Proceedings of the 2nd International Workshop on Research 2.0

At the 5th European Conference on Technology-Enhanced Learning: Sustaining TEL

Barcelona, Spain, September 28, 2010

Erik Duval
Fridolin Wild
Peter Scott
Thomas Ullmann
Stefanie Lindstaedt
(Eds.)

Please refer to these proceedings as

Erik Duval, Thomas Daniel Ullmann, Fridolin Wild, Stefanie Lindstaedt & Peter Scott (eds.): Proceedings of the 2nd International Workshop on Research 2.0. At the 5th European Conference on Technology Enhanced Learning: Sustaining TEL. Barcelona, Spain, September 28, 2010, CEUR-WS.org/Vol-675, ISSN 1613-0073.

© 2010 for the individual papers by the papers' authors. Copying permitted for private and academic purposes. Re-publication of material from this volume requires permission by the copyright owners.

Address of first editor:
Erik Duval
Katholieke Universiteit Leuven
Department of Computer Science
Celestijnenlaan 200a
B-3000 Leuven, Belgium

erik.duval@cs.kuleuven.be

The front-cover was created by Harriett Cornish (The Open University, KMi).

Editorial: Research 2.0 for TEL - Four Challenges

In recent years, Web 2.0 has become manifest in new types of applications causing fundamentally new experiences of large-scale social interaction. It has affected the way people communicate, share, collaborate, and - ultimately - participate on the Web. The technologies associated with "the Web 2.0" have a focus on broadened participation by lowering the technical barriers for users. Over the years, the ability to publish content on the Web with little technical knowledge has created not only a new level of public accessible data, but also created the dynamic world of the social Web. The openness of the Web also allowed building new services based on old ones, fostering the development of a mash-up culture.

The philosophy underpinning reflects back on the practice of researchers, not only in tech savvy areas of research. However, what does this really mean? Is it about the adoption of existing tools and services? Is it about the (re-)development of applications based on success criteria of Web 2.0 applications? Is it about the distillation of good practice and their diffusion amongst researchers, either bottom-up or top-down? What type of methodology is appropriate to investigate Research 2.0 phenomena?

As concluded during the workshop, at least four challenges are vital for future research. The first area is concerned with availability of data. Access to sanitized data and conventions on how to describe publication-related meta-data provided from divergent sources are enablers for researchers to develop new views on their publications and their research area. Additional, social media data gain more and more attention. Reaching a widespread agreement about this for the field of technology-enhanced learning would be already a major step, but it is also important to focus on the next steps: what are success-critical added values driving uptake in the research community as a whole?

The second area of challenges is seen in Research 2.0 practices. As technology-enhanced learning is a multi-disciplinary field, practices developed in one area could be valuable for others. To extract the essence of successful multi-disciplinary Research 2.0 practice though, multi-dimensional and longitudinal empirical work is needed. It is also an open question, if we should support practice by fostering the usage of existing tools or the development of new tools, which follow Research 2.0 principles. What makes a practice sustainable? What are the driving factors?

The third challenge deals with impact. What are criteria of impact for research results (and other research artefacts) published on the Web? How can this be related to the publishing world appearing in print? Is a link equal to a citation or a download equal to a subscription? Can we develop a Research 2.0 specific position on impact measurement? This includes questions of authority, quality and re-evaluation of quality, and trust.

The tension between openness and privacy spans the fourth challenge. The functionality of mash-ups often relies on the use of third-party services. What happens with the data, if this source is no longer available? What about hidden exchange of data among backend services?

This year's Research 2.0 Workshop at the EC-TEL 2010 Conference in Barcelona had an emphasis (a) on tools, applications, and infrastructure components supporting researchers and (b) on insights into how practices of researchers change. It combined quantitative and qualitative approaches shedding light on different facets of Research 2.0.

Kraker, Fessl, Hoefler, and Lindstadt present in their paper "Feeding TEL: Building an Ecosystem Around BuRST to Convey Publication Metadata" a system fostering the exchange publication meta-data. They propose to use a semantically enriched RSS format, which allows institutions to exchange publication meta-data and to make this meta-data accessible for research. The paper also presents complementing services and widgets, and outlines the benefit of the approach for institutions.

Parra and Duval describe in their paper "Filling the Gaps to Know More! About a Researcher" a mobile application called More! that serves the discovery of researcher profile information about a speaker at a conference. Their approach takes into account the various identities of researchers on the Web to present relevant information for researchers with a unified interface. The mobile application presents information about the researcher, the current work, and social handles.

Joubert and Sutherland look at the practice of collaboratively writing a deliverable about vision and strategy for the STELLAR network of excellence. In their paper "Research 2.0: Drawing on the Wisdom of the Crowds to Develop a Research Vision" they outline their experiences with a wiki software in the collaborative writing process. They discuss risks and outline strategies to overcome them. They especially highlight the importance of the engagement of the contributors, the discussion features of wikis, and the clarification of the overall goal of the collaboration.

Vandeputte and Duval report on a multi-touch table, called the ScienceTable, in their paper "Research at the Table". They focus on the support of researchers in finding scientific papers. Researchers can explore the co-author space of publications. Two tasks are supported. Researchers can either use the multi-touch table to explore the publication world top-down or they can use the table with a bottom-up approach, exploring the neighbourhood of authors.

The interactive visualization Muse is described in the publication "Muse: Visualizing the Origins and Connections of Institutions based on Co-authorship of Publications" of Till Nagel and Erik Duval. The focus on this visualization is on exploring the collaborations between institutions. Therefore, they geo-locate the affiliation of authors. This gains insights into the collaboration network of institutions, regions, and countries. Same as the ScienceTable, Muse runs on a multi-touch table.

The paper "Tools to Find Connections between Researchers - Findings from Preliminary Work with a Prototype as Part of a University Research Environment" of Hensmann, Despotakis, Brandic, and Dimitrova presents tools of the JISC Brain (Building Research and Innovation Networks) project with emphasis on identifying connections between researchers, as well as researchers and business and other wider partners. The tools described in the paper provide facilities for researchers to search for other researchers by keywords, which are related to own work, and to find links between researchers. Central to their work is a Research 2.0 approach supporting researchers in several stages of their research carrier.

Wild and Ullmann explore the collaboration networks of deliverables of the STELLAR network of excellence in their paper "The Afterlife of "Living Deliverables": Angles or Zombies". It focuses on collaboratively authored online project reports, that use a wiki software to support the writing, but also serve to enable knowledge exchange after the submission deadline. While wikis tend not to emphasize authorship of individuals, versioning history data of the wikis allow drawing conclusions on the nature of the collaboration and particularly on which authors collaborated on text passages and topics. In their empirical investigation, they describe the collaboration on a deliverable before and after the deadline. They state that most of the deliverables are used also after the deadline, while others only exist for the purpose of writing up and delivering.

The microblogging platform Twitter is subject of the paper "@twitter Try out #Grabeeter to Export, Archive and Search your Tweets" by the authors Muehlburger, Ebner, and Taraghi. Starting with the problem that Twitter streams usually are not available anymore after an event, they propose a solution in the form of an application called Grabeeter, which stores the tweets locally, allowing analysing the tweets also after the event. They discuss the architecture of the application client and server aspects and they focus specifically on how to use the system for conducting research.

The paper "Connecting Early Career Researchers: Investigating the Needs of Ph.D. Candidates in TELWorking with Web 2.0" from Heinze, Joubert, and Gillet, reports on a case study about the needs of young TEL researchers. The authors asked 21 doctoral candidates and three senior researchers about how they would wish to receive support for their doctoral work: regarding personal support, awareness support, and tools for collaboration. The three major findings were that first, it is unlikely that even a larger community of practice can survive on its own; second, a community of practice is highly dependent on individuals dedicated to it; and third, tools or services should mainly support collaboration and communication.

We want to use this opportunity to thank the authors for their contributions. The work in organising the workshop and producing these proceedings has been financially supported by the European Union under the ICT programme of the 7th Framework Programme in the project STELLAR.

October 2010

Erik Duval, Thomas Daniel Ullmann
Fridolin Wild, Stefanie Lindstaedt
Peter Scott

Organizing Committee

Erik Duval, Katholieke Universiteit Leuven, Belgium
Stefanie Lindstaedt, Know-Center Graz, Austria
Peter Scott, The Open University, United Kingdom
Thomas Daniel Ullmann, The Open University, United Kingdom
Fridolin Wild, The Open University, United Kingdom

Program Committee

Xavier Ochoa, Escuela Superior Politecnica del Litoral (ESPOL) at Guayaquil, Ecuador
Wolfgang Reinhardt, University of Paderborn, Germany
Nina Heinze, Knowledge Media Research Center Tuebingen, Germany
Peter Kraker, Know-Center Graz, Austria
Frederik G. Pferdt, University of Paderborn, Germany
Johannes Metscher, University of Augsburg, Germany
Andreas S. Rath, Know-Center Graz, Austria

Contents

Feeding TEL: Building an Ecosystem Around BuRST to Convey Publication Metadata

Peter Kraker[1], Angela Fessl[1], Patrick Hoefler[1] and Stefanie Lindstaedt[1]

[1] Know-Center Graz, Inffeldgasse 21a,
8010 Graz, Austria
{pkraker, afessl, phoefler, slind}@know-center.at

Abstract. In this paper we present an ecosystem for the lightweight exchange of publication metadata based on the principles of Web 2.0. At the heart of this ecosystem, semantically enriched RSS feeds are used for dissemination. These feeds are complemented by services for creation and aggregation, as well as widgets for retrieval and visualization of publication metadata. In two scenarios, we show how these publication feeds can benefit institutions, researchers, and the TEL community. We then present the formats, services, and widgets developed for the bootstrapping of the ecosystem. We conclude with an outline of the integration of publication feeds with the STELLAR Network of Excellence[1] and an outlook on future developments.

Keywords: science 2.0, web 2.0, mashups, services, widgets, feeds

1 Introduction

Recently, developments under the paradigm of Science 2.0 have received a lot of attention [1]. Researchers are embracing the capabilities of Web 2.0 tools and technologies, such as blogs, wikis, and social networking sites, to support their research. Using Web 2.0 for scientific work has numerous potential advantages: it possibly leads to shorter feedback cycles, enhances the communication between researchers, and yields a higher penetration of ideas. One of the prerequisites for the introduction of a modern Science 2.0 in the field of Technology Enhanced Learning is the wide-spread access to resources, data, and publications for the whole community [2].

In this paper we present an ecosystem for the exchange of publication data based on existing Web 2.0 infrastructure. At the heart of this ecosystem, semantically enriched feeds based on the popular RSS format [3] are used as a means for lightweight exchange of information on the web. They can easily be combined, aggregated, visualized, and republished. Hence, publication feeds have the advantage

[1] STELLAR [4] is an EU-funded Network of Excellence, which aims at unifying the diverse community in the field of Technology Enhanced Learning in Europe.

to provide important scientific data in a format widely used by existing Web 2.0 infrastructure.

To facilitate the opening of institutional archives, easy-to-use tools are needed. Web services are especially apt for this, since they are the cornerstone of Web 2.0, allowing for loosely coupled systems and simple syndication [5]. Whereas the services aid the producer in generating a publication feed, widgets let the recipient consume and manipulate these feeds. Users can collectively contribute to the data base by adding their own feeds; they can help identify good publications by rating them, and interact with each other by leaving comments. A visualization widget provides them with filtering and searching facilities for the aggregated data.

This paper consists of three sections. At first, we introduce two scenarios for the usage of publication feeds in research from a personal and an organizational perspective. Then, we present the pillars of the ecosystem, namely the adapted BuRST format, a suite of web services for feed producers, and several widgets for feed consumers. Finally, we conclude with an overview of the integration of the ecosystem into the STELLAR Network of Excellence and an outlook on future developments.

2 Scenario

In the following section we present two scenarios which illustrate the benefits of the presented ecosystem. These scenarios emphasize on lightweight dissemination, visualization, and navigation of semantically-enriched scientific publication feeds in the style of Web 2.0.

2.1 Scenario 1: Semi-automated dissemination of publication feeds

Sandra is a supervisor at a TEL research institution dedicated to professional learning. She is responsible for collecting the publications of her group. Therefore, her assistants keep a BibTeX file of their publication metadata, which is periodically uploaded to a common server. Sandra is interested in a wider dissemination of this data, but unfortunately she cannot get her assistants to enter the publication data over and over again into other repositories. Hence, she is looking for a way to automate dissemination. Since publication data is already available in several BibTex files, she uses a dedicated BibTeX converter to convert these files into publication feeds. The resulting individual feeds are then merged into a single feed with the help of the Publication Feed Merger. Due to the fact that there are also publications not related to TEL in the feed, a Publication Feed Filter is applied. Sandra now publishes this feed so that all interested parties that support the BuRST format can subscribe to it.

2.2 Scenario 2: Explorative research on publication feeds

Kurt is an early-career researcher interested in professional learning. He wants to find out about the most influential publications, recently trending topics, and interesting conferences in the field. Therefore, he joins a special interest group dedicated to professional learning on a social networking platform. Sandra and other users have already added their institutions' publication feeds to this group. The individual publications are presented as blog posts, which can be rated and commented on. Kurt now has an overview of the top rated publications and the discussions revolving around them.

Kurt then opens the "Publication Visualization" widget from within the special interest group. He is presented with a faceted browsing view containing all publication metadata from the feeds. A tag cloud aggregated from the keywords is additionally shown to Kurt. He then restricts the data to certain years to see the changes in the tag cloud. This allows him to reflect on the trending topics.

Next, Kurt restricts the publication type to conference proceedings. Now, all proceedings titles are presented to him, alongside the corresponding articles. From the keyword tag cloud, he chooses a topic that he finds interesting. This supplies Kurt with a list of conferences that are important for that specific topic.

3 Publication Feed Ecosystem

In this section, we present the three initial pillars of the publication feed ecosystem: the adapted BuRST format, a suite of web services for feed producers, and several widgets for feed consumers.

3.1 Publication Feeds

Publication feeds are RSS 1.0 feeds, enhanced with elements from the SWRC[2] and DC[3] ontologies. These feeds are an adaption of the BuRST[4] format, proposed by Peter Mika [6]. The bases for BuRST [7] are RSS 1.0 [2], RDF [8], DC 1.1 [9], and SWRC 0.3 [10]. Modifications were applied where the format was outdated or underspecified. It is, for example, not possible to express affiliation in FOAF[5] other than by providing the URL of the institution. As this is not always feasible, the affiliation attribute of SWRC is suggested to represent this data in free text. A complete reference of the publication feed format can be found at [11].

[2] Semantic Web for Research Communities
[3] Dublin Core
[4] Bibliography Management using RSS Technology
[5] Friend of a Friend

See below for an exemplified item representation. The item is divided into two parts:

1. A native RSS part
2. A RDF extension part (highlighted in grey)

Both parts are linked through the burst:publication property. Information given in the RSS part of the item is mainly intended for display purposes (e.g. in RSS feed readers or widgets), and for processing in other tools which can deal with RSS (e.g. Yahoo! Pipes). The RDF extension part describes the publication in a semantically much more sophisticated way. This part is intended for tools and services that are able to process and display BuRST feeds (see sections 3.2 and 3.3), as well as semantic web applications that understand RDF.

Example of a publication represented in a BuRST feed.

```
<item rdf:about="http://know-center.tugraz.at/papers/16"
xml:lang="en">
  <title>A Storyboard of the APOSDLE Vision</title>
  <link>http://www.aposdle.tugraz.at/content/download/288/1411/file/l
  indstaedt_mayer_APOSDLE_poster_p.pdf</link>
  <description>Lindstaedt, S. N., Mayer, H. (2006): A Storyboard of
  the APOSDLE Vision.</description>
  <dc:date>2009-10-27T14:40:18+01:00</dc:date>
  <burst:publication>
    <swrc:InProceedings>
      <swrc:title>A Storyboard of the APOSDLE Vision</swrc:title>
      <swrc:author>
        <swrc:Person>
         <swrc:name>Lindstaedt, Stefanie N.</swrc:name>
        </swrc:Person>
      </swrc:author>
      <swrc:author>
        <swrc:Person>
         <swrc:name>Mayer, Harald</swrc:name>
        </swrc:Person>
      </swrc:author>
      <swrc:booktitle>Proceedings of the First European Conference
      on Technology Enhanced Learning</swrc:booktitle>
      <swrc:year>2006</swrc:year>
      <swrc:month>10</swrc:month>
    </swrc:InProceedings>
  </burst:publication>
</item>
```

The publication feed format serves two purposes: firstly, it can be understood by existing Web 2.0 infrastructure, which is capable of processing and visualizing RSS feeds. Secondly, it has the expressive power of RDF to describe publication metadata and to link entities through URIs. The example given contains a minimum set of attributes, especially addressing the "what?", "who?", "where?", and "when?". The available vocabulary is much larger, because the whole SWRC ontology can be used to markup publication metadata.

3.2 Publisher Services

The Publication Feed Publisher Services are a suite of helper services aiding individuals as well as institutions in producing, aggregating, and refining publication feeds. Services are one of the cornerstones of Web 2.0, allowing for loosely coupled systems and simple syndication [3]. The publisher services were designed according to the needs of institutions as described in scenario 1. At the moment there are three services available (via [12]):

1. The *BibTex Converter* translates BibTex to the publication feed format. It takes any BibTex file as input and converts it into a publication feed. Optionally, certain other metadata can be set, e.g. the publisher of the feed.
2. The *Publication Feed Merger* combines two or more publication feeds and ensures that item URIs are unique. If two items have the same URI, but different content, the more recent version prevails. It takes two or more publication feeds as input and provides a single publication feed as output.
3. The *Publication Feed Filter* selects relevant publications from a feed, according to a given taxonomy. It follows the "filter in" approach, which means that all publications containing one or more keywords in the taxonomy are included in the filtered feed. The Publication Feed Filter takes a publication feed and a taxonomy file as input and returns a filtered publication feed.

All publisher services were written in PHP. They are free for everyone to use, and there is no registration or API key required. To help with the orchestration of these services, a DERI Pipes [13] Installation is available at [14], along with a frontend to the BibTex converter [15].

3.3 Subscriber Widgets

The Publication Feed Subscriber Widgets are a suite of widgets for the visualization of and the interaction with publication feeds. They were designed according to the needs of researchers described in scenario 2. Specifically there are two widgets already implemented:

1. The *Publication Feed Integration Widget* was designed as a plugin to the social networking platform system Elgg [16]. It is based on Blogextend [17] and the Simplepie RSS Feed Integrator [18]. The widget allows members of an Elgg platform adding publication feeds to groups. The publications contained in these

feeds can be accessed via a common group blog. As pictured in Figure 1, individual publications are being visualized as blog post entries. Users are able to rate each publication and engage in discussions with each other by posting comments.

2. The *Publication Feed Visualization Widget* is available as a native Elgg widget and in a Wookie [19] version. It visualizes publication feed items in a faceted browser view based on Simile [20]. The faceted browser currently allows for filtering the publication feeds along the dimensions authors, publication years, and keywords, but this could easily be expanded to include other fields contained in the feeds. The filtering mechanisms are complemented with a full text search. Furthermore, a timeline visualization orders publications chronologically and allows users to intuitively browse through them. A tag cloud helps with detecting the most important keywords for a given collection of publications.

Fig. 1. Rating and commenting features of the Publication Feed Integration Widget

4 Integration into the STELLAR Network of Excellence

The publication feed ecosystem is being integrated with the STELLAR Network of Excellence. See Figure 2 for an overview of the proposed concept.

As a first step, all partners within STELLAR are asked to produce a publication feed. In the process, they are able to use the publisher services described in section 3.2 to generate their feeds. The published feeds are in turn being used to update the STELLAR Open Archive (SOA) [21], an open access platform dedicated to collecting and distributing TEL-related publications as well as the accompanying metadata. Therefore, the SOA subscribes to all of the feeds generated by the partners. The SOA is not only an archive, but it also acts as an aggregator of feeds, allowing to export all or parts of the collected publications as publication feeds. As shown in Figure 2, other tools, which are able to process RSS (such as feed readers) are able to subscribe to the publication feeds as well.

At the same time, the subscriber widgets described in section 3.3 are being deployed to TEL Europe. TEL Europe [22] is a social networking platform based on Elgg for all stakeholders in Technology Enhanced Learning in Europe, operated by STELLAR. With these widgets, users on TEL Europe are able to add relevant publications to a group in subscribing to any publication feed. The feeds might be coming from the SOA, from individual partner institutions, or indeed from any publisher of such a feed (e.g. a special interest group). The members of the group are then able to start a discussion around particular publications, and they may also add a rating. Additionally, they can visualize all feeds available on the platform for search, exploration, and trend scouting.

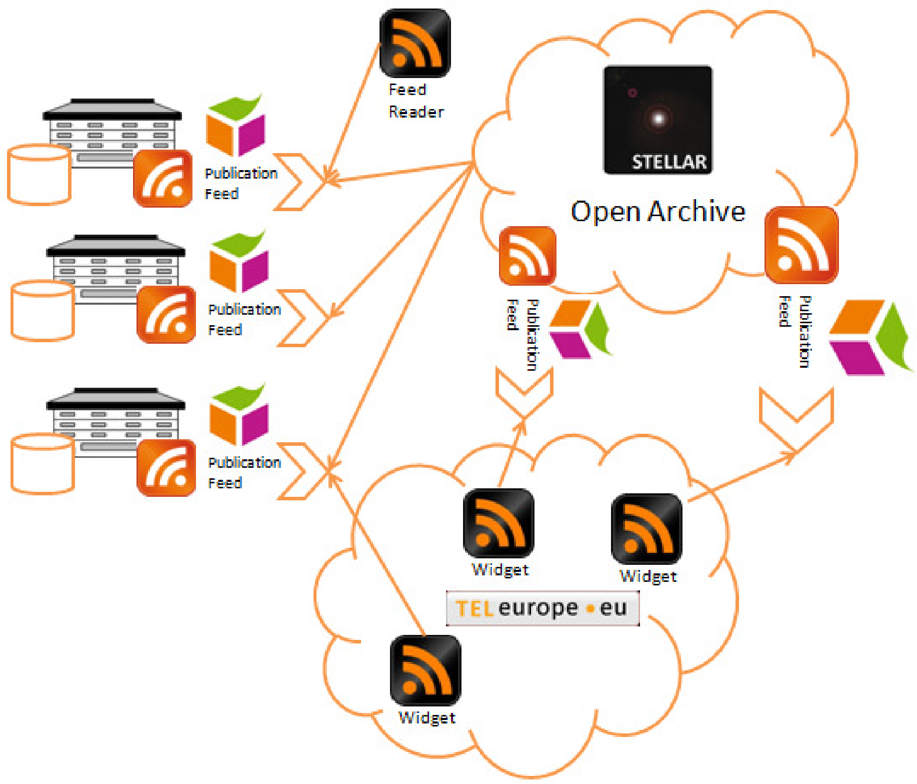

Fig. 2. Overview of the integration of the ecosystem in STELLAR

5 Conclusion and Outlook

In this paper, we presented an ecosystem for the lightweight exchange of publication metadata contributing to the prerequisites for a modern Science 2.0. In two scenarios, we showed how publication feeds can benefit researchers, institutions, and the TEL community. We described the main building blocks of the ecosystem, being (1) the feed format, (2) publisher services, and (3) subscriber widgets. Lastly, we outlined the adoption of the ecosystem by the STELLAR Network of Excellence.

