integrated data and information we construct the LOs implemented as views on the data (as a selection of properties about a resource) and select the best performing method for each specific purpose. We personalize the view for each researcher by exploiting explicit information such as each researcher’s own content or search queries and implicit information such as reviews, ratings, citations, mentions or other (social) features.

II. CROSS DISCIPLINARY COLLABORATION USE CASE

In research it is common to explore several solutions instead of searching for a specific solution to a specific well defined problem. This is a complex process as there is no knowledge a priori what the solution could look like and all forthcoming results have to be evaluated to which extent they meet a certain research hypothesis. A ready to use answer to the hypothesis can only rarely be found. Especially in cross-disciplinary research, the background of a researcher might not extend to be able to indicate or fully understand a possible solution. Noteworthy is that work is increasingly accomplished through collaboration among interdependent groups of disciplinary specialists [9]. By leveraging data from social media, we make intelligent use of the ‘previous path’ already explored by other (more experienced) researchers in a related discipline. This is shown in Fig. 1 as publish, share and review actions for each experienced researcher. Especially in case when it is not exactly clear where to look for the next viable piece of information or to find the solution for a problem that requires ‘out-of-the-box’ thinking (e.g. when the solution requires background knowledge from a discipline unfamiliar to the researcher).

Users define and select their ’intended’ research objective, from which we analyze and combine various learning objects. In case users have no idea which object/topic to investigate next, they can get an overview of possible objects of interest (similar to points of interest on a street map). The objects can contain entities that represent:

- **Events**: such as scientific conferences, seminars, classes
- **Documents**: such as (journal) articles, reports, tutorials or posts
- **Concepts**: topics, categories, classifications
- **Locations**: both real-world and online (web pages, webinars)
- **People**: as mentions, authors or reviewers

For example, Research 2.0 and by extension TEL is an area that is itself cross-disciplined. In the case of TEL researchers from multiple disciplines such as social scientists, educators, pedagogy experts, business and system engineers or computer scientists work together to create engaging environments using the latest technology. TEL is perfect to develop a use case. For example, suppose we want to do research on novel grading systems to support students and educators. The purpose is that the student gets instant insight and a detailed overview of the feedback for each test and question within a test. How do we get started? Since we are not familiar with existing systems and implementations nor the context of test reviewing and feedback, it is necessary to collect literature and relevant material in a personal library. While doing some of the early reading, we make notes and annotations about the documents we read. In the meantime we also post comments online and share documents with fragments of interest. We connect (by following their updates for example) with authors and other people in the community of TEL. At a certain point, when we feel we have enough background knowledge, we need to work toward solving the grading system problem.

![Cross Disciplinary Collaboration use case](image)

During resource exploration and online activities, we have read lots of information. This information has been collected and integrated in our environment. Around certain entities we select predefined views (or adapt our own) on the collected LOs. We could be interested specifically in Events with a seminar and tutorials, but not in journal articles or conference proceedings. When we have participated in an event or have written a positively received (following a discussion in the research community) review about a journal article, we can state that we have learned and assessed these LOs.

The LOs are extracted from the integrated data according to pre-defined or our own user-defined settings. They are centered around entities such as applications, course subjects, topics, events, people etc. A single learning object can contain links to an online tutorial, details about a seminar, links to fragments of related papers and tutors or people who are known to have contributed to the entities of this specific object. More specifically we define a learning goal or research target. In this case this is centered around the topic "student administration" and "grading systems". Based on this research target, learning objects are filtered and aligned with our registered activities (online comments, sharing, saved publications and previously assessed learning objects).

The environment where learning objects are grouped and mapped according to our research activities, achieved learning objects and defined research target improves cross disciplinary collaboration by making it very easy to discover who is responsible for each object. It allows us to interact immediately and give feedback or comments on the complexity, structure or content of each learning object. If applied correctly, a community centered on certain topics ensures quality of
encountered entities (such as people, locations, events, documents or other objects). The environment allows the extraction of entities and synthesizes learning objects to facilitate interaction between researchers from different disciplines and experience levels.

III. SEMANTIC INTERLINKING MODEL FOR RESEARCH DATA

State-of-the-art semantic technologies are very useful in cases where data integration is required, when the data should be reusable and is made available in a decentralized way or there is no single person responsible for the data in its entirety. Our model aligns and integrates information from the different resources using a process called triplification. This results in a graph based representation of the information and metadata. The data in the set is being described using community approved vocabularies. Relations between different vocabularies and ontologies are resolved and entities originating from different resources referring the same entity are detected and combined.

We combine related entities based on common links they share such as being related to and containing more information about an entity. The users manage the different types of reusable descriptions of views (like 'templates') that are being used. The users generate their own views and objects and can share or compare those with others. All such views of a researcher together make up a personalized environment. This will boost interaction and grouping of similar views and objects onto bigger packages. We also map all objects for users based on their "researcher profile", determined by personal preferences, produced content and interactions with the environment. Thus creating an object-centered view on and representation of the stored data, see Fig. 2.

A. Extraction and Triplification of Data

Our approach provides a solution for three types of data resources: Linked Data (structured data with links), structured data (databases with tables or spreadsheets) and unstructured data (such as peer-reviewed journal articles). We want to end up with all data linked. This means that the metadata is structured and all the data is semantically linked but some of the contents may still be unstructured. All data is accessible from the main endpoint with a query, typically SPARQL.

Linked Data is typically described according to "Resource Description Framework" (RDF), a general method for conceptual description or modeling of information that is implemented in web resources, using a variety of syntax formats. There is already RDF formatted data available for research. For example Bio2RDF is a system to overcome the important issue that datasets needed for the applications might not be available as Linked Data. Bio2RDF is built from a triplifier, RDF storage technology and described in a well-defined ontology. With Bio2RDF, all documents from public bioinformatics databases are now made available to the scientific community in RDF format [10]. Expert search and profiling systems aggregate and analyze certain types of data depending on the kind of expertise hypotheses they use. Traditional approaches tend to retrieve their data from closed or limited data corpuses. Typical Linked Data approaches on the other hand allow querying the whole Web like a huge database, thus surpassing the limits of closed data sets, and closed online communities. This opens new possibilities for traditional expert search and profiling systems which usually only rely on data from their local and limited databases or on unstructured data gathered from the Web. For being a simple collection of linked triples [11], Linked Data also wants to deliver multipurpose data that can be used to find experts in many domains and with many different expertise hypotheses [12].

Much research information is not available as RDF but typically in relational databases as tables or spreadsheets. To make data available in RDF we use two types of processes: Pre-defined (static) annotations using the API of the resource provider to load the information from the data repository (1); A dynamic mapping between the ontology and the data repository, for example with a tool like D2RQ [13] in case of relational databases (2).

In our approach we do not focus on the structuring or entity recognition of unstructured data. Instead we make use of the current state-of-the-art approaches and we assume that they are advanced enough to extract usable entities such as people, locations, topics and other types from the source data. Such a tool is OpenCalais [14] or the NERD framework that aggregates multiple tools and leverages the results as linked data [15]. The results from the third-party tool, the extracted information, are used to describe the unstructured information resource.

B. Describing Data using Relevant community approved Vocabularies

Data in RDF is typically described with the use of a domain specific vocabulary or ontology. There exist some frequently
used vocabularies. They have the advantage that they are the same over many resources making it convenient to make the data integration: SKOS (concept modeling), FOAF (social media), DCTERMS (context and metadata), MOAT (tags), SIOC (people and relations) and SWRC (research communities ontology). These vocabularies are only sufficient to describe metadata itself and more common data. More complex or specialized data requires a more advanced and maybe dynamic ontology or sub-ontology. Maintaining ontology or extracting sub-ontology requires special focus and for research we will track evolutions in this area but we have no plans to carry out research in this area ourselves. Promising advances to describe research data are for example a comprehensive ontology for call for papers. That would be relevant and useful in Research 2.0 context for two main target groups: authors involved in collaborative writing of academic papers, and conference organizers or journal editors [16].