The adoption process has not been finished yet, but the first results are promising. Four partners in STELLAR are actively developing BuRST feeds. Some of them have already been submitted to the STELLAR Open Archive which recently experienced a boost in the number of publications to 1038[6]. The two subscriber widgets have been deployed to TEL Europe and the first special interest groups are starting to use them.

There are certain challenges regarding the publication feed format, which have not been explicitly addressed in the first version. First, the vocabulary of SWRC could be enhanced to include more metadata, e.g. the Digital Object Identifier (DOI) of a publication. Secondly, URIs for authors and institutions would help to manage the entities in the network, and to detect duplicates. URI assignment can either be carried out by the individual institutions or a central repository. With a central repository there is no need to match corresponding entities from various sources, but it also imposes the burden of creating and maintaining said repository.

There are some possible enhancements concerning the existing services and widgets as well. For the Publication Feed Merger, it would make sense to implement a more sophisticated conflict management. This could be done by taking into account the richness of the metadata, as well as the source of information. In the Publication Feed Visualization Widget, additional fields will be added to the existing facets. Furthermore, there is no possibility for end users to correct errors in feed entries. This functionality, however, would rather have to be implemented with a large aggregator of feeds, such as the SOA.

Generally, harvesting and processing of RSS is an open issue. RSS feeds need to be fully retrieved under most circumstances; one is not able to restrict the data to just the new/updated items like in dedicated harvesting protocols, such as OAI-PMH[7]. To overcome this deficiency, we are investigating the integration of the PubSubHubbub protocol [23] into the ecosystem. In the PubSubHubbub protocol, each publisher declares a hub. Subscribers register with that hub, which in turn notifies the subscribers of new and updated items. This avoids repeated polling of the publisher's feed and relieves the subscriber from retrieving the whole feed on update.

Due to its decentralized architecture, the publication feed ecosystem can be extended by anyone. In the future, we expect to see other interested parties contributing their own components. This openness helps making the ecosystem adaptable by other research communities and is a precondition for its sustainable future.

---

[6] On 24/06/2010

[7] Open Archives Initiative - Protocol for Metadata Harvesting

6 Acknowledgement

This work was carried out as part of the STELLAR Network of Excellence, which is funded by the European Commission (grant agreement no. 231913). This contribution is partly funded by the Know-Center, which is funded within the Austrian COMET program – Competence Centers for Excellent Technologies – under the auspices of the Austrian Federal Ministry of Transport, Innovation and Technology, the Austrian Federal Ministry of Economy, Family and Youth, and the State of Styria. COMET is managed by the Austrian Research Promotion Agency FFG.

7 References

1. Waldrop, M: Science 2.0: Is Open Access Science the Future? Scientific American, 5 (298), 46--51 (2008).
2. Kieslinger, B., Lindstaedt, S.N.: Science 2.0 Practices in the Field of Technology Enhanced Learning. Science 2.0 for TEL Workshop, EC-TEL (2009)
3. RDF Site Summary (RSS) 1.0, http://web.resource.org/rss/1.0/spec
4. STELLAR: The Network for Technology Enhanced Learning, http://www.stellarnet.eu/
5. O'Reilly, T.: What is Web 2.0: Design Patterns and Business Models for the Next Generation of Software. Online at http://oreilly.com/web2/archive/what-is-web-20.html (2005)
6. Mika, P., Klein, M. and Serban, R.: Semantics-based Publication Management using RSS and FOAF. In: Proceedings of the Poster Track, 4th International Semantic Web Conference, Galway (2005)
7. Mika, P.: Bibliography Management using RSS Technology (BuRST). Online at http://www.cs.vu.nl/~pmika/research/burst/BuRST.html (2005)
8. RDF - Semantic Web Standards, http://www.w3.org/RDF/
9. DCMI Metadata Terms, http://dublincore.org/documents/dcmi-terms
10. SWRC Ontology v0.3, http://ontoware.org/swrc/swrc_v0.3.owl
11. Publication Feeds Format 1.0, http://www.stellarnet.eu/d/6/3/Publication_feeds_format_v1.0
12. Publication Feeds Publisher Services, http://stellar.know-center.tugraz.at/services
13. DERI Pipes: Open Source, Extendable, Embeddable Web Data Mashups, http://pipes.deri.org/
14. DERI Pipes@Know-Center, http://stellar.know-center.tugraz.at:8080/pipes/
15. BibTeX to Publication Feed Converter, http://stellar.know-center.tugraz.at/html/convert.php
16. Elgg – Open Source Social Networking Engine, http://elgg.org/
17. Blogextended, http://community.elgg.org/pg/plugins/antifm/read/230708/blogextended-132
18. SimplePie RSS Feed Integrator, http://community.elgg.org/pg/plugins/costelloc/read/37480/simplepie-feed-integrator/
19. Apache Wookie, http://getwookie.org/
20. SIMILE Widgets, http://www.simile-widgets.org/
21. Stellar Open Archive, http://oa.stellarnet.eu/
22. TEL Europe, http://www.teleurope.eu/
23. pubsubhubbub, http://code.google.com/p/pubsubhubbub/

Filling the Gaps to Know More! About a Researcher

Gonzalo Parra[1], Erik Duval[1]

[1] Dept. Computerwetenschappen, Katholieke Universiteit Leuven, Celestijnenlaan 200A, 3001 Heverlee, Belgium
{gonzalo.parra, erik.duval}@cs.kuleuven.be

Abstract. As one of its main goals, the Research 2.0 concept focuses on the improvement of the connection and collaboration between researchers. Within this short paper we present More!, a mobile social discovery tool for researchers. We describe the application itself and present some initial results obtained by using the tool on small scenarios. Later we describe the current challenges of the tool and the future developments. Finally, we state open problems of the field and the application itself.

Keywords: research2.0, web2.0, human computer interaction, mobile devices.

1 Introduction

Research 2.0 is the result of applying Web2.0 tools and approaches on regular research processes in order to improve practices and increase participation and collaboration [1,2]. The connection of researchers in order to nurture future collaboration is one of the key goals of the Research2.0 concept. To support this goal, social networking approaches used on commercial Web2.0 platforms are being applied for research purposes. Tools like Scopus, 2collab [3], ResearchGATE [4], Mendeley [5], Academia.EDU [6] are some examples of supporting tools to achieve this goal. Taking a closer look, academic communities are also spending some efforts to create such tools and encouraging participation of researchers. As an example in the Technology Enhanced Learning community, tools like TELeurope.eu [7] or Academic Experts [8] are being developed and used.

Due to the availability and heavy use of many Web 2.0 and Research 2.0 platforms, the users have to deal with the problem of keeping and sharing with others several electronic identities [9]. This digital identity problem is also observed in the scenario where a researcher is attending a conference presentation and is interested in finding more information about the topic and the speaker. We have addressed this need and bootstrap collaboration between researchers through a mobile application, called “More!” [10].

The structure of this short paper is as follows: we first present the implemented application and its current outcomes and limits. In the following sections, the proposed solutions to two different limitations are discussed. Later, we present the open problems and opportunities for further work. Finally, we include some initial conclusions of this work in progress.

2 The More! Application

More! is a mobile web application that groups relevant information about a speaker in a way that can be easily exposed and integrated in the normal workflow of the audience of an academic event. The application exposes the following information from the speaker:

- researcher: full name, photo, e-mail and affiliation;
- work: current paper, slides, and publications list;
- social tools handles from: Twitter, SlideShare, blog, Delicious, LinkedIn, and Facebook.

In this way, the attendee can access some regular information about the speaker; as well as the paper and slides of the current presentation; and his previous publications. Moreover, he can 'identify' and 'follow' the speaker on some of the more mainstream Web 2.0 social tools, to get access to previous, current and future work. The workflow of the application in a conference scenario is as follows:

1. The speaker exposes a QR code [11] (resolvable to an URL link) to the audience.
2. Attendants capture and decode the QR code by using any code reader application available on their smart phones. After decoding, they are redirected to the "More!" web application.
3. "More!" presents the data on the client tool.

After evaluating the usability and the functionality of the tool in a real life scenario, we noticed two big limitations in this workflow [10]. The first limitation is related to the metadata needed to feed the tool. The More! application requires research and social tools metadata, and relies heavily on the availability of such data. The problem encountered was related on how to obtain this metadata.

The second problem encountered is related to: how the QR code is exposed to the audience, the extra work required by the speaker to make the codes visible to the audience, and the poor image quality of photos for the QR decoding applications on mobile devices.

Finally, the backend and the frontend of the More! application required different approaches to efficiently solve the original problem for which the application was made.

3 Improving the Back-end: Research.fm

As presented in the previous section, we identified the need to have a common entry point and a unified metadata sharing approach to feed the application. Currently, More! is using a local database where this data is stored, but this approach is neither scalable nor aligned with the Research2.0 concept of open data. For this reason an initial approach is being developed to expose and share research metadata: the research.fm API.

The research.fm is a RESTful API that will give access to social networks and publications data of scientific authors in a standardized way. This service exposes common data requirements for applications by following the Cool URI approach in order to provide readable, logic and persistent endpoints. On the other hand, the metadata will be exposed in a standardized results format, in order to be interoperable. Table 1 shows some URIs example calls to retrieve author, publication and social tools metadata.

Table 1. URI examples to retrieve research metadata.

Social Tool	/<author_id>/social_tools
Author	/<author_id> /<author_id>/publications /<author_id>/lastpublication

The URIs provide us a logic and readable URL to obtain different kind of metadata elements from a researcher, such as: list of publications, current publication, and social tools handles. Currently, there are some discussions about how to correctly identify the authors among different platforms and how to link his digital identities. On the other hand, we are also discussing which is the way to represent the metadata and the output format for the API, in order to provide the desired interoperability. For this purpose, we are revising some publications and online community ontologies such as: SWRC [13] and SIOC [14]; together with social network approaches to share data as: OpenSocial [12] and FOAF [15].

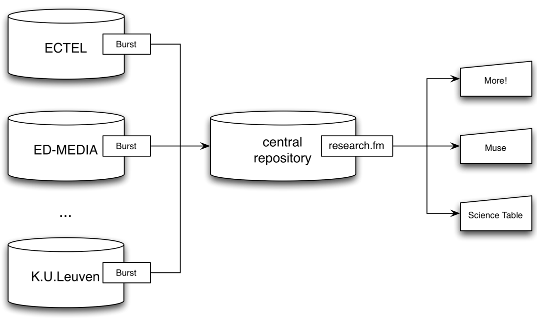

Fig. 1. Architecture where Research.fm is used.

Figure 1 presents the intended architecture to support the desired data sharing approach. Different publication sources, like publications archives and social media repositories will be included in a central repository where the metadata will be exposed through the research.fm API to different Research 2.0 tools like More! and others.

4 Improving the Front-end: Image Recognition rather than QR-Codes

The QR codes and its resolution to the More! application is crucial in order to engage the audience to use the application. As explained previously, the QR codes became a small barrier between the researchers and the solution offered by our application. Nowadays, with the large amount of open data authored and shared by users over the Internet, new possibilities are available in order to apply different approaches for the required initial fingerprint of the speaker. To be more precise, the voluntarily shared photos and tags of users in social networks, such as Facebook; can be used to apply face recognition algorithms to identify a person [16].

In order to provide the face recognition capabilities to the More! application, an external facial recognition system will be tested. Face.com provides a face recognition service that allows the analysis of facial information from photos, and identify faces from a known set of users [17]. The site provides a REST API for detection, recognition and tagging of faces in photos. The system's algorithm can be connected to a Facebook account in order to obtain the training set of photos from predefined users.

Currently we are experimenting with the different requirements to make this a successful approach. We need to find out in average how many images an author makes available in Facebook, or how many images are necessary to be able to train the face recognition algorithm. Also, we need to find the maximum distance where a smart phone camera can provide a good quality picture that can be used to detect faces. Figure 2 presents initial results of applying the face recognition algorithm to a photo captured by a mobile device.

Fig. 2. Architecture where Research.fm is used.

5 Conclusion and Future Work

The More! web application is a working prototype that is currently on its second development cycle, where the improvements described on the previous sections are being implemented and tested. On the other hand, we need to understand how tools like More! can increase the awareness about related work, or even collaboration between researchers. Does it help the research community to perform in a more effective and efficient way? To answer these questions, we are currently planning a second evaluation in practice of More! to find answers from measurable characteristics.

Currently, there is still some work needed from the Research 2.0 community regarding automatic gathering of information from the scientific publications and researchers Web 2.0 footprints and identities. Regarding scientific publications approaches like the Stellar Scientific Portal [18], DBLP [19], Mendeley are important to the scientific community in order to obtain structured and clean publication metadata. On the other hand, there are some approaches to identify and make searchable users over the Internet, like: 123people [22], Yasni [23], zoominfo.com [24], and ArnetMiner [25]. Even though there are some efforts to solve these problems, there is still much room for improvement and a long way for a sustainable solution

Acknowledgements. We gratefully acknowledge the support of the STELLAR Network of Excellence on Technology-Enhanced Learning

References

1. Waldrop, M. M. Science 2.0 - is open access science the future?. Scientific american. (2008)
2. Shneiderman, B. Computer science: Science 2.0. Science, 319(5868) (2008), 1349-1350
3. Katzen, J. Connecting researchers boosts collective intelligence. Research Information. (2008)
4. Hamm, S. ResearchGATE and Its Savvy Use of the Web. Bloomberg Business Week, (2009)
5. Mendeley. http://www.mendeley.com
6. Academia.edu. http://academia.edu
7. TELeurope.eu. http://www.teleurope.eu/
8. Academic Experts. http://academicexperts.org/
9. Neuenschwander, M. User-Centric Identity Management and the Enterprise: Why Empowering Users is Good Business. The Burton Group. (2005)
10. Parra, G., Duval, E. More! A Social Discovery Tool for Researchers. In Proc. ED-MEDIA 2010, AACE (2010), 561-569
11. ISO/IEC. ISO/IEC 18004:2006, Information technology - Automatic identification and data capture techniques - QR Code 2005 bar code symbology specification. (2006)
12. OpenSocial. http://www.opensocial.org/
13. Semantic Web for Research Communities. http://ontoware.org/swrc/
14. Semantically-Interlinked Online Communities. http://sioc-project.org/
15. The Friend of a Friend (FOAF) project. http://www.foaf-project.org/
16. Sorensen, C. Has Facebook fatigue arrived?. Toronto Star, (2008)
17. Face.com. http://face.com/
18. Stellar Scientific Portal. http://oa.stellarnet.eu/
19. The DBLP Computer Science Bibliography. http://www.informatik.uni-trier.de/~ley/db/
20. 123people. http://www.123people.com/
22. Yasni. http://www.yasni.de/
23. ZoomInfo. http://public.zoominfo.com/search
24. Tang, J., Zhang, J., Yao, L., Li, J., Zhang, L., and Su, Z. ArnetMiner: Extraction and Mining of Academic Social Networks. In Proc. ACM SIGKDD 2008, 990-998

Research 2.0: Drawing on the Wisdom of the Crowds to Develop a Research Vision

Marie Joubert and Rosamund Sutherland

University of Bristol, UK
{marie.joubert, ros.sutherland }@bristol.ac.uk

Abstract: This paper describes and reflects upon taking a 'Research 2.0' approach to developing a 'vision and strategy statement' for a network of researchers involved in researching Technology Enhanced Learning (TEL). It relates how the statement was developed first by collecting content from colleagues within the network through face to face meetings and contributions to a wiki and then by creating a coherent linear text document which further developed the content on the wiki. It discusses the risks inherent in the approach and outlines the strategies taken to address the risks. It suggests that, although the approach taken was successful, the success was limited owing to factors including a) limited engagement by the community with other people's contributions, b) a reluctance to amend other people's contributions and c) the difficulty of aggregating the multiple voices within the community while retaining faithfulness to the philosophies underpinning a 'Research 2.0' approach.

Keywords: deliverables, wiki, collaboration, analysis, #stellarnet

1 Introduction

This paper describes, and reflects on, the approach taken to developing a research 'vision and strategy' statement for the European Network of Excellence, STELLAR. The statement needed to reflect the views of a diverse community of researchers in Technology Enhanced Learning (TEL), represented by individuals from a variety of backgrounds such as computer science, engineering, education and psychology. The representatives of the community work in sixteen different labs in nine different countries in Europe and work within a wide range of research and cultural traditions. Given the diversity of backgrounds of the individuals within the community, producing a joint vision and strategy was a significant challenge.

This paper reflects on how the deliverable was produced using a 'Research 2.0' approach, critically examining the process and the products. The paper develops a use case scenario, discusses the influence of Research 2.0 on the scientific practice of developing the statement and evaluates the use of Research 2.0 tools. The paper describes the novel approach adopted and the successes and failures of the endeavour.

2 Background

STELLAR is a multi-disciplinary consortium (Network of Excellence) which aims to bring together the different research traditions and disciplines within TEL. The cornerstone of the work of STELLAR is the Description of Work (DoW), which was developed by drawing on knowledge and expertise of members of previous Networks of Excellence, Kaleidoscope and Pro-Learn.

The DoW identified three themes (called ‘Grand Challenges’) intended to be a starting point for providing a framework to identify and formalise the visons and strategies for TEL: 1) Connecting learners 2) Orchestrating learning 3) Contextualizing virtual learning environments and instrumentalising learning contexts. For each theme, the DoW also posed a number of related research questions.

One of the early deliverables for the consortium was to produce a document outlining the vision and strategy of the whole STELLAR consortium, by developing the themes in the DoW. One partner of STELLAR (University of Bristol) had ultimate responsibility for the document, but considered the vision and strategy for the consortium to be the responsibility of all partners, and wanted to find a way to for the whole consortium to contribute to the joint vision and strategy. As such, the enterprise could be seen as successful if all partners were actively engaged of in the construction of the vision and strategy.

As a Network, STELLAR subscribes to the idea of ‘Science 2.0’ as a way of working; this approach draws on ‘Web 2.0’ and can broadly be described as being underpinned by the democratic principle in which members of a community have the opportunity to contribute to a collaborative project and the contributions of all individuals are valued and become aggregated to represent the ‘wisdom of the crowds’ [1].

> ‘… Web 2.0 has been ushered in by what might be a thought of as rhetoric of 'democratisation'. This is defined by stories and images of 'the people' reclaiming the Internet and taking control of its content; a kind of 'people's internet' … This, we are led to believe, has led to a new collaborative, participatory or open culture, where anyone can get involved, and everyone has the potential to be seen or heard.’ [2]

> ‘ The Internet is enabling an unprecedented number and variety of individuals to contribute knowledge, by authoring content individually or collaboratively and by helping one another directly in online forums. [3]

We argue that, because the research approach parallels the ‘Web 2.0’ approach, it could be called ‘Research 2.0’. Research 2.0 uses tools and technologies as appropriate for the tasks involved in the research process, and these may include Web 2.0 tools, such as wikis, blogs, micro-blogs, podcasts, reference management and sharing (e.g. Delicious and Mendeley), photograph sharing (e.g. Flick*r) and social networks. (For example, see [4] and [5]). However, we argue that Research 2.0 can also use more traditional non-digital research tools to generate content such as face to face discussion, focus groups and interviews. Our key concern was knowledge creation using appropriate methods and tools.

3 Quantity and Quality of Knowledge Produced in Wikis

We suggest that we have much to learn about knowledge creation within a 2.0 approach from the use of Web 2.0 tools and hence we draw on literature relating to Web 2.0, and in particular wikis, to inform us. We focus on the literature concerning wikis for two reasons: first because Wikipedia is generally agreed to be a successful example of knowledge creation (e.g. see [6], [7]) and second because we chose to use a wiki for our knowledge creation project. This literature falls into two key areas: the first is concerned with the processes of collaborating to produce knowledge and the second with the nature and extent of knowledge itself. The literature review below is framed within these two key areas.

Processes of collaborating: Producing knowledge collaboratively using Web 2.0 technologies (wikis) is still relatively new and the concern of much literature in the area is about 'what works'. We argue that understanding online collaboration is at the heart of the 'what works' question. Coleman and Levin, 2008, put forward their view on collaboration:

> Collaboration is, we believe, primarily about people, about trust, and about the willingness to share information and work in a coordinated manner to achieve a common goal [8] (p 25).

We agree; collaboration is between people, who coordinate to achieve a common goal; in the context of this paper, this coordinated working involves sharing knowledge and building knowledge together. Those concerned are willing to share knowledge and want to share knowledge.

> Contributors' motivations seem to be critical for sustaining Wikipedia and other collaborative user-generated content outlets. [9], (p1)

As Coleman and Levine (ibid) point out, it is important to establish trust between the collaborators. This seems to be particularly important in online collaboration:

> Web 2.0 is built upon Trust, whether that be trust placed in individuals, in assertions, or in the uses and reuses of data. [9].