Designing involved in collaborative writing of academic papers, and in Research 2.0 context for two main target groups: authors involved in collaborative writing of academic papers, and conference organizers or journal editors [16]. Designing ontology for call for papers. That would be relevant and useful in Research 2.0 context for two main target groups: authors involved in collaborative writing of academic papers, and conference organizers or journal editors [16]. Designing ontology for call for papers. That would be relevant and useful in Research 2.0 context for two main target groups: authors involved in collaborative writing of academic papers, and conference organizers or journal editors [16]. Designing ontology for call for papers. That would be relevant and useful in Research 2.0 context for two main target groups: authors involved in collaborative writing of academic papers, and conference organizers or journal editors [16]. Designing ontology for call for papers. That would be relevant and useful in Research 2.0 context for two main target groups: authors involved in collaborative writing of academic papers, and conference organizers or journal editors [16]. Designing ontology for call for papers. That would be relevant and useful in Research 2.0 context for two main target groups: authors involved in collaborative writing of academic papers, and conference organizers or journal editors [16]. Designing ontology for call for papers. That would be relevant and useful in Research 2.0 context for two main target groups: authors involved in collaborative writing of academic papers, and conference organizers or journal editors [16]. Designing ontology for call for papers. That would be relevant and useful in Research 2.0 context for two main target groups: authors involved in collaborative writing of academic papers, and conference organizers or journal editors [16]. Designing ontology for call for papers. That would be relevant and useful in Research 2.0 context for two main target groups: authors involved in collaborative writing of academic papers, and conference organizers or journal editors [16]. Designing ontology for call for papers. That would be relevant and useful in Research 2.0 context for two main target groups: authors involved in collaborative writing of academic papers, and conference organizers or journal editors [16]. Designing ontology for call for papers. That would be relevant and useful in Research 2.0 context for two main target groups: authors involved in collaborative writing of academic papers, and conference organizers or journal editors [16]. Designing ontology for call for papers. That would be relevant and useful in Research 2.0 context for two main target groups: authors involved in collaborative writing of academic papers, and conference organizers or journal editors [16]. Designing ontology for call for papers. That would be relevant and useful in Research 2.0 context for two main target groups: authors involved in collaborative writing of academic papers, and conference organizers or journal editors [16]. Designing ontology for call for papers. That would be relevant and useful in Research 2.0 context for two main target groups: authors involved in collaborative writing of academic papers, and conference organizers or journal editors [16]. Designing ontology for call for papers. That would be relevant and useful in Research 2.0 context for two main target groups: authors involved in collaborative writing of academic papers, and conference organizers or journal editors [16]. Designing ontology for call for papers. That would be relevant and useful in Research 2.0 context for two main target groups: authors involved in collaborative writing of academic papers, and conference organizers or journal editors [16]. Designing ontology for call for papers. That would be relevant and useful in Research 2.0 context for two main target groups: authors involved in collaborative writing of academic papers, and conference organizers or journal editors [16]. Designing ontology for call for papers. That would be relevant and useful in Research 2.0 context for two main target groups: authors involved in collaborative writing of academic papers, and conference organizers or journal editors [16]. Designing ontology for call for papers. That would be relevant and useful in Research 2.0 context for two main target groups: authors involved in collaborative writing of academic papers, and conference organizers or journal editors [16]. Designing ontology for call for papers. That would be relevant and useful in Research 2.0 context for two main target groups: authors involved in collaborative writing of academic papers, and conference organizers or journal editors [16]. Designing ontology for call for papers. That would be relevant and useful in Research 2.0 context for two main target groups: authors involved in collaborative writing of academic papers, and conference organizers or journal editors [16]. Designing ontology for call for papers. That would be relevant and useful in Research 2.0 context for two main target groups: authors involved in collaborative writing of academic papers, and conference organizers or journal editors [16]. Designing ontology for call for papers. That would be relevant and useful in Research 2.0 context for two main target groups: authors involved in collaborative writing of academic papers, and conference organizers or journal editors [16]. Designing ontology for call for papers. That would be relevant and useful in Research 2.0 context for two main target groups: authors involved in collaborative writing of academic papers, and conference organizers or journal editors [16]. Designing ontology for call for papers. That would be relevant and useful in Research 2.0 context for two main target groups: authors involved in collaborative writing of academic papers, and conference organizers or journal editors [16]. Designing ontology for call for papers. That would be relevant and useful in Research 2.0 context for two main target groups: authors involved in collaborative writing of academic papers, and conference organizers or journal editors [16]. Designing ontology for call for papers. That would be relevant and useful in Research 2.0 context for two main target groups: authors involved in collaborative writing of academic papers, and conference organizers or journal editors [16]. Designing ontology for call for papers. That would be relevant and useful in Research 2.0 context for two main target groups: authors involved in collaborative writing of academic papers, and conference organizers or journal editors [16]. Designing ontology for call for papers. That would be relevant and useful in Research 2.0 context for two main target groups: authors involved in collaborative writing of academic papers, and conference organizers or journal editors [16]. Designing ontology for call for papers. That would be relevant and useful in Research 2.0 context for two main target groups: authors involved in collaborative writing of academic papers, and conference organizers or journal editors [16]. Designing ontology for call for papers. That would be relevant and useful in Research 2.0 context for two main target groups: authors involved in collaborative writing of academic papers, and conference organizers or journal editors [16]. Designing ontology for call for papers. That would be relevant and useful in Research 2.0 context for two main target groups: authors involved in collaborative writing of academic papers, and conference organizers or journal editors [16]. Designing ontology for call for papers. That would be relevant and useful in Research 2.0 context for two main target groups: authors involved in collaborative writing of academic papers, and conference organizers or journal editors [16]. Designing ontology for call for papers. That would be relevant and useful in Research 2.0 context for two main target groups: authors involved in collaborative writing of academic papers, and conference organizers or journal editors [16]. Designing ontology for call for papers. That would be relevant and useful in Research 2.0 context for two main target groups: authors involved in collaborative writing of academic papers, and conference organizers or journal editors [16]. Designing ontology for call for papers. That would be relevant and useful in Research 2.0 context for two main target groups: authors involved in collaborative writing of academic papers, and conference organizers or journal editors [16]. Designing ontology for call for papers. That would be relevant and useful in Research 2.0 context for two main target groups: authors involved in collaborative writing of academic papers, and conference organizers or journal editors [16].