> … in and of themselves, these technologies cannot ensure productive online interactions. Leading enterprises that are experimenting with social networks and online communities are already discovering this fact and along with it, the importance of establishing trust as the foundation for online collaboration [10]

A further point made by Coleman and Levine is that in successful collaboration the goal is shared and that members of the collaboration have the same (or similar) end point in mind. This point was also made by Wagner and Majchrzak [11], who developed a set of enabling characteristics for successfully engaging 'customers' in a wiki through a detailed study of three cases: "Boomtown Times" (a pseudonym) wiki editorial experiment, Novell's Cool Solutions wiki, and Wikipedia. They found that if users' goals were aligned, the endeavour was more likely to succeed.

A factor that is sometimes reported in the literature as contributing to successful online collaboration concerns explicit rules related to contributing content. Wikipedia includes a page of 'rules' and 'guidelines' which are described as a 'policy, a widely accepted standard that all editors should normally follow. Changes made to it should reflect consensus.' (see http://en.wikipedia.org/wiki/What_Wikipedia_is_not). Wagner and Majchrzak (ibid) suggest that these guidelines ensure quality:

> Wikipedia has strong editing guidelines that are motivated by the refactoring rules of software development and principles of objectivity. This ensures that articles, which might have suffered in readability from the disjointed work of multiple contributors and commentator, ultimately becomes very readable again. [11]

However, while there are some who consider that rules encourage contribution to the wiki, such as Wagner and Majchrzak (ibid), others have found that the presence of rules makes little difference, (e.g. [12]).

Finally, it seems that constructive engagement could encouraged by allowing different levels of participation; 'lurking', commenting on others' contributions, making original contributions, editing and asking for explanations of others' ideas and organisation of content for better structure. [11,12]

Quality of knowledge: Wikis can be successful tools for collecting and aggregating knowledge. As pointed out above, WikiPedia, probably the best known wiki, is generally seen as a success. At the time of writing this paper (July 2010) it had over 3 million articles in the English version, and it is in the top ten web sites accessed anywhere. This demonstrates that it is possible to create a wiki that 'works' in terms of community engagement. There is debate, however, about the quality of the knowledge on wikis.

Whereas wikis sometimes have rules of engagement, the knowledge produced on wikis is usually not subject to editorial control which leads to concerns over the provenance of information posted. Concerns relate to various aspects of knowledge, largely to do with the accuracy of knowledge. For example, Don Fallis (2008) suggests that:

> serious concerns have been raised about the quality (e.g., accuracy, completeness, comprehensibility,etc.) of the information on Wikipedia [13] (p 1663)

Fallis' article suggests that Wikipedia has been dismissed by much of the library and information science communities because it is seen as unreliable. He presents a thorough analysis of potential different types of inaccurate information in terms of factual accuracy, completeness, currency and comprehensibility and he demonstrates that Wikipedia fails rigorous tests of accuracy in these respects. However, he continues by arguing that Wikipedia is 'quite reliable' and 'quite verifiable' and that it contains 'quite a lot of high-quality accurate information' (p 1669). He makes the point that 'it is probably epistemically better … that people have access to this information source'. (p 1669). He argues that there are ways in which the reliability of information on Wikipedia can be improved, but points out that the cost of this would undermine some of the values on which the project is based, such as the number of contributions and the speed with which entries are added and updated. His key point is that ultimately it is the responsibility of readers 'to decide whether to believe what they read on Wikipedia' (p 1671) and he concludes by suggesting ways in which to help readers in this respect (e.g. signaling evidence of the quality of articles, directing readers to further reading, flagging omissions).

Concerns over the accuracy of information on wikis and Wikipedia in particular frequently relate to factual content (and this is to be expected in the case of Wikipedia which collects 'facts'). However, there are other concerns which relate to the quality of knowledge built using online collaboration. For example, Anderson [5] argues that the 'Web of Content' (WoC) discourages 'a deep level of critical thinking' because development of content is influenced by a 'powerful zeitgeist'. The computer scientist, Jaron Lanier, in an essay about the dangers of elevating collectivism above merit and thus lowering standards, describes a similar concern:

> What I've seen is a loss of insight and subtlety, a disregard for the nuances of considered opinions, and an increased tendency to enshrine the official or normative beliefs of an organization. [14]

This section has outlined some of the key issues relating to the collaborative production of knowledge within an online environment, with a focus on the use of wikis. It demonstrates the keys risks associated with using a wiki in terms of the amount of knowledge produced and the quality of the knowledge. In terms of the former, the main risk seems to be non-participation in the process of knowledge building and we recognised within our project that we may need to take steps to encourage our colleagues in STELLAR to contribute to the wiki. In terms of the latter, the risk for us was less clear. Our project was not essentially about collecting facts, as Wikipedia is, and we did not consider that we risked inaccurate contributions. Our project was more about developing arguments, debate, insight and vision and did, perhaps, run the risks described by Anderson and Lanier above. These risks were less clear to us at the beginning of the project but as it developed we put strategies in place to encourage high quality debate.

4 Developing the Vision and Strategy Statement

4.1 Starting Points

The text from the DoW was used as a starting point to create a 'Grand Challenges' wiki. The text was pasted into three main pages, one for each of the three Grand Challenge themes. At the same time, the wider STELLAR community was asked to recommend reading related to producing a TEL vision and strategy statement. The recommended readings and were put together and distributed to the STELLAR network and posted onto the STELLAR web site. Members of STELLAR were asked to engage with the readings prior to the face to face meeting described below.

4.2 Face to Face Meeting

A day-long face-to-face meeting was set up in Bristol in May 2009 (month 4 of STELLAR). 33 members of STELLAR participated and worked in three groups, each with a chair and a note-taker. The groups were constructed to include individuals who represented the diverse research interests and perspectives within STELLAR.

In the morning there were two discussion sessions. Participants remained in the same groups for both these sessions although the chairs and note takers were different.

In the first session groups discussed questions relating to the Grand Challenge theme 'connecting learners'. Each group was given one of three questions to discuss:

- What are key enabling and success factors for learner networks?

- What impact could web 2.0 technologies have on learning in educational institutions and what are the implications for a) professional development b) design and organisation of learning spaces c) policy makers?
- What are the changing demands for workplace knowledge and skills and what are the implications for a) leaders and managers and b) the workforce?

In the second session groups discussed questions relating to the Grand Challenge theme 'orchestrating learning':

- What is the role of the teacher/more knowledgeable other in orchestrating learning and how does this relate to collaboration and the knowledge of students?
- What is the role of assessment and evaluation in learning and how can technology play a role?
- From the point of view of the learner what is the relationship between higher-order skills and learning of a particular knowledge domain and what is the role of technology in this respect?

For the third session (which took place in the afternoon), participants were put into new groups. These groups discussed questions relating to the Grand Challenge theme 'Contextualising virtual learning environments and instrumentalising learning contexts':

- How can new forms of technology-enhanced learning enable novel experiences for learners and for development of human competences and capabilities?
- How can the mobility of the learner in distributed and multi environment learning settings be supported, to include the transition between a) real and virtual contexts b) informal and formal learning contexts?
- Which standards are needed to achieve interoperability and reusability of learning resources in this field? How can we harmonise the existing learning standards?

The main purpose of the meeting was to expand the collective understanding of the community concerning the three research themes, through knowledge contributed by experts within the community and discussion and development of related research questions. The meeting was set up using an adaptation of the 'knowledge café' methodology (Firestone and McElroy, 2005). Within this methodology discussion is not driven by an agenda, and this is seen to encourage groups to develop discussion in line with the expertise and interests of the individuals in the group.

Note-takers were told that the notes would be added to the wiki but otherwise were not given any specific instructions or guidelines. They adopted different approaches but generally attempted to capture as many of the points being made as possible, not attempting to organise the points into coherent prose. The examples below are taken from discussion starting from the questions 'What are key enabling and success factors for learner networks?' and 'What is the role of the teacher/more knowledgeable other in orchestrating learning and how does this relate to collaboration and the knowledge of students?' The examples demonstrate different approaches taken to note taking.

Example 1

This is the first set of aspects created in the first grand challenge vision workshop on May 20th, 2009 in Bristol:

- Connections with people with whom you interact
- Merging of Formal & informal, Lifelong, Self-organised / self-constructed,
- One holistic network per person, not a private one, professional one…

- Medium used for communication is fundamental; Software can support maintenance and building of network
- Challenge: Integrate networks with learning processes
- Most prominently: Social network; but not only people: Networks of people, artefacts (e.g. paper), and tools (distributed cognition, actor-network theory)
- Sense of being in control essential (when to use, how to use, …) / responsibility

Example 2

What does a more knowledgeable other offer? A frame of reference/organised state of mind, knowledgeable other takes a scaffolding role - metalevel role - from research on expertise. Not just content knowledge - pedagogy as a whole - mediating content - children in school unlikely to have pedagogical expertise, but just more content knowledge. Teacher required to facilitate knowledge transfer/representation. Maybe there is a changing role of teacher within 21st century - but not necessarily to do with technology.

In one group, the notes were entered directly into the wiki and in the others they were written in a word-processed document and pasted into the wiki. These notes were seen as the starting point for extending the community's understanding of the Grand Challenges and the plan was to develop them into a more coherent whole over a period of weeks to form a substantial part of the vision statement. Importantly they were faithful to the spirit of the Research 2.0 approach in that contributions from all individuals were valued and the notes represent the collective responses of the community to the nine Grand Challenge questions chosen as the starting point.

4.3 Online Collaboration

After the Bristol meeting STELLAR partners were invited to join a small team to coordinate the ongoing contributions to the wiki (to be called the D1.1 team). Apart from the Bristol team (UB), five partners volunteered: Istituto Tecnologie Didattiche in Italy (ITD), Ludwig-Maximilians-Universität München in Germany (LMU), Centre for Social Innovation in Austria (ZSI), Know Centre in Austria (KC) and Université Joseph Fourier in France (UJF). UB took a leadership role, with other team members taking responsibility for provoking STELLAR members to contribute to a particular subsection of the wiki related to:

- connecting learners (ITD and ZSI)
- orchestrating learning (LMU and KC)
- contextualising virtual learning environments and instrumentalising learning contexts (UJF and UB)

In the first half of June 2009, the D1.1 team met once online (using FlashMeeting, see http://flashmeeting.open.ac.uk/home.html) to discuss how to proceed. Following this, UB put together a written plan which outlined a tight time-frame for the development of the wiki:

- 22/6/09 to 6/7/09 – intensive work by all D1.1 team to get contributions from the whole STELLAR community.

- 6/7/09 to 30/7/09 UB will take responsibility for developing the wiki into a deliverable. Other D1.1 team members will be asked to contribute by a) writing sections b) reviewing sections and c) clarifying sections where necessary

UB also suggested strategies for the D1.1 team to use to provoke colleagues to contribute to the wiki. For example, written suggestions included:

> For example, there might be a part of the wiki which you think requires further development; you could use this as a basis to develop a question for people to answer. You might make a sub page with this one question and invite people you know have expertise in the area to contribute a paragraph.
>
> You might find that two people are making similar points, or two people are disagreeing, it might be worthwhile pointing out the synergies and encouraging further debate. However it could be important to find a way of keeping the 'disagreements' in the document.

The team met online again in the third week of June to discuss progress and to kick-start the phase during which the D1.1 worked intensively with colleagues to encourage them to contribute. Towards the end of this phase, one member of the UJF team came to work intensively on the wiki with the UB team for three days in the final week of July 2009.

This section has described the ways in which the online collaboration was organised. The next two sections reflect on the results of the online collaborations in terms of a) the extent of engagement of the STELLAR community and b) the nature of the contributions.

5 Reflections

5.1 Extent of Engagement with the Wiki

The wiki includes functionality to record the editing history of pages; an example covering the editing history of one page over the period of eight days is provided below:

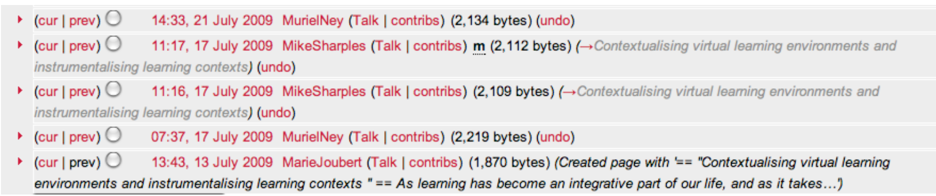
(cur | prev) 14:33, 21 July 2009 MurielNey (Talk | contribs) (2,134 bytes) (undo)
(cur | prev) 11:17, 17 July 2009 MikeSharples (Talk | contribs) m (2,112 bytes) (→Contextualising virtual learning environments and instrumentalising learning contexts) (undo)
(cur | prev) 11:16, 17 July 2009 MikeSharples (Talk | contribs) (2,109 bytes) (→Contextualising virtual learning environments and instrumentalising learning contexts) (undo)
(cur | prev) 07:37, 17 July 2009 MurielNey (Talk | contribs) (2,219 bytes) (undo)
(cur | prev) 13:43, 13 July 2009 MarieJoubert (Talk | contribs) (1,870 bytes) (Created page with '== "Contextualising virtual learning environments and instrumentalising learning contexts " == As learning has become an integrative part of our life, and as it takes…')

Figure 1: Editing history of a wiki page

This information allows us to analyse the extent of engagement. Overall about 20 people from STELLAR contributed to the wiki in the period of development from 22nd June to 6th July 2009. However, sometimes a contribution under one name represented a collation of several contributions from an institution so it could be argued that there were more contributors. The majority of the contributions were made by a small number of people, usually within a short time frame. For example, the three main pages: ‘Connecting Learners’, ‘Orchestrating Learning’ and ‘Contextualising Virtual Learning Environments and instrumentalising learning contexts’ pages had the following contributions:

Table 1: Contributions to the three main pages of the wiki

Page	**Name**	**Date (in 2009) and Number of Contributions**
Connecting Learners	Marie Joubert	13 July (1)
		16 July (1)
	Rosamund Sutherland	14 July (1)
		28 July (2)
		30 July (2)
	Nicolas Balacheff	26 July (5)
		30 July (2)
	Stefanie Lindstaedt	28 July (3)
Orchestrating Learning	Marie Joubert	13 July (1)
		16 July (1)
	Rosamund Sutherland	14 July (1)
		28 July (1)
		30 July (4)
	Nicolas Balacheff	26 July (5)
		30 July (2)
	Stefanie Lindstaedt	28 July (3)
Contextualising Virtual Learning Environments and instrumentalising learning contexts	Marie Joubert	13 July (1)
	Muriel Ney	17 July (1)
		21 July (1)
	Mike Sharples	17 July (2)

When individuals were asked to contribute by adding content, explanation or examples, generally they were very willing to do so. For example, when UB approached the Open University of the Netherlands (OUNL) asking for a clarification of what is meant by ‘interoperability’, the response was immediate and detailed.

Most people who contributed used the ‘Edit’ function to enter text directly into the wiki, either by adding in new text or amending text already present. A few used the ‘Discussion’ function.

The D1.1 team made concerted efforts to encourage contributions, but as their comments suggest, this was not always easy:

‘We have done really our best to obtain inputs and feedback, but it has been a hard task’ (email communication).

They went on to suggest that it had been difficult because people were not motivated to contribute because they did not understand the origins of the wiki and did not know what its purpose was.

Authorship was also seen as an issue for a number of reasons. There were conflicting ideas about whether or not to acknowledge individual contributions,

I am working on the wiki this week (until Friday). Although everything will appear under my name, I am integrating contributions from different people of my group. Thus I would like to let you know that VL and JP should also be mentioned in case there is a list of authors in the end (email communication).

Others were concerned about the extent to which it was appropriate to edit/modify/add to/ delete the contributions of other people. There seemed to be a tension between valuing and respecting other people's contributions (and not vandalising the wiki) but at the same time building the best possible document. As one contributor suggested, he was happy as an academic to use a word processor and the 'track changes' tool to write collaboratively. He suggested that using track changes can be seen as a way of checking with the original author that changes are acceptable; in other words track changes points out the suggested changes (which can then be accepted or rejected). In a wiki, however, the changes are not so obvious and anyone interested in the changes made would have to make a small effort to access the trail of devlopment.

Many of those who did make changes seemed to need to check the changes they had made with the original authors. For example:

> 'I have done a bit of re-organisation, tell me if I am barking up the wrong tree' (email communication).

There was some debate about writing IN the wiki as opposed to writing in a word processor. There were some who thought that it was much easier to do the latter, but others who argued that this meant that the full authoring trail would be lost. Again, there was some debate about the authoring trail and about how important it is to retain the trail. On a similar note, there was a comment that sometimes people try to be the 'last author' in a wiki that is going to be frozen at a given time, because then their voice will be heard.

Finally, a possible barrier to contributing to the wiki may have been the technical difficulty of logging in to the wiki. We do not consider it to be very difficult, but it seems that some people found it confusing. For example, one STELLAR emailed to say:

> 'Unfortunately, it appears that I can't log in to edit it despite I can log in http://www.stellarnet.eu/'. (email communication)

5.2 Nature of Contributions

The contributions varied in style and length. In general, they tended to take the form of paragraphs setting out the perspective of an individual. The first example, below, takes the form of an explanation about the meaning of 'interoperability', provided in response to a direct request from the D1.1 team (mentioned above). This response was sent by email.

> Essentially this is about sharing resources and tools and system spanning. Within the community several specifications/standards are used. Basically there are several standards of content exchange that allow for exchange of learning content between different platforms. Furthermore interoperability is an important topic that considers more the functional integration of different learning services.

The D1.1 team found this sort of explanation to be very helpful as a starting point but found that contributions were seldom expanded, by either the original authors or other colleagues, with arguments, examples or references.

The second example below starts from 'taken as read' assumptions (contexts are more fluid) to suggest a change in focus for educational theory. It goes on to wrap up the paragraph by arguing against polarisation of educational theories.

> When the context was relatively stable (in the case of fixed classrooms) educational theory tended to focus on content. However now that contexts are more fluid there is a shift from a focus on 'content' to a focus on 'context'. However such a polarisation of 'content' and 'context' might be unhelpful in terms of understanding issues related to learning and knowledge construction.

The D1.1 team found this paragraph helpful and interesting, but again noticed that there were no further contributions to the paragraph.

In general, the D1.1 team found that the contributions on the wiki were individually valuable but that the levels of engagement with other people's contributions was disappointing. There was little evidence of individuals challenging other people's contributions or questioning what they had said, but typically were more concerned with phrasing and style. This is demonstrated by the example below, in Figure 2, which was taken from the editing history of the page on Orchestrating Learning. The text on the left is the earlier version, and the text on the right is an edited version.

- The development of digital ~~technology~~, their ~~interface~~ and association ~~to~~ communication technology, has ~~open~~ the possibility ~~to access to~~ a large diversity of learning tools and all ~~kind~~ of resources, as well as new infrastructures to support interactions and communications among learners and teachers or trainers -- or in more general terms, among learners and knowledgeable others. This evolution is supported by the ~~mergence~~ of theoretical frameworks which ~~provides~~ new means to understand learning and to design more efficient and more relevant environments to support it. Situated cognitiion and situated learning theories, collaborative learning, exploratory learning as well as mobile learning theories are stimulating new approaches to learning, pedagogy, didactics and assessment.

+ The development of digital technologies, their interfaces and association with communication technology, has opened up the possibility of accessing a large diversity of learning tools and all kinds of resources, as well as new infrastructures to support interactions and communications among learners and teachers or trainers -- or in more general terms, among learners and knowledgeable others. This evolution is supported by the emergence of theoretical frameworks which provide new means to understand learning and to design more efficient and more relevant environments to support it. Situated cognitiion and situated learning theories, collaborative learning, exploratory learning as well as mobile learning theories are stimulating new approaches to learning, pedagogy, didactics and assessment.

Figure 2: Example of edited text

6 Producing the Deliverable

In order to produce the final document – a linear text document – the text was copied from the wiki into a word processor document. A UB team of two took responsibility for editing it. This involved forming it into a coherent narrative, removing repetition, adding references, examples and explanations and amending text to achieve consistency in language and style.

A draft final document was completed. Once again, the UB team felt that it was important, even at this late stage, to work within a Research 2.0 approach and so the document was distributed to the whole STELLAR community with a request for feedback. In particular, the community was asked to check that any contributions they had made had been represented in the way they wanted.

Two members of the community were asked to provide internal peer reviews and a final version was produced, taking into account the feedback from the community and from the internal peer reviewers.

7 Conclusions

The aim of the project described in this paper was to use a Research 2.0 approach to develop a vision and strategy statement for the STELLAR network. This paper described the processes and reported on the outcomes. This concluding section reflects on the project and ends with some recommendations.

We claim that the project was successful in many respects; members of the community did make contributions and the D1.1 editors were able to produce a deliverable based on the contents of the wiki. We suggest that the success of this way of gathering the views of the community can be explained by the existing 'pre-conditions' for a successful online collaborative venture, as outlined in the 'Quantity and quality of knowledge produced in wikis' section above. In particular the members of the community were willing and able to share knowledge and had, by the end of the Bristol meeting, developed a level of trust. On the whole, we could claim also that the community had a common goal, although – as reported above – perhaps this was not clear to all colleagues.

However, we were slightly disappointed that the D1.1 team had to work so hard to encourage the community to engage more deeply with the wiki and that many of the contributions were less well developed than we had hoped. As described above, the D1.1 team realised, as the project unfolded, that there was a risk that contributions may be less well formed and debated than hoped for, and made efforts to encourage deeper engagement.

Finally, we reflect on the Research 2.0 approach we took. This approach aimed to draw on the wisdom of the crowds (in this case STELLAR) and to aggregate the multiple voices of the individuals in the community in order to develop a coherent and unified vision and strategy for the community. However, the crowd had many voices and the spirit of 2.0 suggests that each should be valued and heard; the problem for us was that we could not aggregate all the voices while remaining faithful to the Research 2.0 philosophy underpinning our project. It may be that listening to the multiple voices of the crowds is at odds with forming an aggregation and it may be that we have to re-think how we conceptualise an 'aggregation' (particularly an aggregation of visions).

As pointed out above, the use of the wiki was perhaps not as successful as we hoped. We suggest that this was the case despite the will and technical ability of the community to contribute. We do not fully understand why we were not as successful as we hoped, but we have some speculative suggestions:

1) Although it seemed that a good level of trust was present at the beginning of the project, STELLAR was a very new community and relationships within the community were still at an early stage. People did not know one another well and may have felt timid about making contributions. This paper has been written almost a year since the D1.1 project came to an end and in the intervening months the community has developed and grown, and (crucially) may be more willing to take the risk of publicly contributing to a growing wiki because of developing trust.

2) The construction of the wiki meant that it was difficult to engage with. There was too much text on each page, often well crafted, which did not seem to encourage discussion.

3) Members of the community did not seem to be clear about the goals of the wiki and how it would contribute to the vision and strategy of STELLAR. They therefore did not know what they should and should not be posting onto the wiki. Importantly, the project was not a research project; it was something different and therefore difficult to engage with.

4) Individuals were reluctant to change text that others had posted and others were reluctant to have their text changed.

In further work on developing STELLAR's vision and strategy, we intend to continue with the approach we used to produce this deliverable, and to experiment in the following ways:

- reduce the amount of text on each page and include prompts to encourage discussion
- make the hopes and intentions of the wiki (and the project) clear
- encourage the use of the 'Discussion' feature of the wiki to overcome the reluctance to change other people's entries
- make it clear that the wiki is a collaborative effort which is based on a Research 2.0 approach and is therefore about building knowledge together in a way that combines the voices of all the community.

References

1. Surowiecki, J., and M. P. Silverman: The wisdom of crowds. American Journal of Physics 75:190 (2007)
2. Beer, D., and R. Burrows: Sociology and, of and in Web 2.0: Some initial considerations. Sociological Research Online 12 (5) (2007)
3. Adamic, L. A., Wei, X., Yang, J., Gerrish, S., Nam, K. K., Clarkson, G. S., et al: Individual focus and knowledge contribution. First Monday; Volume 15, Number 3 - 1 March 2010. Retrieved from http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/2841/2475 (2010)
4. Ochoa, X., & Duval, E.: Quantitative Analysis of User-Generated Content on the Web. http://journal.webscience.org, 34(1). Retrieved from journal.webscience.org/34/1/WebEvolve2008-03.pdf (2008)
5. Andersen, P. : What is Web 2.0?: ideas, technologies and implications for education: JISC (2007)

6. Kittur, A., Suh, B., Pendleton, B. A., & Chi, E. H.: He says, she says: Conflict and Coordination in Wikipedia. Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '07, 453. New York, New York, USA: ACM Press. doi: 10.1145/1240624.1240698 (2007)
7. Ebner, M., Zechner, J., & Holzinger, A.: Why is Wikipedia so Successful ? Experiences in Establishing the Principles in Higher Education. In Proceedings of I-KNOW ’06 (2006)
8. Coleman, D and Levine, S: Collaboration 2.0: Technology and Best Practices for Successful Collaboration in a Web 2.0 World Happy About; 2008:320 (2008)
9. Nov, O.: What motivates Wikipedians. In Communications of the ACM 50 (11) (pp. 60-64) (2007).
10. Miller, P.: Web 2.0: Building the New Library. Ariadane. Retrieved from 43131693cd2b7525be068ce24c2596cb031e9884 @ www.ariadne.ac.uk (2005)
11. Kwan, M., & Ramachandran, D.: Trust and Online Reputation Systems. In J. Goldbeck, Computing with Social Trust - Human Computer Interaction Series (pp. 287-311). London: Springer-Verlag. Retrieved from http://dx.doi.org/10.1007/978-1-84800-356-9_11(2009)
12. Wagner, C and Majchrzak, A: Enabling Customer-Centricity Using Wikis and the Wiki Way Journal of Management Information Systems. 23(3):17-43 (2007)
13. Goldspink, C., Edmonds, B. and Gilbert, N.: Normative Behaviour in Wikipedia. 4th International Conference on e-Social Science, Manchester, June 2008. http://epubs.surrey.ac.uk/cress/24/ (2008)
14. Fallis, D. Toward an Epistemology of Wikipedia Journal of the American Society for Information Science. 59(May):1662-1674 (2008)
15. Lanier, J. “Digital Maoism: The Hazards of the New Online Collectivism,” Edge 183 (May 30, 2006), http://www.edge.org/3rd_culture/lanier06/lanier06_index.html (accessed October 25, 2010).

Research at the table

Bram Vandeputte and Erik Duval

Katholieke Universiteit Leuven, Department of Computer Science
Celestijnenlaan 200a, 3001 Leuven, Belgium
{bram.vandeputte,erik.duval}@cs.kuleuven.be

Abstract In this paper we describe how we want to take advantage of the rapid developments in technology to assist researchers in doing research. More specifically in exploring the publication space. For this purpose we have designed and developed a prototype application to take advantage of large displays with multi touch enabled input. We describe the current state, the next steps and how we will to evaluate it. To conclude we give an outlook on further possibilities and challenges that lay ahead.

Key words: research2.0, information visualization, multi touch, large display, research.fm

1 Introduction

How great would it be to integrate the process of exploring publications, finding them and reading them in an almost seamless way ? This sort of idea was already described in *As we may think* in 1945 by Vannevar Bush [2]. The Memex was described as the perfect desk of a researcher, having all the knowledge of the world readily available. At that time personal computers were not even invented, but since then technology has advanced tremendously and become very common. Using current state of the art technologies, we want to find out how we can ease the process of exploring publications. This process is an important part of a researcher's job, as he wants to know what is going on in his field of research.

To be able to get this kind of understanding, Russell et al [10] have pointed out that it is imperative that the right representation is found for exploring a network of (publication) data.

The idea of visualizing publication networks has been inspired by the work of Klerkx et al, where they explore learning object repositories [8] and social bookmarks [7] in a visual manner.

In this paper we first introduce and describe the problem. We then motivate our hardware platform, describe the origin of the data and we explain the detailed workings of the application. In the next section we compare our work with existing studies. Then we describe how we evaluate this and finally we propose the next steps to be taken. To conclude we summarize our findings and discuss further possibilities.

2 Problem statement

An important part of a researcher's job is reading scientific papers. This ensures that the researcher is up to speed of what is going on in his research field. It is also a prerequisite for writing scientific papers, as handbooks such as the one from Robert A. Day [3] emphasizes.

There are three basic ways of dealing with scientific papers. There is active search, where you search for a particular paper or a 'good' paper on a specific topic you have in mind. There are dozens of websites that serve this purpose really well, such as Google Scholar[1], ISI web of Knowledge[2], DBLP[3] etc. There is also what we can call passive search, where you get alerted whenever new publication material is available. Google Scholar has recently added a feature where you can be alerted whenever something new comes up that matches certain keywords. Also many of the journal magazines let you subscribe to a list to send you the table of contents when a new issue is available. Finally one can focus on relations between papers and authors. There are existing tools where this is possible, but we think that there is not enough technical support available for exploring these networks.

To explain the problem we want to solve, we will briefly describe the use cases we want to tackle with this work. The use cases can be grouped into two categories. In the first category the use cases have a mainly *top down* approach, while the second category holds the use cases that typically need a *bottom up* approach.

2.1 World overview

Typically, in this use case a user would like to start with a complete overview of all nodes laid out in a graph. The user then wants to zoom in on parts of the graph that draw her attention. This can be used to find out patterns or clusters. In this case the user usually is already an expert in the field, trying to understand or improve his knowledge about the field.

2.2 Explore your neighbors

In this case you might want to start from a view with a focus on yourself, or the author or paper that you want to start from. Then you want to browse to nodes in your 'neighborhood', which are likely to be related and/or interesting. Here you can try to find answers to questions like : Where am I in the research publication space ? Who should I talk or connect to ?

[1] http://scholar.google.com
[2] http://apps.isiknowledge.com
[3] http://www.informatik.uni-trier.de/~ley/db/

3 The application

3.1 The hardware

The input modalities We chose for supporting a multi-touch setting, as we want to explore direct and multi touch capabilities. This to find out whether these relatively new input methods can help to make it easier for researchers to interact with the fairly complex graph like structures.

The display The application will entail a visualization of a deeply connected network containing up to hundreds (maybe thousands) of nodes. This property feeds the need for using a large display. These large displays, with increasingly higher resolutions, are also rapidly becoming cheaper and more common, which makes it easier to include them in our study and makes this study more relevant.

A problem that sometimes arises on multi touch input devices is when one touches the screen to give input, the finger or hand occludes information one wants to see at that moment. This can be solved in two ways, either we make the information appear next to the touch point, or we make the information bigger so it is less likely to be occluded. Both solutions can benefit from a larger display, as you have simply more space to put the information.

Studies by Forlines et al [5] and Kin et al [6] have already shown that on tabletop displays multi touch input has performance and spatial awareness advantages over the traditional mouse, which reinforces our choice of hardware. From a research perspective, we want to explore if and how a large screen estate can influence the possibilities of this kind of visualization.

3.2 The data

EC-TEL conference Our first scope was to visualize all the publications from all editions of one conference. We extracted metadata from papers and put them in a database. Unfortunately this extraction process is still very error prone and a lot of semi-manual cleaning up needed to be done. The approach took quite a bit of effort and is not very scalable.

To try and make access to these publication data easier, we propose an open architecture for exchanging these publication metadata. This architecture is currently being discussed and developed in the STELLAR project[4], with both suggestions for collecting these data using BuRST feeds[5] and a webservice API, called research.fm[6], to make them available for tools and widgets like the one we are describing in this paper.

[4] http://www.stellarnet.eu/
[5] http://stellarnet.eu/d/6/3/BuRST_format_adaption_discussion
[6] http://www.stellarnet.eu/d/6/3/KULDocumentation

3.3 The network and the visualization

The obvious relations to visualize are the paper-author relations, and also co-authorship. To build up this network, we want to have a self-organizing and self-declutering algorithm. We chose to use Traer Physics[7], an implementation of a simple particle system physics engine, which allows to combine a spring-graph algorithm with physical forces. This combination will take care of the organizing and declutering of the network, so we don't have to care about where to put the nodes. After experimenting with the parameters such as force, drag, mass of the particles, spring length and strength, ... We could see a clear network-like graph appearing when the network is stabilizing after a few seconds.

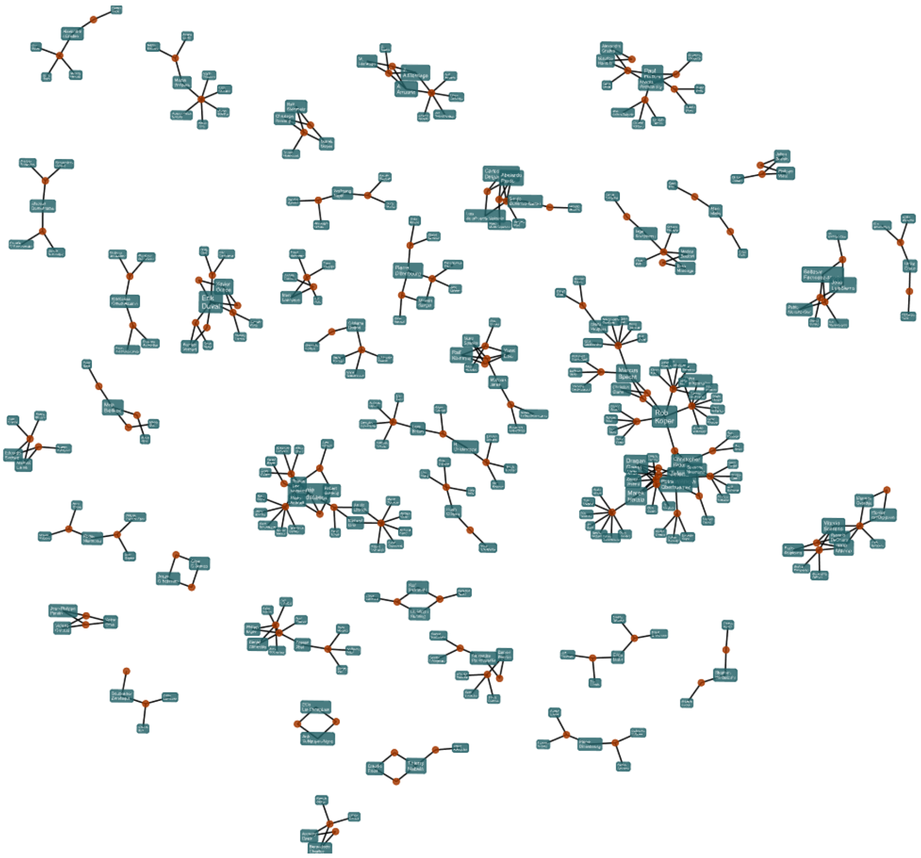

Figure 1: Overview of the whole publication network. The green nodes are authors, while the red ones are papers.

[7] http://www.cs.princeton.edu/~traer/physics/

Figure 1 shows a screenshot of the visualization in the overview state. All nodes present in the network are shown. This state addresses the first use case we described in section 2.1. It can help a researcher to find out whether there is a lot of collaboration going on in this field, where the biggest clusters can be found or who the most active authors are.

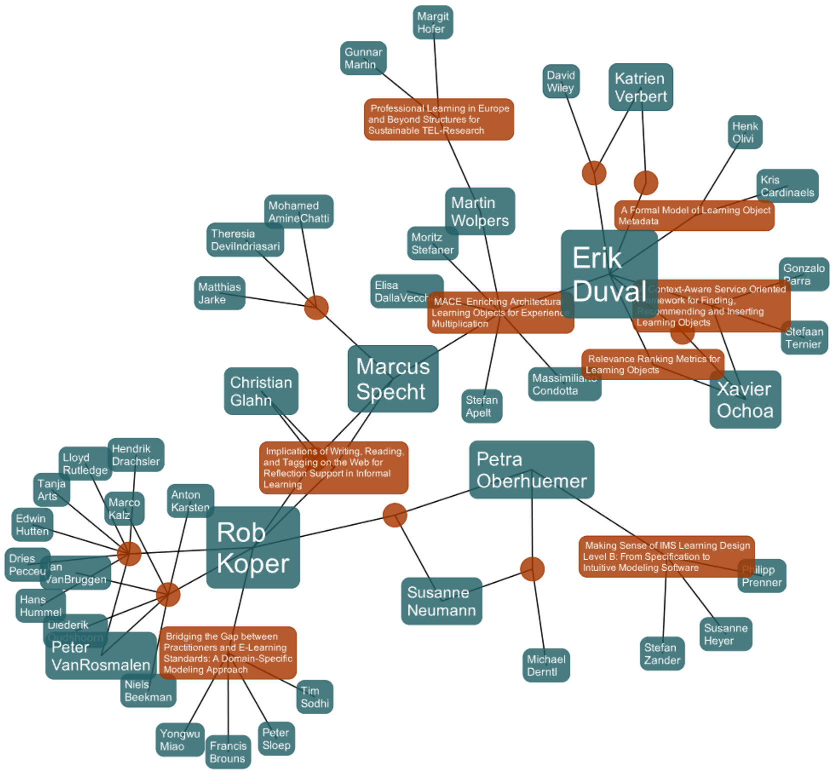

Figure 2: A detailed view of related authors. The green nodes are authors, the bigger they are, the more papers they have published. The red nodes are papers, where some of them have been expanded to show the title of the paper.

The second use case described in section 2.2 benefits from the view as shown in Figure 2. Here the visualization is zoomed in on a specific target. All the author names become clearly visible, so you can find an author very relevant for your work. One can also click on some paper nodes to get more information on the paper itself, so to find papers that are interesting, for example because they are closely related to your work. As you can see we are already experimenting with varying the node size of the author, based on his number of publications, to denote importance of this author.

4 Related Work

There are numerous other visualizations of publication data existing already. In this section we will highlight some visualizations that try to solve similar problems, and we will shortly describe how each of them differ from our approach.

4.1 Papercube

When this web application[8] first opens up, it immediately shows you a search box. This is useful when you are looking for something more specific, but it does not help when you want to explore the publication space and don't have a specific entry point in mind. There are quite some possibilities both in terms of relations and type of visualizations, so it can take a while for someone to get used to the interface and find what one actually wants. In our approach on the other hand, we want to make it easy for starting the exploration phase by directly showing the data. In this visualization the data is shown in a spring graph with a good lay-out. When you hover over a paper, the relations to other papers are highlighted, which is very helpful. One can also directly click through to the paper itself, so if you have found an interesting publication you can directly retrieve it online. Bergström et al [1] evaluated this application, and found that the users unanimously said that this kind of visualization can usefully augment existing digital libraries.

4.2 Ed-Media Relation Browser

The Ed-Media Relation Browser[9] is also an interactive, browser based, author visualization. In this approach they focus on one person and its direct relations, assisted with a strong filtering mechanism. The visualization only starts after you have entered a name. This emphasizes their focus on solving the problem of getting to know closely related authors. It does not allow one to study the field nor to discover the indirect relations between authors and papers. In our approach we try to solve this problem by allowing to zoom in on a specific person, but with a global navigation strategy so that the overview does not get lost. This visualization does not allow to rearrange the graph. To help the spatial memory we allow the user to organize the papers and authors however he likes. The authors, Ochoa et al [9], have also studied the complete publication space of a conference, but only with non interactive visualizations, where we allow to do so with a highly interactive visualization.

4.3 Microsoft Academia Search Visual Explorer

The Microsoft Academia Explorer[10] is similar to the Edmedia Relation Browser. Here you can drag the authors around to get a better view if something is not

[8] http://papercube.peterbergstrom.com
[9] http://ariadne.cs.kuleuven.be/edmedia/
[10] http://academic.research.microsoft.com/VisualExplorer.aspx

clear. Once you click on an other author, the graph keeps the link with the previous author but unfortunately all not directly related authors get thrown away. Thus also this visualization only displays direct relations. This application is also only targeted at visualizing authors. One can click through to see all the details of an author, but it is not possible to see the publication which make authors related. Our approach makes the transition from exploration to reading papers easier by bringing the papers visually in the network. If a paper draws attention, one can immediately retrieve more information from it.

5 How to evaluate ?

Due to the early stages of this work, there has not been any evaluation yet, but we are planning to do a complete evaluation and here we outline how we will approach this. The evaluation would be done on two levels :

Macro level We will introduce the test subjects to the application, explain them the purpose, how it works and what are its functionalities. On this level we want to get answers to questions like : Is this application useful ? Does it address an actual need ? And if so, are the people aware of the existing need ?

Micro level In another evaluation, we focus more on the micro level. We want to know if the application is usable, which functionalities and features work well and which do not. In this evaluation the subjects would get specific tasks and we would then record how and how fast these tasks are completed. The specific tasks are not defined yet, but one example could be : Find the most interesting paper written by author x.

Public spaces In order to get more feedback, we also plan to deploy this visualization at one or more conferences, where we can observe the people discovering the tool and see what the initial thoughts are.

6 Future work

At the time of writing, a first working version of the application has been developed with some basic functionalities. But before we can do a real evaluation of this visualization, we need to improve the functionality of the application. In this section we describe the next steps that will be taken to achieve this.

An important feature that is missing at the moment, is being able to search for a certain author or paper to use as a starting point for the visual exploration. At a first stage we will add a keyboard like possibility to enter part of an author or a paper. To show the results there are several options that can be tried out. The found results can be highlighted in some way, or once a single result is found the visualization can center the result and zoom in on it.

At the moment it is not very visually clear yet which papers or authors are the most important or the most relevant. We are already exploring the possibilities to improve this by trying out filtering mechanisms and visual improvements. These visual improvements can be highlighting certain nodes or areas, varying the size of the nodes based on these factors, varying the strengths of the connections, etc.

7 Conclusion

In general, the fundamental issue is to understand in a deeper way how we can support the work of researchers with the technology that is available and how we can evaluate that our efforts make a difference. The design based research presented in this paper tries to move that agenda forward.

A major problem we face is getting clean data. At the moment this is too hard: we had to invest considerable effort in extracting the bibliographical data from the PDF version of the papers and in manually cleaning up the result. Initiatives like DBLP[11], Citeseer[12], bibsonomy[13], citeUlike[14] and others are targeting the same issue and we need to leverage their results in the context of our research.fm framework (see section 3.2) to create sustainable and scalable services for basic bibliographical data provision.

Assisting the user with navigation through the publication space is crucial. It is hard to figure out the correct way to combine navigation and search for manipulation of this information space. Currently, we only provide navigational access and we need to augment this with search facilities to locate relevant locations in this space: these can be papers or authors or relationships between them. We also need to add filtering facilities to reduce the complexity and size of this space to only that part that is relevant to the information need of an author.

We only use a fraction of the available metadata at the moment: our current visualization focuses on (co-)authorship relations between authors and papers. There is plenty of opportunity to also include other kinds of metadata in our scope: this could include forward and backward citations, geospatial information about the affiliations of the authors, textual relationships based on concept extraction techniques, etc. Assessing which kinds of such data help to address which kinds of problems researchers face and how we can exploit the data to make them useful and usable to that audience is a deep design challenge.

Finally, we do not exploit time information yet. However, especially as we start adding more of the metadata to our visualization, this will become an important concern. If we are able to integrate time information, then we can help users understand how a domain or publication outlet (conference, journal,

[11] http://www.informatik.uni-trier.de/ ley/db/
[12] http://citeseer.ist.psu.edu/
[13] http://www.bibsonomy.org/
[14] http://www.citeulike.org/

...) evolves, how a paper gains in influence, how the collaborative relationships between authors evolve, etc.

Acknowledgements We gratefully acknowledge the financial support of the European Commission through the STELLAR project and the STELLAR Network of Excellence.