C. Interlinking of Entities and Resources

We define relations between items and their descriptions. The most important relation is similarity. In its most basic version entities which are linked to many of the same entities are considered similar. This is a simple and strict approach. More advanced version might take into account weights for certain type of links. These can be determined and tweaked based on input from user or by learning from feedback. Links found to be important gain a higher weight over time. Related entities are items which are linked to many similar entities. We also derive dependencies if the ontologies specify RDFS, OWL or SKOS based class hierarchies (e.g. generalization and specification) as dimension to be used for dedicated filters. We discover and keep track of known entities and allow the user to manage the links between vocabularies. This resolves the issue of properties that refer to the same concept but use a term from a different vocabulary.

To resolve synonyms we rely on the good definition of each vocabulary and a dictionary or thesaurus. We make use of semantic (e.g., hyponymy or meronymy) and/or lexical (e.g., synonyms) relationships within the available network (graph) of concepts to determine semantic proximity between the concepts and between literals. It might occur that for some concepts little or even no relationships at all are defined in the thesaurus or vocabulary. These are typically literal property values where the property itself has little meaning. In this case, finding a matching concept among the other datasets and entities can be very difficult as there is no additional context present that can be used for disambiguation. For this reason, we will first ignore such cases and assume that these are rare cases. In case we notice such cases frequently in a certain context we could use an approach like Debevere et al. where statistics are collected, in order to get a set of keywords that most frequently co-occur with another keyword [17]. These keywords will then be used as additional context information in order to find a matching concept among the other datasets and entities.

D. Quality Dissemination of Individual Data Entities

Each of the integrated entities is evaluated based on primary and secondary characteristics. Primary characteristics such as social features (tags, ratings, authors…), creation date and content length can be extracted immediately from the metadata. Secondary characteristics such as reliability, factuality and topicality are analyzed using algorithms based on the actual content and additional information generated by the community around the main content.

If an entity is linked to several other entities with content, then the weights for each of the factors in the basic version can be calculated using a weighted average, configurable by the user for each type of node. Fig. 3 shows an example of such an entity can be for example a journal that consists of many journal articles with each their own quality characteristics. Of course the entity itself has direct influence on the score as it may also be subject to social discussion and its metadata is being taken into account.

Content differs highly in quality and could even be useless (like spam). Social features are an addition to the content itself, there is a wide array of metadata available, such as explicit quality ratings from members of the community. Other important social features can be tags, likes, shares, feedback, ratings, reviews, citations, referrals, certifications… An important difference between user-generated content and traditional content that is particularly significant for knowledge-based media (such as for example question/answering portals) is the variance in the quality of the content. A classification framework for quality estimation in social media based on such features was implemented using a comprehensive graph-based model of contributor relationships and combined with content- and usage-based features. It has been successfully applied to identify high quality items in a web-scale community [18].

Figure 3. Quality of each entity depending on their linked entities
We primarily focus on difficulty level, topicality and the reliability of the content as it is very important in the presentation of the final learning objects to the user. Every object should contain entities of the same difficulty level. When the users get more experience they gain access to more advanced objects that consist of more difficult entities. For every resource and type of entity the user can determine a difficulty level (for example tutorials, journal articles or online posts). The topicality of an object depends on the creation date, modification date and social activity (how discussed a resource belonging to an entity is). For example a news article could spark commotion and lead to weeks of opinions, forum activity etc. Users can select to only include or prefer ‘hot topics’ in their overview. Provenance (keeping track of the sources and history) is assured by presenting to the user all resources responsible for each entity and each piece of information. Furthermore references to different resources and repositories containing the original data of an entity are being kept alongside each entity. The level and amount of trustworthy resources for each entity determines its reliability. The value of a resource is also influenced by the number of valuable citations because important resources, for example publications, in a scientific community are characterized by a high degree centrality in its citation network [19]. The feedback on entities for each specific resource by the community will affect the trustworthiness of the resource itself. The user can select at any time resources and people he trusts as a filter on the presented learning objects. For example a filter can be applied when a researcher is only interested in learning objects from entities extracted of publications from its own university.

E. Viewing Data and Converting Entities to Learning Objects

The Semantic Web technology stack and LOD principles are powering a new approach that is being developed for the discovery of LOs. Preliminary evaluation of related systems has shown that the system is effective in terms of LO disclosure [20]. We implement a semantic engine to extract learning objects with SPARQL queries. The engine looks up and includes all related entities centered on a topic or a concept to construct each single object. The engine interprets user actions and updates the view on the new desired objects. The engine also converts entities stored in the basic layer to learning objects. Objects are represented in Fig. 4(a) as blocks in the research space relevant to the dimensions with graph and the result selection of objects.

Every researcher has a previous set of assessed learning objects. For every object or the entire cluster of objects it is possible to map other objects, selected according to the learning or research target in dimensions broadening (determined by similarity) or deepening (determined by required dependencies). Researchers can slice this space with filters on topicality and difficulty level for example. Each learning goal or research target is translated as a set of learning objects and the engine selects all the objects between this set and the already assessed objects, as shown in Fig. 4(b). If the resulting set is too big or to varied then the researcher, then possible objects of interest on the path to the target object gain preference. A research target or learning goal is set by exploring all available objects while applying filters in suggested dimensions such as topicality, difficulty level, topic relevance etc.

IV. Socially Enhanced Resource Alignment to Support and Trigger Collaboration

Objects, more specifically the research target (for example: understanding a disease in a clinical trial) and the infrastructure (email, documents, learning objects) support cross-disciplinary collaboration: they provide the motives and drive for collaboration to emerge, they allow participants to work across different types of boundaries, and they constitute the fundamental infrastructure of the activity [21]. We derive up-to-date researcher profiles out of the data explicitly provided by researchers such as content of their microblogs, articles, notes or other documents combined with implicit or initially unobservable information about the researcher such as search actions or activity logs. A user profile with implicit information such as feedback can dramatically improve the quality of personalized views on sets of data [22]. When users interact with a result set, they leave a “semantic fingerprint” they are interested in the content of this item to a certain degree. A single profile is useful because it can improve the view of entities out of all available resources in the integrated repository. This method of profiling also helps in the understanding of scientific relevance and importance of entities or learning objects to other researcher’s specific needs. The extent of social network data is massive and individual researchers are only likely to be interested in specific parts of the overall knowledge on the basis of their area of specialization.

We build result sets (other researchers, communities, conference suggestions, paper recommendations) around common entities where researchers link to and may not be aware of. It is unlikely that researchers have always sufficient knowledge of other disciplines to be able to explicitly identify or detect the right learning objects they want to assess. An interactive approach helps researchers to find the right people behind each resource. For example, Fig. 5 shows researcher Anna and she has set a learning goal that requires assessing objects already assessed by researcher Bob. For Anna, Bob will be a reference during the assessment of the objects and is
encouraged to interact with Bob. With the help of Bob, Anna can assess required objects much more efficiently. Bob gains more experience by helping Anna. This dynamic group forming around a single research goal or learning objective allows faster assessment of objects with some overlap where necessary but opens up opportunities for more complex research problem solving.

Figure 5. Research Object Alignment: the objects researchers want to assess can overlap with objects assessed by other researchers.