References

1. P. Bergström and D.C. Atkinson. Augmenting the Exploration of Digital Libraries with Web-Based Visualizations. In *Fourth International Conference on Digital Information Management, 2009. ICDIM 2009*, pages 1–7. Citeseer, 2009.
2. V. Bush and A.W.M. Think. The Atlantic Monthly. *AsWe May Think*, 176(1):101–108, 1945.
3. R.A. Day. How to write and publish scientific papers. *Memórias do Instituto Oswaldo Cruz*, 93, 1998.
4. T. Dwyer, B. Lee, D. Fisher, K.I. Quinn, P. Isenberg, G. Robertson, and C. North. A Comparison of User-Generated and Automatic Graph Layouts. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):961–968, 2009.
5. Clifton Forlines, Daniel Wigdor, Chia Shen, and Ravin Balakrishnan. Direct-touch vs. mouse input for tabletop displays. In *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 647–656, New York, NY, USA, 2007. ACM.
6. Kenrick Kin, Maneesh Agrawala, and Tony DeRose. Determining the benefits of direct-touch, bimanual, and multifinger input on a multitouch workstation. In *GI '09: Proceedings of Graphics Interface 2009*, pages 119–124, Toronto, Ont., Canada, Canada, 2009. Canadian Information Processing Society.
7. J. Klerkx and E. Duval. Visualizing Social Bookmarks. In *conference Workshop on Social Information Retrieval for Technology-Enhanced Learning (SIRTEL07) at the EC-TEL conference. September*, pages 17–20. Citeseer, 2007.
8. J. Klerkx, E. Duval, and M. Meire. Using information visualization for accessing learning object repositories. In *International Conference on Information Visualisation*, volume 8, pages 465–470, 2004.
9. Xavier Ochoa, Gonzalo Mendez, and Erik Duval. Who we are: Analysis of 10 years of the ed-media conference. In *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications ED-Media 2009*, pages 189–200, 2009.
10. Daniel M. Russell, Mark J. Stefik, Peter Pirolli, and Stuart K. Card. The cost structure of sensemaking. In *CHI '93: Proceedings of the INTERACT '93 and CHI '93 conference on Human factors in computing systems*, pages 269–276, New York, NY, USA, 1993. ACM.

Muse: Visualizing the origins and connections of institutions based on co-authorship of publications

Till Nagel[1,2] , Erik Duval[1]

[1] Dept. Computerwetenschappen, Katholieke Universiteit Leuven, Celestijnenlaan 200A, 3001 Heverlee, Belgium
[2] Interaction Design Lab, University of Applied Sciences Potsdam, Pappelallee 8-9, 14469 Potsdam, Germany
nagel@fh-potsdam.de, erik.duval@cs.kuleuven.be

Abstract. This paper introduces Muse, an interactive visualization of publications to explore the collaborations between institutions. For this, the data on co-authorship is utilized, as these signify an existing level of collaboration. The affiliations of authors are geo-located, resulting in relations not only among institutions, but also between regions and countries. We explain our ideas behind the visualization and the interactions, and briefly describe the data processing and the implementation of the working prototype. The prototype focuses on a visualization for large tabletop displays, enabling multiple users to explore their personal networks, as well as emerging patterns in shared networks within a collaborative public setting. For the prototype we used the publication data of the EC-TEL conference.

Keywords: geo-visualization, tabletop, research, human computer interaction

1 Introduction

There has been vast amount of research in the areas of bibliometry and scientrometry to extract and specify the metrics of scientific publication and citation networks. Several works used approaches to visualize these networks (e.g. [1], [2]). In the field of TEL, [3] analyzed and visualized ED-Media publications.

The objective of the presented visualization is not to study individuals and their personal co-authorship networks, but rather to enable analyzing the connection network of universities and research centers. The inter-institutional relationships are based on co-author data, as “co-authorship seems to reflect research collaboration between institutions, regions, and countries in an adequate manner” [4].

Our intention is to focus attention on the spatial relations by creating an easy-to-understand geo-visualization with an emphasis on affiliations and collaborations between these institutions. Studies have shown geographic proximity is important and does positively influence the intensity and frequency of scientific collaboration [5]. However, there has been little research on using geo-visualization for inter-institutional and inter-country collaboration based on publication data (e.g. [6]).

This work focuses on an interactive geo-visualization on large display, enabling multiple users to explore the networks of their affiliations, as well as emerging patterns in shared networks within a collaborative public setting.

We envision several use cases for the application, from which we briefly describe three, exemplarily. (1) A visitor wants to get an overview of the spatial characteristics of scientific collaboration. He starts exploring the institutions and their locations, with the application showing the number of co-authored publications over the years. This visualization supports him understanding whether there is a correlation between proximity and the amount of collaboration. (2) An attendee is interested in finding future partners for writing a proposal. She sees that a colleague from her institution once co-authored a paper with someone from a university department in her field. She writes down the author's name, to later ask her colleague to introduce her. (3) Two persons stand at the table and both are exploring their own affiliations. The application highlights the respective publications, thus enabling them to see shared publications of colleagues, by serendipity. They start talking about these former projects, and find out they have mutual research interests.

The paper introduces *Muse*[1], a working prototype, whose main purpose is to ease the exploration of collaborations between institutions. In addition, the use of a large display tabletop, as well as the aimed-for simplicity of visualization and interaction intend to invite attendees to participate, and engage in discussions at a conference location. The following chapter gives a short overview on the data set. A description of the prototypes' visualizations and interactions follows. The paper closes with short conclusions and comments on future work.

2 Data Set

We are using the EC-TEL dataset as first illustration to show the connectivity in the scientific TEL community. With a young conference as EC-TEL we will not be able to show long-term transformations. Instead, here our aim is twofold: Showing how a striving conference evolved over recent years, and enabling attendees to explore their scientific neighborhoods in the TEL domain.

We harvested the publication data from the website of Springer, the proceedings publisher. We used Web-Harvest [7] to collect all titles, authors, and affiliations including their postal addresses (as well as further data). As the data originally is provided by the authors, using various languages, formats, and accuracies of data, we needed to apply different aggregation and unification heuristics, trying to reduce unintentional duplicates or other skewed data entries. First, the affiliation line is split up into the affiliation's name and its address, to allow a better unification of affiliations, and to display a shorter and more readable name in the visualization. The simplistic, language agnostic approach was to concatenate all text segments up to and including the last segment containing one of a set of specific keywords, selected for high probability of matching institutional name segments (e.g. "universi",

[1] The name of the application was chosen to reflect the meaning of "to look thoughtfully at". Secondarily, Muse, the greek goddess, presides over literature and science.

"a[c|k]adem"). Second, the affiliations were to be unified based on the similarity of the name[2]. After geo-coding the addresses, we also incorporated the spatial proximity to ensure not unifying institutions with very similar names but different locations, e.g. "Dept. of Preventive Medicine, Korea University, South Korea" and "Dept. of Preventive Medicine, Konkuk University, South Korea".

Generally, it is difficult to structure real-world objects in a way to map all possibilities and special cases, thus we utilized a good-enough approach. Before realizing the prototype we probed into the data and looked for patterns to establish the visualization will be able to reflect those inherent relationships. Some of our analysis for the EC-TEL conferences 2006-2009 can be found at [8].

3 Prototype

We designed two working prototypes, with an iterative development approach to refine the visualizations, and to increase the usability of the interactions. The first interactive visualization was presented at the Science2.0 for TEL workshop at EC-TEL 2009. The presentation, and the public display at the venue thereafter allowed us to gather informal responses of attendees. We tried incorporating the given feedback into the second version, and aimed for improving the clarity of the visualization and the overall user experience in an on-location conference setting.

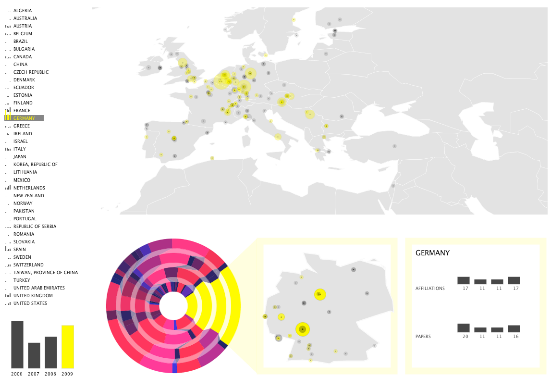

Fig. 1. Screenshot of first prototype with Germany and 2009 as selected country and year.

The first application consists of a static world map showing institutions as colored circles with its overall publication number mapped as size (see Fig. 1). Several further visualizations are in juxtaposition: An overview list shows the names of the countries with contributing authors, with a small *sparkline* [9] signifying the absolute

[2] This simplistic approach results in some false positives (e.g. "Av. Universidad 30"), which are recognized as part of the name, and some false negatives (e.g. "ETH Zürich"), which are regarded as part of the address. Furthermore, some entries could not be unified automatically, such as "Lehrstuhl Informatik V" with "Informatik 5 (Information Systems)".

publications over the years. The concentric rings represent the relative distribution of publications of every participating country over the years, starting from older (inner) up to the latest conference (outer).

These multiple displays are connected, and every user interaction is reflected in all other views. After selecting a country in one of the displays the application provides details-on-demand on that country, and its respective publications and institutions in simple bar diagrams. When the user selects a year the publications are filtered to highlight the data of that specific conference (i.e. as yellow circles and bars).

While the multiple displays allowed looking into the dataset from different perspectives, they also tended to clutter the screen. To effectively communicate the data in a concise visual manner some of the useful but distracting displays have been eliminated in the second prototype. The main improvements were to reduce the visual and interaction complexity by focusing on one main visualization, and the employment of an interactive tabletop with the aim to facilitate multi-user scenarios.

With the large interactive surface, the user not only views and manipulates data on a single user system, but operates in a collaboratively created and used information space (see Fig. 2). In this setting, co-located users, who may or may not be associated with each other, explore the visualization together. Users can arrive or leave at any time, and have the ability to interact as an individual, or as a member of a group with similar interests, goals or attitudes. Cooperative interaction can involve periods of tightly coupled activities by groups with similar but diverging goals, alternated with more loosely coupled individual work. Such collaborative threads can close, split off and merge repeatedly.

Fig. 2. Users exploring institutions with the tabletop prototype.

A single large world map showing all institutions and their relations based on co-authorship are displayed. The user is able to select the region she is interested in by panning and zooming the map (while in the first prototype a user only could switch between World and Europe). Even though more complex map manipulations are possible, we chose this interaction approach, as by reducing the prototype to a single visualization the user can concentrate on the map, thus lessening her efforts. The user can select a country she is interested in. That country is selected, and additional information and diagrams are shown, similar as in the first prototype. These info-windows can be moved to any point on the table. When two countries are selected the prototype displays the diagrams besides each other, allowing the user to compare them.

4 Conclusion and Future Work

Although we have utilized only a small dataset with a rather small significance for general scientific network analysis, we see the Muse prototype with the used data set as beneficial case study. Through interactive filtering the user is able to explore the temporal as well as spatial relations between institutions, and can gather insights into the conference. The collaborative usage of the interactive tabletop display fosters communication among participants.

We intend to broaden the data set to other conferences. Currently, we see two possibilities: Besides using the harvesting tool to scrape further publications from Springer and other official sources, we plan to integrate publication data services, such as pub.fm [10]. Second, querying Web2.0 applications such as Mendeley [11] to gather social network data of the authors.

Furthermore, we are planning an evaluation on intelligibility of the visualization, and usability of the interactions. As direct response from users in a real-world setting can be worthwhile, we intend to create a brief questionnaire to gather feedback from attendees at the EC-TEL 2010.

References

1. Ponds, R., Van Oort, F. G., and Frenken, K. The geographical and institutional proximity of research collaboration. Papers in Regional Science, 86(3), 423–443. (2007)
2. Henry, N., Goodell, H., Elmqvist, N., Fekete, J.–D.: 20 Years of Four HCI Conferences: A Visual Exploration. International Journal of Human-Computer Interaction, 23:3, 239-285. (2007)
3. Ochoa, X., Méndez, G., Duval, E. Who We Are: Analysis of 10 Years of the ED-MEDIA Conference. In: G. Siemens and C. Fulford (eds.), Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2009, 189-200. Chesapeake, VA: AACE. (2009)
4. Glänzel, W., Schubert, A.: Analyzing Scientific Networks through Co-authorship. In: H.F. Moed et al. (eds.), Handbook of Quantitative Science and Technology Research, 257-276. Kluwer, Dordrecht (2004)
5. Katz, J. S. Geographical Proximity and Scientific Collaboration. Scientometrics 31 (1), 31–43. (1994)
6. Cyberinfrastructure for Network Science Center. Research Collaborations by the Chinese Academy of Sciences. http://scimaps.org/maps/map/research_collaborati_110/ (2009)
7. Web-Harvest Project. http://web-harvest.sourceforge.net/
8. Nagel, T. http://tillnagel.com/2010/09/ectel/
9. Tufte, E. Beautiful Evidence. Graphics Press. (2006)
10. Stellar project. http://www.stellarnet.eu/d/6/3/KULDocumentation
11. Mendeley. http://www.mendeley.com/oapi/

Tools to Find Connections Between Researchers – Findings from Preliminary Work with a Prototype as Part of a University Virtual Research Environment

Jim Hensman[1], Dimoklis Despotakis[2], Ajdin Brandic[1], Vania Dimitrova[2]

[1] Coventry University, UK
{j.hensman, a.brandic}@coventry.ac.uk
[2] University of Leeds, UK
vania@comp.leeds.ac.uk, scdd@leeds.ac.uk

Abstract: This paper describes development work in progress on tools to identify connections between researchers, as well as between researchers and business and other wider partners. The work is being carried out as part of a project, the Building Research and Innovation Networks (Brain) project, based at Coventry University in the UK with Leeds University as partners, and is part of a JISC funded Virtual Research Environment Programme. The Brain project aims to facilitate the building of Communities of Practice and networks of researchers and business and community partners to help enable the collective intelligence that potentially exists if these participants could be suitably engaged. In this endeavour, the project has explicitly identified the Research 2.0 approach as being central. Within the scope of this paper, only certain aspects of the project will be considered in any depth. The wider project includes work on business and knowledge related processes which impact on the nature and validity of the data used by tools such as those described here. Also central to the project is the building of Communities of Practice of researchers and other partners and the development of physical and virtual networks to support these. The development work on the tools described here was carried out by Dimoklis Despotakis and Ajdin Brandic.

Keywords: virtual research environment, brain project, research 2.0

1 Introduction

The tools discussed in this paper are being developed in response to ongoing user requirements identified by the project as well as conforming closely to the identified strategic institutional need to facilitate collaborative research focused around 8 themes. The techniques used can be considered part of those concerned with finding commonality between items and the tools discussed provide two main functions: Searching for researchers by keywords related to their work, and finding links from a specified researcher to others. The key system components are the user input interface, a means of expanding keywords - using synonyms for example, the search mechanism, a means of filtering/weighting of results and the user output interface including suitable visualisation of information. The person linking tool adds an additional component which generates appropriate search keywords for an individual which can then be processed using the search functionality.

An example output for a person connection search is shown, illustrating the identification of expected close connections as well as ones from other disciplines. Also shown is an illustration of the use of the tools to create a map of linked topics and individuals around one of the broad strategic themes. Early evaluation of results shows a favourable reaction from users and considerable promise. Important future work seeks to extend the scope of coverage to wider research and business information, to using additional techniques including more adaptable and semantic methods and to looking in more depth at the underlying principles behind establishing connections, including applying pattern language-based approaches.

2 Requirements and Use Cases

The need for internet based services and tools to support communities of researchers has been generally recognised. Several national and international co-ordinating organisations and projects, such as JISC in the UK, Surfnet in the Netherlands and the EU Stellar Network have identified generic facilities and services which can facilitate collaborative research and complement discipline specific applications. Extensive user requirement analysis with researchers and other stakeholders at Coventry University has confirmed the need for certain functionality that correlates with requirements identified more widely. One such set of requirements relates to tools to support researchers finding potential collaborators or links to potential partners in business and the community. This arises in various forms in different stages of the research process. For example, at the inception stage of a possible piece of research, a typical need was expressed by a user as, “How do you find the people to talk to about an idea?”

At a later stage, when more detailed formulation of a research proposal or the writing of a paper is taking place, specific expertise, that could be outside the discipline of the main researcher or set of researchers, could be needed – perhaps in the area of data analysis or project evaluation. This would especially arise in cases of multi-disciplinary or inter-disciplinary research and in work that combined academic research with external activities in business or the community. A particular use case analysed by the Brain project illustrates the potential complexity of creating a suitable research team. This was a research call funded jointly by the Science and Social Science Research Councils in the UK on the theme of “Energy and Communities”1. This call involved subject areas ranging from environmental science, civil engineering and computer simulation through to psychology, sociology, economics and politics. A

[1] http://www.esrcsocietytoday.ac.uk/ESRCInfoCentre/Images/Energy_and_Communities_Call_Specification_tcm6-34922.pdf

particular set of use cases has acted a central driver for the project and arose from the need to create collaborative networks to take forward work on 8 cross-disciplinary themes prioritised by the University as “Grand Challenges”. A similar, but less clearly defined requirement arises when trying to identify groupings or clusters of researchers that may have the potential of working together or where the objective is to identify sub-disciplines within a larger area, but where the connecting themes are not known in advance. Examples of this which the Brain project has been engaged with concern finding connections between specific research groups and wider groupings of researchers for the purpose of the (UK) Research Evaluation Framework exercise that requires this for funding allocation purposes.

The basic methodology of the Brain project is to identify requirements, construct a structured model of the processes and services that could fulfill these requirements and then develop a prototype integrated environment based on this. These three parts of the work are closely intertwined and the Brain project adopts an Agile programming development methodology with short iterative cycles of development closely integrated with user requirements gathering, testing and evaluation. This paper describes some of the initial work carried out in the area of developing tools to identify connections, which forms part of the wider system to support collaborative research and innovation including discussion, networking and other services.

3 Techniques and Functional Components

It is not possible in this brief paper even to begin to indicate the considerable volume of research relevant to this area of work and this cursory introduction will only mention a few examples of work to set the wider context. Analysis of scientific and research networks and the connections between researchers that constitute them can reveal important characteristics and trends, such as the well known “six degrees of separation” property, described in one piece of research [1] as, “collaboration networks form “small worlds,” in which randomly chosen pairs of scientists are typically separated by only a short path of intermediate acquaintances.” Important examples of this work include an analysis of the Edmedia conferences [2] and an analysis of TEL Research Communities [3]. An extensive amount of software exists in the general field of social network analysis and is documented by organisations such as the International Network for Social Network Analysis[2].

The requirement considered here is about finding connections for the purposes described earlier and thus has a specific focus in comparison to the field in general. Numerous systems for finding experts exist, ranging from systems within individual organisations or particular membership networks, to those that aim to cover the web as a whole. A comparative evaluation of a number of these systems is made by Becerra-Fernandez [4]. Although a diverse variety of complex techniques are used by systems of this kind, it is possible to identify an underlying core set of functional elements used to implement them. One key generic component is to be able to

[2] http://www.insna.org/

identify what could be termed commonality – which could be between search terms and a document, between different researchers or researchers and businesses, and so on. This could be based on explicit or implicit characteristics. A search term has commonality with a document that contains it within the text – an explicit indication. Two researchers have commonality if they have read, cited or co-authored a particular paper – an implicit indication that arises from an aspect of their behaviour and which forms a central part of the analysis of the work mentioned above. Some quantifiable relative measure may be associated with this commonality. For example, two researchers who have referenced a paper would generally be considered to have a higher degree of commonality to two who have read the same paper, and two researchers who have co-authored a paper would be considered to have a still higher degree of commonality. A further simple metric may be the number of matches – the number of times a search term occurs in a document, the number of commonly referenced papers etc. This may need adjustment or normalisation in some form however – so that long and short documents or someone who has written a few papers can be compared with someone who has written many, for instance. In a more general sense this can include other features of adaptability that adjust the results to the characteristics of the data or the context. In some cases quantification of commonality or other analysis could be used to exclude certain results or weight them in some form that could be used in the visualisation – for instance grouping more strongly related items closer together.

More complex techniques that take into account indirect and secondary effects can sometimes be crucial to the success of this approach. A relatively simple and then more complex example will illustrate this. If we are searching for a keyword in a set of documents, we may wish to include synonyms or apply stemming techniques so that related terms are also searched for. This can be extended to include applying the concept of semantic distance [5], so terms which are more similar according to some criterion based on their place in some subject taxonomy, for instance, have a greater weighting. Perhaps the best known example of a more complex technique is the PageRank algorithm [6] used by Google and in various forms by other web search engines. The objective here is to quantify and thus rank the importance of a web page that contains a search term. The number of links to a page is used as the metric for this, but weighted by the number of links to those pages in turn and so on. Another well known technique will illustrate a different important aspect to using techniques of this kind. Recommendation engines used by businesses like Amazon base themselves on the commonality of customers reflected in the past purchases they have made to suggest new ones. An easily quantifiable success metric is available to the business in this case – what proportion of what customers actually buy are recommended items. Having a metric of some form like this is important to evaluate the success of techniques used and help choose and improve them. In the case of the research related examples considered, this will usually be more difficult and directly measurable metrics often not available. Nevertheless, having processes available to serve the same purpose - through user feedback and interaction for example – are still important, and incorporating these into the overall design is necessary to improve and evolve the systems implemented.

Although techniques such as these mentioned provide a powerful set of methods and a number are part of current project development activities, this paper focuses on some of the initial development work which concentrated on the rapid creation of functional prototypes that could be deployed with users to meet real requirements. Although the techniques used for these were simple, they nevertheless provided usable functionality and allowed engagement of the project with researchers and others – a key priority at this stage, although improving and optimising the implementation is being carried out as a parallel process.

4 Implementation

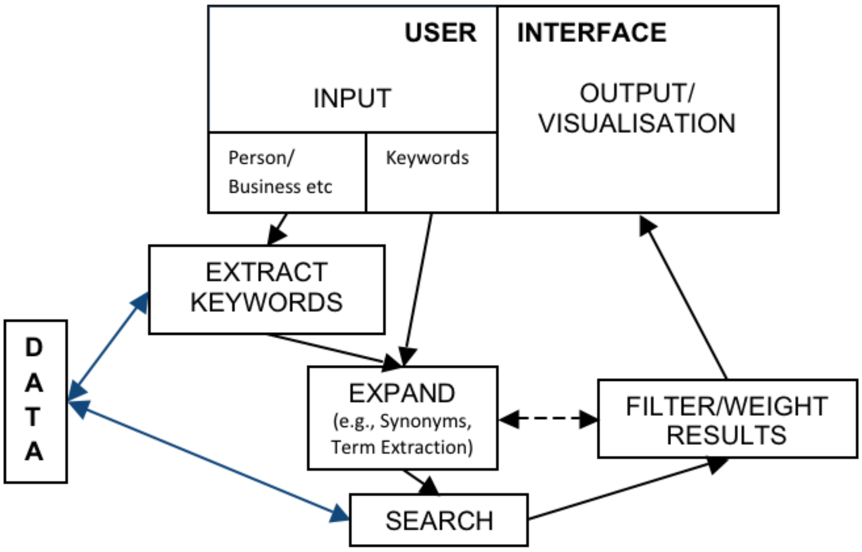

Fig. 1: Functional System Components

The diagram shows in outline form the basic functional components of the tool which incorporates in a simple form the key elements outlined in the previous section. For the initial prototype, these were implemented as follows.

4.1 User Input

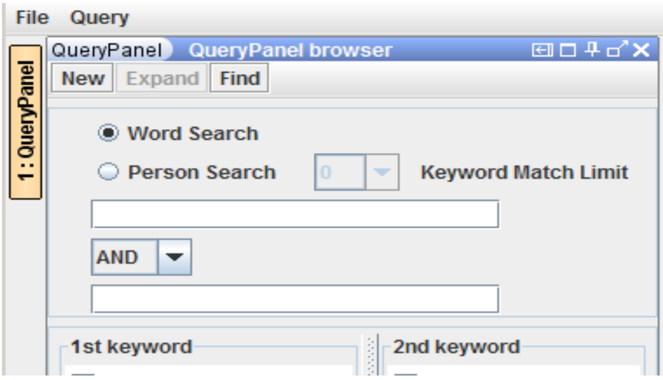

Fig. 2: User Input Interface

Shown above is the user input panel. Two facilities are provided in the prototype, searching by keyword or topic and grouping by person. For the keyword search, simple Boolean combination of terms is available.

4.2 Data

A central part of the wider Brain project was looking at the data relevant to research and the processes associated with it. In the general case, even for a particular requirement such as finding links between researchers, a very wide variety of data could be used in various ways. The project is interested in connections within the academic community as a whole as well as with wider business and community engagement. For the initial work however, data was restricted to the University, to provide a more limited scope for the requirement that could be evaluated more rigorously and then generalised appropriately. How this is being extended wider in developments currently taking place will be mentioned later. Data from a variety of sources has been used, providing information about researchers' expertise, interests, publications, projects etc. Linking information from these different data sets was done for the first time at the University by the project and proved a very considerable challenge. Some information was not available in an online form previously and special work had to be carried out to clean up and to link data where appropriate key fields did not exist. However, carrying out these tasks allowed valuable knowledge about research to be available for the first time, irrespective of the techniques used to process and analyse the information. The part of the project not detailed here relating to process has concentrated on how appropriate information can be made available in an up-to-date and reliable manner.

4.3 Commonality analysis

A brief indication of the range of functionality that could be used in this area was discussed earlier. Although some of these techniques are now being used in the ongoing development, the simpler techniques used in the initial prototype will be discussed here. The keyword search facility implemented was based on a simple string matching in the available data with the search words and selected synonyms, implementing simple Boolean combinations of these appropriately. Synonyms for the terms entered were generated using WordNet[3] and Disco[4] facilities and a checkbox facility provided for the user to choose these as desired.

For the person matching facility, the aim was to re-use a number of the components used in the keyword search. The keyword search process finds individuals whose associated information matches the keywords entered. Therefore, if appropriate keywords can be associated with an individual, an aggregation of the results from these as separate keyword searches can be used to determine the required person links. This raised the problem of how to generate these keywords. Where an explicit list of expertise areas was available in the data for an individual, for example, applying this approach would be trivial. However, applying this to other information was not as easy. Using the title of an academic paper as a keyword, for instance, would usually be very specific and therefore only usually match another academic if they were co-authors of the paper.

The synonym facility provided to expand the keyword search was not appropriate in this case and a different technique was used to implement this facility to generate keywords from sources of information such as the titles of papers. Among other techniques, two in particular using available web services were tried as part of this, the Yahoo Term Extraction service5 and the OpenCalais6 semantic metadata service . The Yahoo service proved to be more appropriate for use with publication titles especially and is the one used for the first prototype development, although the OpenCalais service is also being used in the system being developed.

Filtering/weighting results was looked at earlier as one of the components in determining commonality. In the early prototype system described here, adequate functionality for the keyword search could be provided without having to consider this area. However, for the person search this was an important consideration. In developing any system of this kind a balance has to be maintained between completeness and usability. When finding matches between people, a certain number of false positives can be expected. However, if these are too large as a proportion of results returned, the system will not be usable. Two techniques were used to tackle this problem. The first was the use of a stop list which filtered out certain words or phrases which were adjudged not to be useful in establishing connections, and was used after the stage of keyword expansion. For example, words like "research" and "university", are obviously too general to be of use. Considerable user testing and feedback was required in refining this stop list to limit matches to relevant ones, and

[3] http://wordnet.princeton.edu/

[4] http://www.linguatools.de/disco/disco_en.html

[5] http://developer.yahoo.com/search/content/V1/termExtraction.html

[6] http://www.opencalais.com/

the current list has over 1200 terms. The second technique used was to provide a user selectable filter parameter which would exclude terms which generated over a specified number of person matches. This allows searches to be run and then this parameter adjusted depending on the results.

4.4 Output and Visualisation

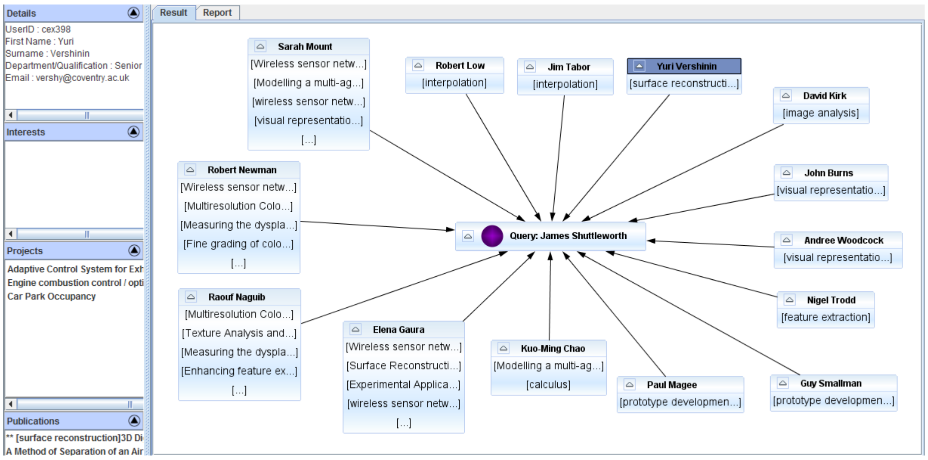

Fig. 3: Example Output

The output from a typical person connection search is shown above. Matched individuals have the items which were responsible for the connection displayed and highlighting an individual allows more detailed information about the match as well as other information about them to be shown in the side window. Individual matched items can be moved and hidden easily allowing particular features to be focused on if required. Multiple searches can be run and then tabbed between to allow different results to be compared and combined as necessary. A separate report view is also available which provides more detail about all the matches and which also provides the output in formats that can be exported into other applications for analysis and visualisation.

The example shown demonstrates one of the aspects of the system which allows new relationships and potential collaborations to be facilitated. In the illustration above, the researchers displayed on the left with a number of matches shown are members of research groups that the selected researcher, James, is part of. Thus their work (in the areas of Wireless Sensor Networks and Computer Analysis of Medical Images) can be expected to be already known to him. However, the tool has also picked up a variety of other researchers and associated research areas that in some cases are quite unexpected but nevertheless possibly relevant. These include mathematicians through the analytical techniques used by James, specialists in visual representation through the visualisation techniques he has used, and specialists in different types of image analysis from disciplines as diverse as automotive engineering, shot peening, and Geographical Information Systems.

5 Use, Evaluation and further Development

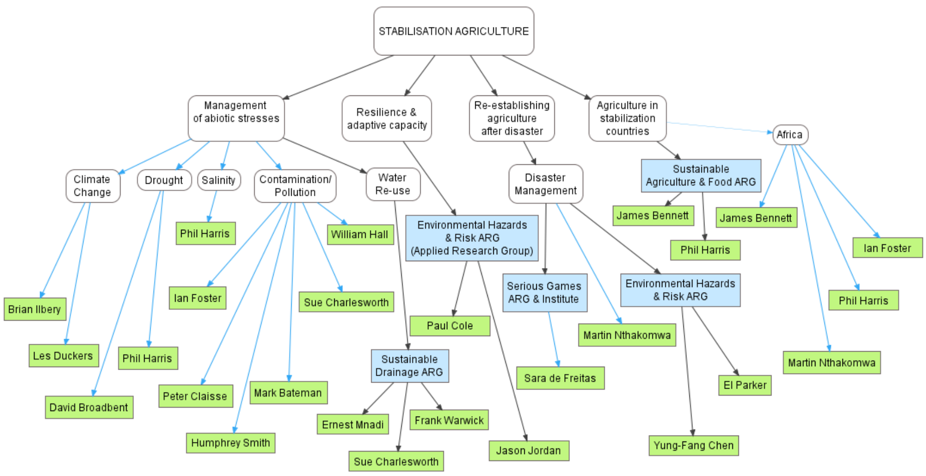

Fig. 3: Example Theme Mapping

As mentioned earlier, a key aim of the project was to be involved in fulfilling real requirements and solving real problems. Even in its relatively early stages the project and the tools it has created have had the opportunity to be embedded in key strategic initiatives and be tried out in practice. A significant amount of feedback and evaluation has been obtained working with individual researchers, research groups and research support staff, which has allowed significant iterative modification and improvement to be implemented, as well as further requirements that are currently being implemented to be identified. Space precludes detailed discussion of the many ways the tools developed have been used, but one example will be shown here. The diagram above shows a small section of one of the many visualisations of disciplinary and researcher links which the project working with appropriate researchers has generated, in this case for one of the "Grand Challenges" mentioned earlier - Sustainable Agriculture. Using both the keyword search and person link facilities together iteratively and the export facility mentioned earlier, powerful visualisations, in this case using the Visual Understanding Environment (VUE) application[7], can be constructed relatively easily. Used together with some of the other facilities that the project has made available, in the social networking field for instance, this provides a very significant capability to assist and facilitate collaborative research and innovation.

[7] http://vue.tufts.edu/

Key lessons learned from user engagement and feedback are summarised below, together with intended developments based on them and on the other aims of the project:

- Formal evaluation of the tools is mainly part of the forthcoming work of the project, but a co-evolutionary methodology utilising detailed feedback from users around specific use cases is the basic approach adopted. User response to the early prototypes has been very favourable in general. The system demonstrated its value from the first time it was used in practice by finding researchers for a particular initiative who were working on a common topic in different faculties but unaware of each other, and this has been repeated a number of times. As referred to earlier, this was partly a consequence of linking together information which had never been linked before as well as the expected and sometimes unexpected aspects of how the tools operate. In comparing previous attempts to manually carry out some of the tasks which the system has been used for it is also apparent that even semi-automated methods save a huge amount of time and make previously impossible analyses relatively trivial. The exercise has also helped to demonstrate the value of a more knowledge-based approach to university information and has fed directly into institutional strategic policy.
- Easy access to the tools and availability of current versions of software and up-to-date data are seen as a necessity, which in practice means implementing the tools as web-based applications. Currently the tools are implemented as a stand-alone PC application, mainly because the synonym generation facility used is only available in this form. Subsidiary web services for facilities like this will need to be developed if necessary.
- Extending coverage to include external information and being able to establish connections with researchers and others generally was both requested and a key aim of the project. This would require generalising how data is accessed and an implementation of the system which uses more general structured search, possibly implemented using Solr/Lucene, is part of the current development. Integrating information in RDF form together with the use of semantic search techniques is an intended further development. Because of the key aspect of the project relating to innovation as well as research, currently also being developed are ways to integrate business and other sources of information, using tools like OpenCalais and screen scraping and mashup tools as necessary. Considering commonality analysis in its more general sense could include facilities to recommend suitable papers to researchers, associate expertise and potential projects with funding etc. Because these requirements are linked, tools and services to deal with one can be used for others and the underlying knowledge set can be common, leading to the potential for a very powerful integrated environment.
- More powerful functionality to allow co-authorship and co-citations etc to be taken into account explicitly was seen as important and including a number of search, clustering and classification algorithms relevant to different contexts and types of information, is also necessary.
- Improvements in the visualisation algorithms and associated commonality techniques, for instance to reflect the strength of a connection by closeness, was a common request, as was the ability to manipulate and aggregate maps more easily, so

that multiple maps could be combined and connections linking to other connections generated automatically.

- Many improvements to the overall user interface and underlying functionality were suggested. Users often compared the tool to services like Google they were familiar with, requesting more flexible searching etc.
- Improvements to a number of auxiliary services used, such as the synonym mechanism, were requested. The current systems used, which are for a general audience, were considered too informal by some users. Work is being done to include more technical sources, thesauri and ontologies, such as the UKAT system8. Using systems of this kind allow more powerful commonality associations to be implemented, which have been found to be important to find less obvious connections – using measures of semantic distance for example.
- A considerable amount of feedback has been about the importance of including informal and tacit knowledge. Again, in considering “commonality” and how research topics and researchers link to each other, a number of assumptions have been made, such as that the closeness of match is the only criterion to be used. More sophisticated approaches are needed for understanding and representing information to take into account complementarity of knowledge and other considerations. A key part of the theoretical basis for the current project derives from earlier work carried out by members of the project team, in particular the Planet project9, which looked at how practice could be shared and represented – taking the use of Web 2.0 techniques in learning as an example, and the Connection project10 linked to this which looked at how connections between projects could be facilitated, particularly carrying this out for the set of projects that were part of the JISC Users and Innovation Programme (Emerge). A number of principles and techniques came out of this work, particularly involving the use of pattern language based approaches. The Brain project is seeking to further develop and extend some of these which are especially relevant to the tools discussed in this paper.

Acknowledgements

The authors wish to acknowledge the contribution of other members of the project team, Peter Haine, Derek Griffiths, Stella Kleanthous and John Tutchings. The contribution of the JISC in funding this work is also acknowledged.

[8] http://www.ukat.org.uk

[9] http://www.jisc.ac.uk/media/documents/programmes/usersandinnovation/planet%20final%20report.pdf

[10] http://cublogs.coventry.ac.uk/innovation/files/2010/08/jisc-connection-final-report.pdf

References

1. M.E.J.Newman (2001), The structure of scientific collaboration networks, *Proceedings of the National Academy of Sciences*, January 16, 2001 vol. 98 no. 2 404-409.
2. Ochoa, X., Mendez, G., Duval, E. (2009). Who we are: Analysis of 10 years of the ED-MEDIA Conference, ED-MEDIA 2009.
3. Marco Fisichella, Eelco Herder, Ivana Marenzi, Wolfgang Nejdl, (2010), Who are you working with? Visualizing TEL Research Communities, Retrieved on 5/7/2010 from: http://www.l3s.de/~herder/research/papers/2010/who_are_you_working_with.pdf
4. Irma Becerra-Fernandez, Searching for Experts on the Web: A Review of Contemporary Expertise Locator Systems, ACM Transactions on Internet Technology, Vol. 6, No. 4, November 2006, Pages 333–355.
5. Sowa, John F., (2000), Knowledge Representation, Logical, Philosophical and Computational Foundations.
6. Sergey Brin, Larry Page (1998), The Anatomy of a Large-Scale Hypertextual Web Search Engine, Proceedings of the 7th international conference on World Wide Web (WWW), Brisbane, Australia.

The Afterlife of ‘Living Deliverables’: Angels or Zombies?

Fridolin Wild, Thomas Ullmann

Knowledge Media Institute, The Open University, UK
{f.wild, t.ullmann}@open.ac.uk

Abstract: Within the STELLAR project, we provide the possibility to use living documents for the collaborative writing work on deliverables. Compared to ‘normal‘ deliverables, ‘living’ deliverables come into existence much earlier than their delivery deadline and are expected to ‘live on’ after their official delivery to the European Commission. They are expected to foster collaboration. Within this contribution we investigate, how these deliverables have been used over the first 16 months of the project. We therefore propose a set of new analysis methods facilitating social network analysis on publicly available revision history data. With this instrumentarium, we critically look at whether the living deliverables have been successfully used for collaboration and whether their ‘afterlife’ beyond the contractual deadline had turned them into ‘zombies’ (still visible, but no or little live editing activities). The results show that the observed deliverables show signs of life, but often in connection with a topical change and in conjunction with changes in the pattern of collaboration.

Keywords: deliverables, wiki, collaboration, analysis, visualisation, #stellarnet

1 Introduction

In standard project management jargon, a ‘deliverable’ refers to a pre-defined, tangible, and verifiable work product such as a feasibility study or a prototype [1]. In research projects, deliverables often document process and outcomes of (more or less) systematic knowledge creation. They report on the progress against the tasks expected to be ‘delivered’ during a defined phase of the project. These documents sum up the focused work of a group or single person.

Within the STELLAR project, we provide the possibility to use living documents for the collaborative writing work on deliverables. They can be continuously updated and revised by all authors, even in parallel, using the popular wiki software MediaWiki (the software on which Wikipedia is based). Compared to ‘normal‘ deliverables, ‘living’ deliverables come into existence much earlier than their delivery deadline and are expected to ‘live on’ after their official delivery to the European Commission. They are expected to foster collaboration in writing. Within this contribution we investigate, how these deliverables have been used over the first 16 months of the project. We will critically look at whether they have been successfully used for collaboration and whether their 'afterlife' beyond the contractual deadline had turned them into a 'zombie' (arguably still some sort of life, but not a really welcome one). A zombie can still be seen, but does not show any signs of vital activity, whereas an angel cheerfully continues editing activities – but with the difference of being relieved from the duty of the mortal to deliver. It is clear that deadlines are typically drivers of activity, so also for angels, afterlife activity should be visibly less hectic and might focus on new or different areas of editing activity.

The analysis of the dynamics of wikis and their flagship Wikipedia is naturally a relatively young research field, since Wikipedia was created only back in 2001 – thereby making available a large public data-set of revision histories. Viegas et al. propose a method called 'history flows' for analysing the social dynamics expressed in the editing of Wikipedia articles [4]. They analyse the relationship between document revisions revealing cooperation and conflict patterns. Nunes et al. [3] use the revision history to visualize revision activity through sparklines in a timeline plot within their system 'WikiChanges', additionally supported by a 'tag-cloud'-like visualisation of term changes in the time frame selected (the font size is scaled by their changed frequency within the time window inspected). Arazy et al. [2] develop a series of glyphs to visualise contribution scores of authors in pages in order to ease the recognition of their work. Suh et al. [5] focus on identifying patterns of conflict with the help of so-called 'revert graphs', visualising the relation between authors of Wikipedia established through revisions that void previous edits. Baumgrass et al. [6] apply social network analysis in order to investigate corporate knowledge exchange processes in wikis. Closely related is also the work of Jesus et al. [7], within which network analysis is applied to study cluster-level collaboration between authors grouped by their work on related articles. Whereas [2,3,4,5] focus on the analysis of collaboration in individual pages, [6] and [7] deploy the same analytical technique – (social) network analysis –, but with a different focus of analysis [7] and in a different cultural and application setting [6].

All of them, however, share with our work the interest to shed light on the authorship relations documented in the revision histories. The user interface of the wikis is designed in a way, which centres the article and not so much the contributions of the single authors: its focus is on content and not authorship [2]. Making the authorship relation visible means extracting the relevant data from the revision histories of the pages and providing an easy to understand view of this data.

While a deliverable is the result of the edits of all authors, the revision history retains information about the contribution of each individual. This makes it easy to spot latest edits or compare changes with previous ones. It helps to keep track of the development of the pages contained in the living deliverable and, for example, make it easy to revert edits.

There are many ways of how to represent writing activity and collaboration of wiki pages. Within the rest of this paper, we first elaborate on our method of analysis used to make the collaborative writing process of living deliverables visible. With this, we analyse the data gathered within the STELLAR project so far: we visualize the overall co-authorship network; we outline the revision frequency over time to investigate if the living deliverables are indeed living; and we show how the collaboration network of authors and their contributions changes before and after a deadline. Finally, we conclude the paper with a summary and an outlook.

3 The data: Stellar's 'living deliverables'

The observed dataset consists of five living deliverables. They have been selected from the set of 14 wikis created so far for 19 project deliverables by excluding 'obvious zombies' and 'small group wikis' such as the coordination manual. Obvious zombies thereby relate to those wikis for which the group of collaborators did not use the offered wiki or abandoned it early in the writing process favouring different solutions to organise collaborative writing: these were mainly google docs and in several cases the exchange of word and excel files via mail with one or several editors consolidating tracked changes. The latter thereby being the main method used for the five management and evaluation deliverables that are much more clerical in nature and contain a lot of spreadsheet data – a task for which MediaWikis are hard to use.

Each living deliverable resides in its own MediaWiki instance. All wikis were initialized at the beginning of each deliverable writing period. While observing the process of the living deliverable evolution, we have to consider the fact that these documents served as input for the 'normal' deliverables (the type-set word or PDF file delivered to the European Commission), and the latter could then again feed back into the living deliverables.

The following Table 1 gives an overview of each of the investigated living deliverables. Among others, it outlines the number of authors, the number of pages contained in the wiki (and their number of page views), and – most notably – the number of edits these pages have received. All in all, the deliverables had an average number of 22.7 users, with a varying number of page views (in average 3,820). Some of them have received a substantial number of edits (such as the grand challenge document d1.1 and the science 2.0 mash-up deliverable d6.3, both earlier deliverables).