V. EVALUATION

We are applying a design based research methodology that involves several iterations containing extensive literature study, analysis of user questionnaires and qualitative system tests using consecutive rapid prototypes of implemented use cases. We choose this approach because it was developed as a way to carry out formative research to test and refine the processes based on principles derived from prior research [23]. Formative research is conducted to help describe the target audience, understand the factors which influence their behavior, and determine the best ways to reach them. The development of our proof-of-concept (POC) consists of several iterations. All iterations refine existing processes and introduce new processes to meet some of the key concepts and to test the environment. The first iteration focused on the integration of data from a single social media resource and extracting entities out of it. These entities were linked to LOD. This version provided an interface for researchers to discover people or conferences related to them based on the number of entities they shared. The second iteration focuses on the interlinking of entities from multiple heterogeneous resources such as social media, notes and digital libraries. The third iteration focuses on the mapping of entities to learning objects and mapping them according to the explained dimensions. Future iterations will implement features as depicted in the future work section of this paper.

As a POC we have successfully implemented an application for the researcher profiling use case: “Researcher Affinity Browser” [24]. The use case is about system suggestions of interesting people or of relevant scientific events to its users. This application exposes affinities, otherwise hidden proximities or to likings for specific elements or in this case learning or research objects. It shows researchers related connections based on common affinities such as hash tags, mentions or conferences. The attendance of researchers at conference is a good indicator. In the past few years many researchers have started to blog and tweet about conferences. We focused on linking the content of a single information source to the LOD cloud: microblogs on Twitter (a.k.a. tweets). We chose Twitter because of its accessible API and its users share relevant content. We focused on creating a reliable personalized view on existing social data using the process of extraction, annotation and interlinking. Conferences were identified and linked to a secondary source and this turned out to be an important indicator of the reliability. We found that more different resources belonging to entities would imply more reliability of these entities.

The second iteration of our POC addresses the integration of information from various Web 2.0 tools and Social Media with different interfaces and APIs. In the first iteration we focused on the triplification and annotation steps of the semantic model presented. Section III explains the steps in detail. We are testing the concept of our environment that makes research data available through the interlinking of multiple resources. We implement the interlinking as a dynamic alignment to improve the reliability of the content presented. It was noted by test users during the first iteration as an important criterion for improvement. We create an overview of all the entities extracted from the input data. As input data we continue to use Twitter as information stream for web resources and user posts. We allow the user to connect with its Mendeley [25] library of collected literature and cross-reference authors of publications with people mentioned on or using Twitter. Mendeley is a tool to manage, discover and share research. We continue to track information about conferences using their hashtags. Now we also use the user provided tags on their notes with Evernote to link their notes with other available resources. Evernote is cross-platform tool to help users remember everything along their way. This includes web snippets, notes, bookmarks and OCR’ed pictures of handwritten notes. We have completely redesigned the architecture of the framework behind the Researcher Affinity Browser to allow updates of the entity space for each researcher within a few minutes rather than in a few hours (like in the previous iteration). The entities supported in the second iteration are Conferences, Authors, Documents (Articles, Web Pages) and Locations. As soon as the prototype web service and back-end is finished, we plan to do evaluation of the environment with simulations with pre-defined scenarios to test and tweak the desired environment behavior. Measure the expected outcome vs. the actual outcome to optimize the content shown in different views.

VI. RELATED WORK

The relevance to students and researchers of personal learning and research environments is not new. From a user-point of view, tests with the AWESOME framework, an awareness Dashboard for Research Communities showed that the integrated environment performed better in covering necessary activities [26]. Their approach was mostly wireframe based and assumed a ‘perfect’ working backend. Important first steps towards the implementation of such an environment contain a mash-up infrastructure allowing for continuous
innovation. This is done by recombining and repurposing existing technology: Ullman et al. proposed building blocks and their concrete implementation to leverage Research 2.0 practice and technologies, including a publication feed format for exchanging publication data [27]. This infrastructure is strongly based on the use of Linked Data. In several application fields and research areas, Linked Data is more and more being used as a fundamental component for the implementation of knowledge intensive applications and environments. Another example is the “3A recommender system” [28]. It targets computer-supported collaborative learning environments. The system models user interactions in a heterogeneous graph: “interactive model” and applies a personalized, contextual, and multi-relational ranking algorithm to rank Actors, Activity spaces, and Assets. An important difference between the 3A approach and ours is the way the data is being modeled, mapped and presented to the user. The 3A system does not model the entities and their relations semantically and thus the data model is mostly interaction based (for example tagging). They rely on patterns between user ratings and activities, but do not use data about items. They do not, for instance, know that “Mathematics” is a discipline and “Algebra” a topic within “Mathematics”.

While the number of linked datasets has grown significantly over the last years, there is still a marked lack of mature applications that enables researchers to explore and query these datasets. Linked data browsers such as SWSE [29] or Sig.ma [30] are currently too generic for most researchers (although they can be very helpful for developers). Organic.Edunet, a federation of learning repositories in the domain of organic agriculture, has been integrated using linked data and a redesign of the export mechanisms that uses an RDF store and several ontologies to browse and search resources [31]. The linked data approach represents an opportunity to evolve existing learning object repository technology towards the open exposure of metadata in a form that enables novel approaches to search and navigation that are not restricted to centralized metadata stores and that enable navigating across repositories without a need for integrating them beyond the provision of links to other linked data sources.

CS AKTive Space (CAS) is a Semantic Web application that seeks to provide the experience of an integrated information overview that allows a user to determine quickly who is doing what where in computer science research in the UK [32]. CAS has demonstrated that it is possible to explore an ontologically-mediated information space gathered from a wide set of heterogeneous sources. Other systems such as EURECOM EventMedia and the ‘Confomaton’ framework encourage collaboration when researchers are ‘on the road’ and looking for conferences to participate [33]. The framework aggregates in real-time social media shared by conference attendees and aligns it with event descriptions.

In above solutions researchers browse typically an environment with a to-the-point view of actual related research that supports solving partly the questions of the research hypothesis or the exploration of interesting information. Locations are typically presented on a geographical map. Unlike these solutions that rank and filters results according to selected features and personal settings. We generate a view where potentially relevant resources are showed. To verify relevancy the researcher is encouraged to interact with the environment and the people responsible for providing the information presented such as the authors of an article or the writers of a post.

VII. CONCLUSIONS AND FUTURE WORK

We presented the concept of a novel research environment that uses the previous path and profile of each researcher to connect researchers with new resources. We encourage interaction with other researchers. We derive data from social media, existing data repositories and LOD. We aggregate this data, extract entities out of it to generate learning objects as views on the aggregated data. We have shown that the enrichment of social media with linked data is a solid foundation for the next generation research environments: “Research 2.0”. There is great value in exploring administrative and research data linkage options, particularly through reduction of costs through decreased administration expenses. It is now very feasible and desirable to implement a blended research and learning object repository because it is still a laborious task for researchers to construct an optimal search query for each of those services one by one. We expect by advanced refinement of the aligning of the different information sources, we will make a leap forward and be able to present to-the-point resources efficiently and effectively.

The physical presence of the university building with books and libraries etc. is an aspect of the university infrastructure which is already becoming invisible and more and more obsolete thanks to latest developments in ICT and current intelligent, interactive research and learning environments. These environments and interactions need time to develop further and require a long phase of idea generation preceding any focused research effort. Collaboration is by itself a complex system of people, scientific theory, and tools that must be intentionally managed. Our contribution is that we are demonstrating and implementing such an environment with current state-of-the art technologies combined with the contributions from a supportive community with enough incentive to remain interactive and contribute.

Future research will focus on three areas: how we can determine and verify content quality, the efficiency of the semantic engine and the evaluation of the environment. The environment needs enough incentives for the users to keep the system up-to-date and review all the content. Content reviews can be assured by boosting interactivity. It is thus important that users can manage their achievements and assessments with a highly trusted scoring, rating and qualification system. Users should be able to demonstrate their experience in certain domains and more experienced researchers should have more influence on the evaluation of other objects, especially their factuality and reliability. The semantic engine will be worked out in more detail to obtain a more sophisticated selection and presentation of new learning objects based on previous assessed objects and traversed paths. We want to implement a version of the engine that can actually realize the concept of mapping learning objects to personalized roadmap according to achieve goal or research objective.
ACKNOWLEDGMENTS

The research activities that have been described in this paper were funded by Ghent University, the Social Learning Department at Graz University of Technology, the Interdisciplinary Institute for Broadband Technology (IBBT), the Institute for the Promotion of Innovation by Science and Technology in Flanders (iWT), the Fund for Scientific Research-Flanders (FWO-Flanders), and the European Union.

REFERENCES