	Users	Total Views	Total Pages	Total Edits	Total Images	Pages/ Users	Edits/ Users
d1.1	78	14813	78	533	4	1	6.83
d1.2	9	1338	86	137	1	9.56	15.22
d6.1	4	677	39	152	28	9.75	38
d6.2	11	712	14	79	10	1.27	7.18
d6.3	21	2818	65	333	1	3.1	15.86
d7.1	13	2563	84	354	48	6.46	27.23

Table 1. Basic statistics of the investigated wikis.

4 Method of analysis: SNA of the collaboration networks

The revision history of the living deliverables is a chronologically sorted list of changes of pages, listing – amongst others – the editing user, the page, the amount of characters changed with the revision, and a timestamp expressing when the revision was applied. One example of this revision history can be found in the snapshot of a revision history visualisation widget we have created to support the work in the deliverables (Figure 1): it shows the revision of one living deliverable in a scrollable timeline, listing the title of the changed page, the date of the change, and the name of the editor (pop-up bubble).

While this way of exploring the revision data has its benefit for following latest changes or browsing through the history of all changes, it does not provide much insight into the nature and vitality of the underlying collaboration, nor much insight into the focus of collaboration.

Collaboration is expressed in the co-authorship relations and can be extracted from the revision history. Co-authorship relations in living deliverables, however, can be investigated in many ways. The simplest form would be a list of authors of the deliverable or a page in it. List-like representations, however, do not show the structure of collaboration between the authors of the living deliverable. This extra dimension of information can provide insights into the collaboration network structure. We used a co-authorship social network analysis, which shows the relations established between authors by editing the same page. Therefore, an incident matrix was constructed listing the pages as incidents in the rows, the authors in the columns, and their number of edits of the respective page in the matrix cells. By multiplying the matrix with its transpose, an undirected affiliation matrix can be constructed and visualised as a network (see **Figure 2**).

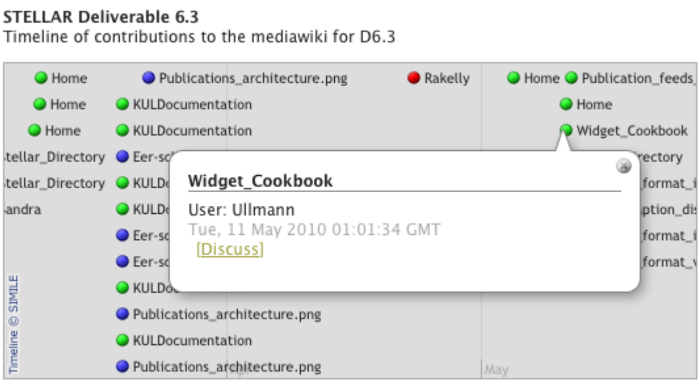

Figure 1. Timeline widget (visualizing the revision history of D6.3).

Since the central jump page (‘home’) of wikis is edited very often and by almost everyone (to, e.g., add links to new sub pages), it may be excluded from analysis in order to expose the clusters of collaborating authors more clearly (see Figure 3).

Figure 2. Collaboration network including edits of the central home page (D6.3).

The graph shows, a cluster of authors who contribute to a shared article. On the periphery of the cluster, the less connected authors are shown. By removing the central home page, two clusters can be seen, which are connected only through shared contributions of two authors. On the periphery there are four authors, who only wrote contributions to the main page or only on pages not edited by others, but not on any of the pages co-edited by the authors in the two clusters.

Figure 3. Collaboration network excluding the central home page.

This co-authorship visualisation has its benefit in showing who collaborated with whom. It does not, however, show the evolution of the living deliverable over time and it lacks information about the content on which the authors collaborated. This can be extended by adding pages as nodes to the network and introducing directed editing relationships pointing from the authors to the pages they have changed. With that, authoring relations on particular pages become more salient.

Additionally, the development of the overall number of non-minor edits over time provides information on the vitality of the wiki and complements the analysis.

5 Discussion: Is there an afterlife after the deadline?

The deadline of regular deliverables marks the end of the writing process. After the deadline, the official writing process ends and there is no formal requirement to modify them anymore. As mentioned above, the purpose of living deliverables is to allow for more continuous collaboration beyond delivery deadlines. The assumption behind living documents is that knowledge construction processes are continuous and deliverables are artefacts of an underlying, continuous collaboration process. By turning these artefacts into living documents, they better reflect the dynamic structure of project work, which is somewhat artificially subjected to a project framework in order to allow for efficient and effective management. Not only in networks of excellence, where a consortium faces additionally the challenge to re-organise an open research network beyond the partnership, but also in other research project types, interdependencies of tasks naturally create feedback loops that should inform already ‘delivered’ work (such as from validation to conceptual design), thus creating an opportunity to update them.

To test whether or not the documents were subject to editing activity also after the submission deadline, we gathered the revisions of each deliverable and cumulated the amount of revisions for each deliverable for each project month. The following line chart shows on the y-axis the amount of revisions and on the x-axis the time frames (16 project months). One deliverable already exists since 13 months, while others are in use for shorter periods of time. The vertical lines at month 3, 6, 9, and 12 represent the submission deadlines.

All deliverables continue their life also after their formal deadline. Even when considering a phase of two months after the deadlines (taking into account possible delays in delivery), still three of the deliverables show lively activity. According to the revision counts, the official deadline raised the number of revisions, while after a deadline the amount of revisions increases mostly less steep. The three deliverables d6.2 (blue), d6.3 (purple), and d1.2 (yellow) show a very steady increase over time, whereas particularly the early deliverables d7.1 (orange) and d1.1 (green) experience their most busy editing processes around the time of their deadline.

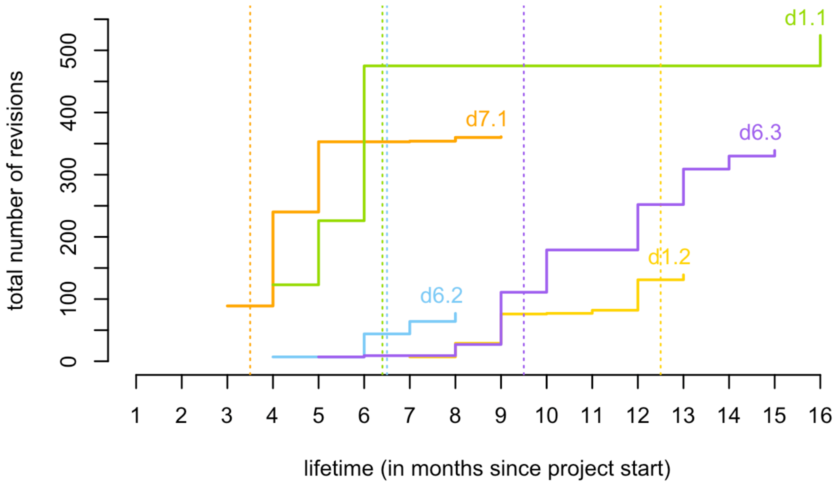

Figure 4. Total number of edits (cumulated) for each living deliverable.

While the line chart visualisation only shows the frequencies of the revisions over time, it does not provide information about the themes of collaboration and the collaboration network created in the co-editing activity – and how they have changed from before to after the deadline.

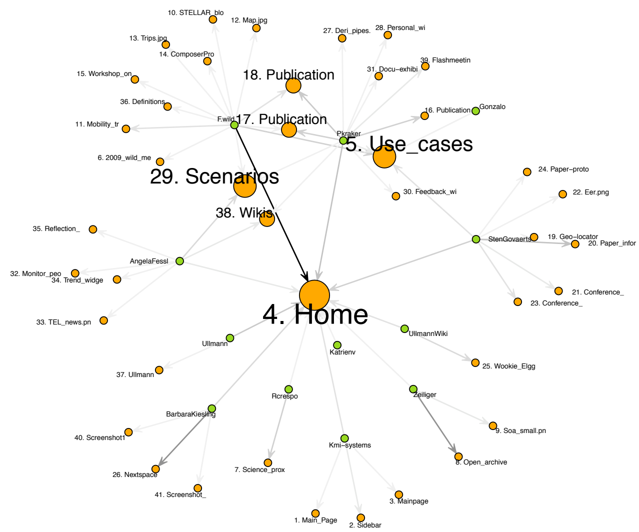

Figure 5. Authors (green) and their contributions to pages (orange): before the submission deadline.

Figures Figure 5 and Figure 6 show the network of authors and their contributions to pages in d6.3 before and after the submission deadline. While the focus before the deadline is clearly on 'use cases', 'scenarios' and the main page of the deliverable, the figure for the network after the deadline shows a change towards more technical topics, like 'Tools', 'Services', and 'Widgets'.

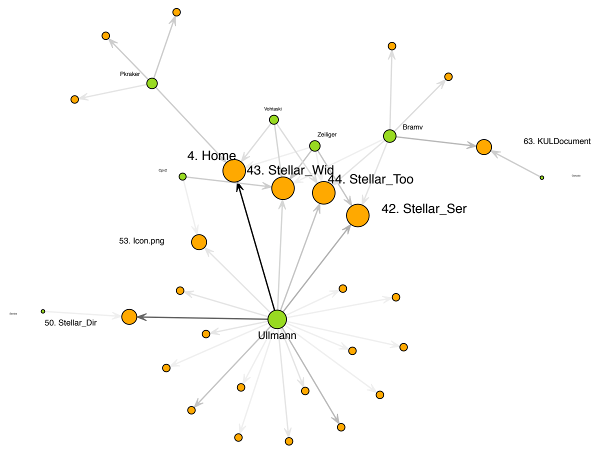

Figure 6. Authors (orange) and their contributions to pages (green): after the submission deadline.

The other deliverables show similar patterns of activity: d7.1 again exposes a larger network of pages (but with a smaller number of contributors), where as d1.1 is significantly reduced in the number of contributors (but still showing a larger number of edits). The deliverable d6.2 shows a star pattern of authors editing the main page and d1.2 ceased its activity with its delivery deadline.

6 Conclusion and outlook

With the analysis presented, the conclusion can be drawn that there definitely is an afterlife for most of the living deliverables. With only one zombie exception, this afterlife is more like a blitheful continuation of activities – relieved of the duty of having a deadline. At least for the one deliverables we have analysed this in more depth and collaboration beyond the deadline exposes a large co-authorship network, accompanied by shift in focus.

As stated the data are extracted from the public revision histories of the living deliverables, made available by MediaWiki. They can be used to show whether wikis show any signs of editing activity and to further investigate the collaboration network structure expressed in these revisions. It is possible to inspect who is collaborating on particular pages. In large projects, like STELLAR, these visualisations can help to make activities more transparent which can create more awareness and accountability – and ultimately offers triggers for new activity.

For living deliverables as such, it provides a way to check for signs of life, especially when their delivery deadline has passed.

There are several limitations this study has. Most notably, collaboration in co-authoring wiki pages cannot be mistaken for the overall collaboration on the (printed) report delivered to the European Commission. All wikis had phases close to the deadline, where an export of the Wikipages into a Word-file served the final polishing and further elaboration. All the deliverables were embedded into collaborative activities of other nature, such as presence and virtual meetings (flashmeetings), reviews (with separate reports), and other forms of collaboration that left no traces in the wikis. Still they are part of the process of creating their content.

Moreover, we have so far looked at only a small number of living deliverables in a limited time period. It will be very interesting to see, whether our findings will be confirmed when repeated in the future with more data and a longer time frame. Not to mention that it will be interesting to see, whether there is an afterlife of the deliverables beyond the runtime of the project.

It is an open question, whether the analysis method used can be matured into a self-explaining visualisation that does not require any insider knowledge about the collaboration in order to correctly read it. Or in other words: an evaluation of usability and accuracy is pending. This might also be helpful further what (wiki-wise) the difference between a living and living dead deliverable is. And it might help to identify driving factors: is it the medium, the collaborators, or the content?

In its current form, the co-editing network plots depict only a holistic view of all contributions. A more flexible approach would be to let the user interactively choose time windows, thereby providing means to investigate collaboration patterns before and after significant events. An animation of the graph change over time would additionally help to understand the development of a living deliverable, emphasizing the process dimension further.

A more fine-grain distinction of the types of contributions and their drivers would serve further analysis: writing passages, proofreading, enhancing with links and media, discussing, altering, and deleting text are all important for the quality of an article, but possibly not all of them trigger further activity by collaborators. This would be equally interesting for life and afterlife of the deliverables.

Additional evidence sources are available to further investigate collaboration among the researchers outside the living deliverable. It would be very interesting to see whether collaboration patterns differ when looking at the accompanying virtual meetings, e-mail exchange, or presence meetings. Does the medium foster certain styles of collaborations or do they converge?

From a project oriented view the proposed type of analysis could serve as a feedback mechanism making achievements visible. This could help to activate discussion about research collaboration.

Acknowledgment

The work presented in this paper was carried out as part of the STELLAR network of excellence, which is funded by the European Commission under the grant agreement number 231913.

Reference

[1] Duncan, W.: Developing a project-management body-of-knowledge document: the US Project Management Institute's approach, 1983-94, In: International Journal of Project Management Vol. 13, No. 2, pp. 89-94, 1995

[2] Arazy, O., Stroulia, E., et al.: Recognizing contributions in wikis: Authorship categories, algorithms, and visualizations. Journal of the American Society for Information Science and Technology. 61, 6, 1166-1179 (2010).

[3] Nunes, S., Ribeiro, C., David, G.: Wikichanges-exposing wikipedia revision activity. Procceedings of the 2008 International Symposium on Wikis (WikiSym) Porto, Portugal. (2008).

[4] Viégas, F.B., Wattenberg, M., Dave, K.: Studying cooperation and conflict between authors with history flow visualizations. Proceedings of the CHI 2004, pp. 575-582, ACM, Vienna, Austria (2004).

[5] Suh, B., Chi, E., Pendleton, B.A., Kittur, A.: Us vs. Them: Understanding Social Dynamics in Wikipedia with Revert Graph Visualizations, In: IEEE Symposium on Visual Analytics Science and Technology 2007, IEEE, Sacramento/CA, USA, 2007.

[6] Baumgrass, A., Mueller, C., Meurath, B.: Analyzing Wiki-based Networks to Improve Knowledge Processes in Organizations, In: Journal of Universal Computer Science, Vol. 14, No. 5, pp. 526-545, ISSN 0948-695x, 2008.

[7] Jesus, R., Schwartz, M., Lehmann, S.: Bipartite Networks of Wikipedia's Articles and Authors: a Meso-level Approach, In: WikiSum'09, ACM, Orlando/Florida, USA, 2009

@twitter Try out #Grabeeter to Export, Archive and Search Your Tweets

Herbert Mühlburger[1], Martin Ebner[1], Behnam Taraghi[1],

[1] Graz University of Technology, Social Learning, Steyrergasse 30/I,
8010 Graz, Austria
{muehlburger, martin.ebner, b.taraghi}@tugraz.at

Abstract: The microblogging platform Twitter is beside Facebook the fastest growing social networking application of the last years. It is used in different ways, e.g. to enhance events (conferences) by sending updates, hyperlinks or other data as a news-stream to a broader public. Until now the stream ends with the end of the event. In this publication a new application is introduced that allows information retrieval and knowledge discovery by searching through local stored tweets related to a corresponding event. The architecture of the prototype is described as well as how the data is being accessed by a web application and a local client. It can be stated that making tweets available after the end of an event, enhances the way we deal with information in future.

Keywords: Knowledge discovery, information retrieval, Twitter, search, microblogging

1 Introduction

Twitter[1] and Facebook[2] are the fastest growing platforms of the last 12 months[3] [12]. On 22nd of February 2010 Twitter hits 50 million tweets per day[4]. Without any exaggeration it can be said that these two social networks are worth to be researched in detail [10] and are of interest for scientists and educators. After a period of testing first results emerge on this form of communication and interaction in science [7] as well as in the area of e-learning [3] [5] [9]. Although Twitter is widely known to be the most popular microblogging platform, a short introduction is given. Templeton [14] defined microblogging as a small-scale form of blogging, generally made up of short, succinct messages, used by both consumers and businesses to share news, post

[1] http://twitter.com (last access: 2010-04)
[2] http://facebook.com (last access: 2010-04)
[3] http://ibo.posterous.com/aktuelle-twitter-zahlen-als-info-grafik (last access: 2010-04)
[4] http://mashable.com/2010/02/22/twitter-50-million-tweets/ (last access: 2010-04)

status updates and carry on conversations. Due to the restriction to 140 characters it can also be compared with a short-message service that is based on an internet service platform. Maybe the factor of success of this application relies on its simplicity - users can send a post (tweet) that is listed on the top of their wall together with messages of their friends. Furthermore any user can be followed by anyone who is interested in that user's updates. By nature Twitter or similar services support the fast exchange of different resources (links, pictures, thoughts) as well as fast and easy communication amongst more or less open communities [2]. In the same way Java [11] defined four main user behaviours why people are using Twitter - for daily chats, for conversation, for sharing information and for reporting news.

Taking a look at the usage of Twitter at conferences we notice the increase of reports, statements, announcements as well as fast conversation between participants. So called Twitter-walls nearby the projection of an ongoing presentation [4] or placed at any other location at the conference support the conference administration, organization, discussions or knowledge exchange. From this point of view microblogging becomes a valuable service reported by different publications [13].

One of the most recent studies on using Twitter at Web 2.0 conferences [1] examined tweets on a semantic basis [6]. The analysis showed that the idea of microblogging usage for distributing or explaining conference topics, discussions or results to a broader public seems to be limited. The authors pointed out that the use of Twitter during conferences should follow logics, like

- Usage as backchannel for conference participants
- Usage of document and illustrate connections
- Usage as a public notepad to collect relevant ideas, quotes or links
- Usage as evaluation tools

Basically there are two core issues - Twitter should be used first for communication between participants instantly and second for documentation on their own. Especially in case of documentation this will be only useful if users are able to create a kind of archive where they can store their tweets.

This publication deals with the research question, what can be the advantages of a web-based application that can also be used offline (without Internet connection) for information retrieval and knowledge discovery based on a micro-content system like Twitter.

"Grabeeter – Grab and Search Your Tweets" is the name of the application that has been developed in order to fulfil these requirements. The next chapter describes Grabeeter in more detail by giving an overview of the system's architecture and its particular features.

2 Architecture of Grabeeter

The architecture of Grabeeter (see Fig. 1) consists of two main parts. The first part is a web application that retrieves tweets and user information from Twitter through the Twitter API[5]. The second part of Grabeeter consists of a client application developed in "JavaFX[6] technology for accessing the stored information on a client side.

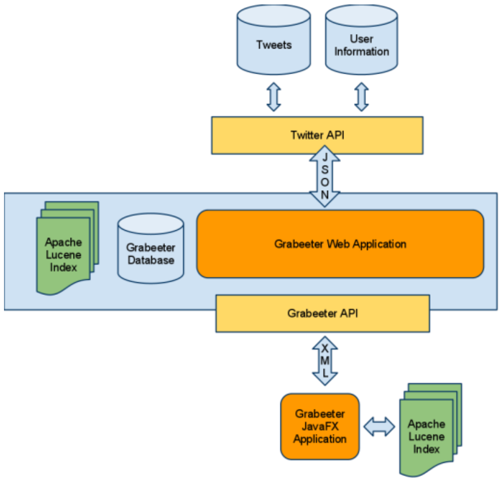

Fig. 1. **Architecture of Grabeeter**

As illustrated in Fig. 1 the Grabeeter web application implements the Twitter API in order to retrieve tweets of predefined users. The tweets are then stored in the Grabeeter database and on the file system as Apache Lucene[7] index. In order to ensure an efficient search the tweets must be indexed. The Grabeeter web application provides access to the Grabeeter database through its own REST style [8] API. This enables client applications to retrieve tweets and user information in an easy way by implementing this API. In difference to the Twitter API Grabeeter API provides all stored tweets and makes no restriction over time.

The Grabeeter client application is developed using JavaFX in order to be independent from different operating systems as well as to provide an easy process to

[5] http://apiwiki.twitter.com/ (last access: 2010-04-21)
[6] http://www.sun.com/software/javafx/ (last access: 2010-04-16)
[7] http://lucene.apache.org/java/docs/ (last access: 2010-04-21)

upgrade the client application using Java Web Start[8]. Furthermore it provides an easy way to store the retrieved tweets on the user's local file system for later offline processing. The following sections describe the different parts of Grabeeter in detail.

2.1 Grabeeter Web Application

The Grabeeter web application enables users to archive their tweets in the Grabeeter database and to perform a search on the stored tweets through a web interface. The tweets are not only stored in the database but also indexed by Apache Lucene in order to support an efficient search on the tweets. These tweets can be accessed then by client applications through the Grabeeter REST style API[9].

As illustrated in Fig. 2 users are able to carry out a search on the stored tweets online or launch the Grabeeter JavaFX Client application by pushing the "Launch" button and search their tweets using the client application.

The workflow of the Grabeeter web application is as follows: At first users register their Twitter usernames at the Grabeeter web application. These usernames are stored in a text file which is parsed later by a cron job. The cron job runs a PHP script that retrieves all accessible tweets for the given usernames. Later another cron job updates the tweets for all monitored users on a scheduled timetable.

[8] http://java.sun.com/javase/6/docs/technotes/guides/javaws/index.html (last access: 2010-04-21)

[9] http://grabeeter.tugraz.at/developers

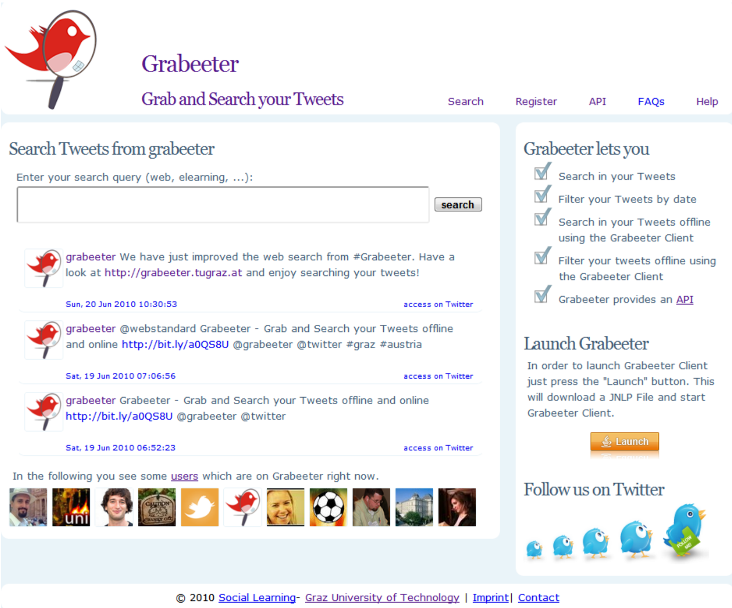

Fig. 2. **Grabeeter Web Application**

Due to Twitter's REST API Limit[10] it is only possible to access the latest 3200 tweets (statuses) via the API of a given user on Twitter. So in case a user has less than 3200 tweets on Twitter at the time of registration on Grabeeter all of the user's tweets are archived. From that time on all future tweets are stored and the entire first (3200 or less) tweets remain accessible and searchable too. In that way all tweets of a user ever become saved and searchable. If a user has more than 3200 tweets on Twitter at the time of registration on Grabeeter it is only possible to retrieve the latest 3200 tweets of this user from Twitter due to the Twitter limit. But from that time on all of the future tweets are archived and searchable through Grabeeter.

Later processing of the stored tweets enables us to achieve more enriched data sets by adding different kind of metadata to the stored information. However this step is not yet implemented and is described in more detail in section 4 regarding future work.

[10] http://apiwiki.twitter.com/Things-Every-Developer-Should-Know#6Therearepaginationlimits (last access: 2010-04-21)

2.2 Grabeeter Client Application

The Grabeeter client application was developed using JavaFX technology. It was tested on different operating systems such as Windows XP, Ubuntu Linux 10.04 and MacOS X running the latest Java SE Runtime Environment.

In order to start the Grabeeter Client application the user clicks the "Launch" button provided on the Grabeeter website (see Fig. 2). While the application starts a shortcut is created on the user's local desktop. Through this shortcut the user is able to restart the application later on instead of using a browser.

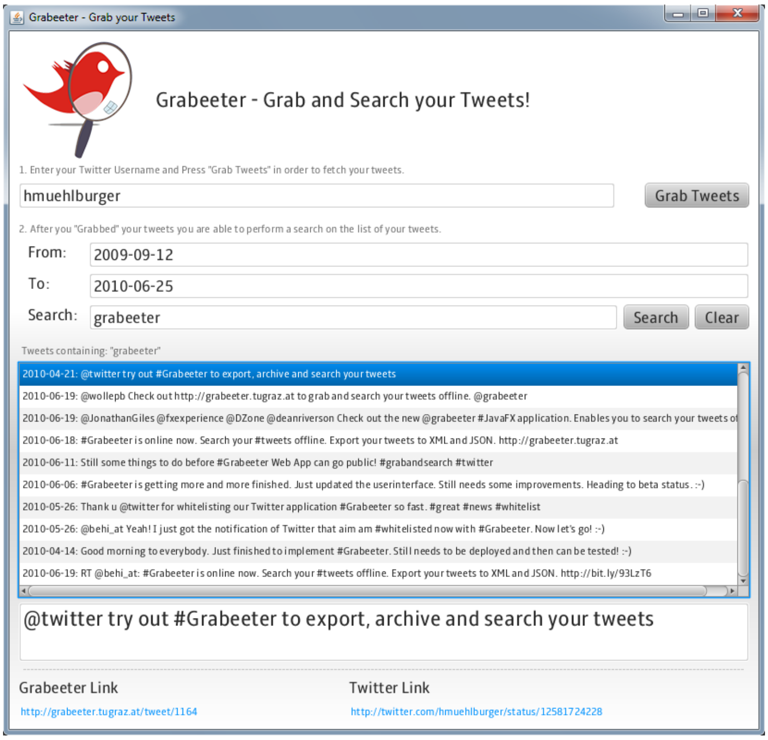

Fig. 3**. Grabeeter Client Application**

The user provides a Twitter username to the client (see Fig. 3) and starts the grabbing of tweets by clicking the button "Grab Tweets". In order to initially grab its tweets the user has to have an internet connection. The Grabeeter Client application then connects to the Grabeeter Database through the Grabeeter API in order to retrieve the tweets. The retrieved tweets are then stored on the local file system in a structured XML format. This enables other applications to access the locally stored tweets for their own purposes.

The Grabeeter application then loads the locally stored tweets and creates an in memory Apache Lucene index. Users are then able to perform a full text search and filter their tweets by specifying a time period.

Initially the Grabeeter Client application works in online mode in order to retrieve and store the entire recent tweets using the Grabeeter API. After restarting the application the locally stored tweets are loaded and indexed again. Therefore users are able to perform searches on tweets without having internet connection and so being independent from web services.

3 Discussion

The following lists interesting aspects that occurred during the development of Grabeeter using JavaFX and the Twitter API.

„Drag-To-Install": One very utile feature of JavaFX is the „Drag-To-Install" possibility. It is the ability for an application to be dragged-out of the browser window and being "installed" on the operating system by dropping it onto the operating system's desktop. The term "installed" means here that a shortcut is created on the desktop and that the JavaFX application is added to the Java application cache on the corresponding operating system. This feature seems not to function properly on MacOS systems so far. From this point of view a new version of the client can be updated in the background without knowledge of the user.

Twitter API restrictions: As already mentioned Twitter REST API requests are restricted to the latest 3200 tweets of a user. There is no chance for any application to access the first tweets, in case the user has already more than 3200 tweets.

Twitter capacity problem: Sometimes the Twitter API is over capacity. In this case no data can be retrieved from the API. This might delay the archive process in Grabeeter web application.

Beside these restrictions Grabeeter may have an interesting effect on the change of writing style: Due to the fact that the suggested tool is able to retrieve data the user is able to document his/her experiences from an event over a time period. This leads to reassess about how we have to use microblogs in general and how we have to write our tweets in order to regain relevant data. Overall this means tweets are written primarily for users themselves and not for a broader public which is a very new aspect to the basic intention of Twitter. With the help of the tool it is now possible to retrieve all tweets concerning a specific hashtag (e.g. event) within a clear defined time frame. Any collected hyperlink can be reused by searching for the specified event and clicking on the appropriate tweet.

If users register on Grabeeter before they reach 3200 tweets on Twitter it is possible to archive and retrieve all tweets from these users. For Grabeeter performs incremental updates and stores all tweets in its archive all tweets of a user are stored continuously from the beginning up to future tweets.

According to our research question in the beginning we like to point out the advantages of the tool Grabeeter:

- Micro-content (tweets) is achievable due to the fact that any tweet can be retrieved at anytime from a local hard-drive
- Micro-content is storable in a way that the user can distinguish between different events
- Micro-content is searchable along keywords, hashtags, time frames as well as different entities (URLs, @, …)

From a technical point of view update process is easily and independence of devices and operating systems is guaranteed.

4 Conclusion and Future Work

Grabeeter was launched in May 2010. The web application as well as the JavaFX client can be accessed at http://grabeeter.tugraz.at.

The rapid improvements in the mobile technology have led to an ascending trend of using mobile applications in recent years. Consequently more users use mobile devices to access online applications. It is planned to build the Grabeeter client as a mobile application for different platforms (Android, iPhone, JavaFX devices …). The adaptations that must be performed are mainly the view adjustment and an appropriate look and feel for the mobile environments.

The next main extension of Grabeeter will be the capability not only to retrieve the search results of a simple search query to the user, but also to combine multiple search queries over multiple users for the analysis of the archived data sets, for data exploration and a better knowledge discovery. Use of semantic technologies and interlinking techniques for this purpose would definitely enrich the data sets and enhance the usefulness of stored tweets. The first step will be to describe the archived data sets semantically, to "triplify" the data sets and convert them to RDF triples by applying the existing vocabularies used for microblogging.

The tweets of each user can be extracted and analysed towards relevant keywords to get a feeling about the main topics for e.g. a specific event. The text fragments in tweets can be extracted and interlinked with resources in the Linked Open Data[11] (LOD) cloud such as DBpedia, Flickr, Geo-names, etc. The Twitter users can be interlinked with FOAF profiles in the LOD cloud too. Having the data sets triplified and interlinked with LOD it will be more efficient to analyse the collected data from the Twitter API. It will become possible to perform a more accurate knowledge

[11] http://richard.cyganiak.de/2007/10/lod/ (last access: 2010-04-16)

discovery and retrieve search results not only within tweets gained from the Twitter API, but also in interlinked resources of the World Wide Web.

Furthermore a SPARQL[12] endpoint can be provided in Grabeeter web application to let different monitoring and analysing client applications to perform SPARQL queries over semantic data sets. As an example searching for tweets containing a geographic term such as “Vienna” would return also the tweets that contain the term “Wien”, which is the German word for Vienna. Search queries can be made even much more complex:

- Get tweets that contain links to photos related to the place where conference xy takes place.
- Get tweets that are related to informatics and semantic technologies.

It can be summarized that the described application allows retrieving status updates from the most famous microblogging platform Twitter for information retrieval on a local hard drive. Furthermore through the combination of tweets from different Twitter users with predefined keywords or hashtags the knowledge discovery seems to be opened up in a new dimension. For the first time the documentation of events by just simply tweeting of statements, hyperlinks or media files becomes possible. Grabeeter is built to enhance the usefulness of microblogging on conferences and allows retrieving data that was produced just on the fly.

References

1. Bernhardt, T., Kirchner, M.: Web 2.0 meets conference – the EduCamp as a new format of participation and exchange in the world of education. In: Ebner, Martin / Schiefner, Mandy (Eds.): Looking Toward the Future of Technology-Enhanced Education: Ubiquitous Learning and the Digital Native. IGI Global, Hershey, 2009
2. Boyd, d., Golder, S., Lotan, G.: Tweet, tweet, retweet: Conversational aspects of retweeting on twitter. In Proceedings of the HICSS-43 Conference, January 2010.
3. Costa, C., Beham, G., Reinhardt, W., Sillaots, M.: Microblogging in Technology Enhanced Learning: A Use-Case Inspection of PPE Summer School 2008. In: Proceedings of the 2nd SIRTEL workshop on Social Information Retrieval for Technology Enhanced Learning, 2008
4. Ebner, M.: Introducing Live Microblogging: How Single Presentations Can Be Enhanced by the Mass. Journal of Research in Innovative Teaching (JRIT), 2 (1), p. 91- 100, 2009
5. Ebner, M., Lienhardt, C., Rohs, M., Meyer, I.: Microblogs in Higher Education – a chance to facilitate informal and process oriented learning?. Computers & Education, ISSN 0360-1315, DOI: 10.1016/j.compedu.2009.12.006., 2010

[12] http://www.w3.org/TR/rdf-sparql-query/ (last access: 2010-04-21)

6. Ebner, M., Mühlburger, H., Schaffert, S., Schiefner, M., Reinhardt, W.: Get Granular on Twitter - Tweets from a Conference and their limited Usefulness for Non-Participants, accepted paper at Workshop (MicroECoP). WCC 2010 conference, Brisbane, Australia, 2010

7. Ebner, M., Reinhardt, W.: Social networking in scientific conferences – Twitter as tool for strengthens a scientific community. In: Proceedings of the 1st International Workshop on Science 2.0 for TEL, Ectel 2009, September 2009

8. Fielding R.: Architectural Styles and the Design of Network-based Software Architectures. 2000, http://www.ics.uci.edu/~fielding/pubs/dissertation/top.htm (last access: 2009-01)

9. Grosseck, G., Holotescu, C.: Can we use twitter for educational activities?. In Proceedings of the 4th International Scientific Conference eLSE "eLearning and Software for Education", April 2008

10. Haewoon K., Changhyun, L., Hosung, P. Moon, S.: What is Twitter a Social Network or a News Media?. Proceedings of the 19th International World Wide Web (WWW) Conference, April 26-30, 2010, Raleigh NC (USA), April 2010, http://an.kaist.ac.kr/traces/WWW2010.html (last access: 2010-04)

11. Java, A., Song, X., Finin, T., Tseng, B.: Why we twitter: understanding microblogging usage and communities. In Proceedings of the 9th WebKDD and 1st SNA- KDD 2007 workshop on Web mining and social network analysis, pages 56– 65. ACM, 2007.

12. McGiboney, M.: Keep on tweet'n. March 2009, http://www.nielsen-online.com/blog/2009/03/20/keep-on-tweetn/ (last access: 2010-04)

13. Reinhardt, W., Ebner, M., Beham, G., Costa, C.: How people are using Twitter during Conferences. In: Hornung-Praehauser, V., Luckmann, M. (Eds.): Creativity and Innovation Competencies on the Web. Proceedings of the 5th EduMedia 2009, Salzburg, pages 145-156, 2009.

14. Templeton, M.: Microblogging defined, http://microblink.com/2008/11/11/microblogging-defined/, 2008, (last access: 2010-04)

Connecting Early Career Researchers: Investigating the Needs of Ph.D. Candidates in TEL Working with Web 2.0

Nina Heinze[1], Marie Joubert[1], Denis Gillet[1]

[1] STELLAR Network of Excellence
n.heinze@iwm-kmrc.de, marie.joubert@bristol.ac.uk, denis.gillet@epfl.ch

Abstract. This article describes the results of a case study conducted amongst 21 doctoral candidates and three senior researchers at the Joint European Summer School on Technology Enhanced Learning 2010. The study aims to analyse the needs of early career researchers working within the field of TEL in geographically distant communities, particularly with respect to online collaboration, communication and information exchange. This study can be seen as a needs analysis on support structures to enable research 2.0 in TEL among young researchers.

Keywords: communities of practice, Research 2.0, Social Media, Web 2.0, case study, awareness support

1 Introduction

Our personal experience suggests that collaboration and communication within the European TEL community usually looks like this: researchers use many offline and web-based tools to work and to share their findings and opinions, there is no standardised way of communicating, and various channels are used to disseminate information. It is difficult to keep up with who is doing what in the field, though many researchers are making a considerable effort to monitor the data that is being spread on the Web by colleagues [1], [2], [3]. Ph.D. candidates new to the field frequently have problems finding relevant information, people, events and platforms to help them in their research endeavours. Recent talks with a number of Ph.D. students we are in touch with have underlined these perceptions.

Some efforts have been undertaken to make it easier for doctoral candidates to stay up-to-date on current topics and events and to enable them to collaborate online. These include the establishment of inter- and transorganisational mailing lists, newsgroups, social media groups or forums[1]. Despite these efforts, however, anecdotal evidence from our discussions with Ph.D. students indicates that doctoral candidates still feel that support in terms of information and collaboration could be improved. To address these concerns, the STELLAR Network of Excellence[2] supports doctoral events that aim to improve collaboration and communication between junior and senior researchers as well as enhance the flow of information. In addition, STELLAR also plans to create a virtual doctoral community of practice (DoCoP) to help Ph.D. candidates stay in touch, share and conduct research, help each other solve problems and get in touch with further junior and senior researchers by means of Web 2.0 technologies, the latter being nowadays referred as social media. We understand Communities of Practice (CoP) to be a group of people who share the same interests and passion for something they do and shape their identity by a shared domain of interest whilst engaging in activities around this domain with other members of the community. They thereby develop a shared repertoire of resources, a shared practice, as Wenger calls it in his explanation of a CoP [4]. For an overview of the implications of CoP's on learning and the possibilities of online CoP's see [4], [5], [6].

We saw it necessary to develop an understanding of the needs of Ph.D. candidates as the starting point for the development of the DoCoP planned in STELLAR. Our first step towards developing such an understanding was to consult with Ph.D. candidates.

An opportunity to do so arose at the 2010 Joint European Summer School on Technology Enhanced Learning, which took place in June 2010, gathering together about 50 Ph.D. candidates working in TEL. We conducted a workshop with focus on students' views on the creation of a doctoral community of practice in the field of TEL. 21 doctoral candidates as well as three senior researchers participated in the workshop. We asked them about what type of information may be of value to them to increase awareness in terms of collaboration, what type of awareness support would be of use to them, what tools they use when collaborating in dislocated research teams and how they believe a sustainable community of practice can be implemented. We report about our findings below.

2 Consulting on a DoCoP with Ph.D. Candidates in TEL – A Case Study

During the workshop at the Summer School the doctoral candidates worked in groups of 5-6 people and were asked to discuss how they would wish to receive support for their doctoral work in terms of personal support, awareness support, tools for collaboration and the characteristics of a doctoral community of practice that would be of value to them. Each group then presented their findings, explained their results and engaged in discussions about their thoughts with the other participants of the

1 Examples include JTEL Summer and Winter Schools, Doctoral Consortia at conferences like EC-TEL or Earli, the STELLAR Mobility Programme or DocNet from the University of St. Gallen, Switzerland

2 http://www.stellarnet.eu

workshop. We recorded the entire session to be able to further analyse the results after the Summer School.

2.1 Results of analysis of needs of Ph.D. candidates in TEL

We analysed their reported needs and categorized them into two levels, each describing the personal involvement or gain of the individual researcher (see Table 1). The *individual level* of needs describes issues that occur on an individual level like review of one's own paper or managing one's own information. Support on this level aids the individual in her endeavour more than it does a larger peer-group. The *community level* is the actual community or peer-group level. Support on this level is useful for more than the individual researcher. A larger CoP would benefit from assistance on this level. Table 1, below, summarises the findings within each of these two categories.

Table 1. Needs of doctoral students on the individual and community levels.

Individual Level	Community Level
Peer-review of artefacts	Information modelling
Methodology	Researcher information
Problem solving	Futuregazing
General feedback	Networking
Jobs / internships / exchange programmes	Guidelines for community management
F2F meetings	Sharing testbeds / datasets
Information management	Peer groups
	Collaborative filtering

As we can see from Table 1 doctoral candidates would, on the one hand, appreciate support on a very individual level concerning the process of finishing their Ph.D. thesis like advice on the methodology they are planning to use, how to solve problems they encounter when doing their research as well as meeting face to face with a senior scientist to discuss their work to be able to better evaluate if they are on the right track. On the other hand, doctoral candidates see the need for a community of peers working in related fields to network, discuss their work, get a notion of where others in the field are, what their work is about and how they cope with writing a Ph.D.. In addition they would like to get feedback from a community of peers on their work and share research findings and data.

When we asked them about how they believe they can be supported in their endeavours and needs on a technical level we received answers related to information gathering like RSS feeds from relevant sites, collaboration tools like a semantic wiki with an ontology as well as information filtering tools like recommender systems and a reputation system to enable them to better match the information with their current needs. The proposed solutions Ph.D. candidates gave revolve around support issues that have a high technical (system) component. They require the provision of some sort of Web 2.0 tool or are in essence already a tool.

What we can see from the distinction we made is that the categorization of needs in two levels is not a sufficient distinction, since some issues on the individual and community levels are at the same time themes that fall into the area of proposed solutions like networking or sharing testbeds. This is not a surprise, though, since communication, collaboration and awareness of a community go hand in hand.

2.2 Results of awareness support of Ph.D. candidates in TEL

In addition, we asked the 4 groups to consider what kind of awareness support may be helpful in research communities with respect to contributing to increased productivity. With awareness we mean the state or quality of being aware of the current themes, projects, events and researchers including their background within the field of TEL and one's own position within it. Again they discussed within their groups and presented their findings in a plenary.

We analysed the plenary discussions and were able to place the findings into two areas. The first area, *personal,* pertains to information available on the personal/professional background of other researchers and contains topics like research background or projects that the person has worked on. The second area of interest in awareness support, *research,* concerns information on the actual output of researchers (artefacts like publications) as well as opinions of others about them. Table 2 sums up the awareness support results of the case-study participants.

Table 2. Awareness support

Personal level	Research level
Research background	Artefacts / publications
Expertise / Competencies	State-of-the-art of topic
Projects	Opinions from peers
Social media handles[3]	

Table 2 shows that doctoral candidates wish to have personal information on people within their area of research in terms of scientific background and expertise, as well as their online handles like Twitter and delicious user names or blogs. On the research level they suggest information on current artefacts and publications, as well as the state-of-the art of research in their field and opinions from peers on research, publications and other researchers.

When asked about technical solutions to make it possible to gather and filter information within the community to increase one's awareness of the field of TEL in terms of people, topics, and events, the Ph.D. students proposed open-source solutions to share datasets as well as reputation mechanisms to increase awareness of and within the TEL community. However, the results on the tool level were low which we believe is due to the fact that there are few good services available and the time we gave the doctoral candidates was too short to come up with productive and creative feedback.

[3] social media handels are usernames for social media services like Twitter, Delicious, Slideshare or URL's to blogs or wikis

2.3 Suggestions for the creation of a doctoral community of practice by Ph.D. candidates in TEL

The last part of the workshop revolved around collecting ideas on how a sustainable virtual doctoral community of practice (DoCoP) amongst former and future Ph.D. candidates participating in STELLAR doctoral events could be established and maintained. We saw a key consideration within this discussion as the tools used to support the DoCoP. Further, participants were also asked which Web 2.0 tools they use in their own practice and for what purposes in order to inform our understanding of what they value. This discussion, again, took place amongst the whole group.

Our analysis of the discussions led to three main results. The first is that the participants in our case study find it unlikely that a larger doctoral community of practice can be sustained in a reasonable manner by itself. Their experience is that events such as, for example, the Summer School, function as an umbrella, or a macro-level of community, out of which several smaller, actual communities of practice arise with about 6 to 10 members. The Ph.D. candidates suggested that these smaller communities of practice should be supported not by a particular tool or service, since the community members would decide on those depending on their needs and habits, but rather by the provision of guidelines on collaboration, including the use of existing Web 2.0 tools for research and community management.

The second conclusion the participants drew was that the sustainability of a community of practice, based on the philosophy underpinning Research 2.0, would be highly dependent on individuals dedicated to it. They concluded that the community is independent of the tools in the sense that tools are used regardless of the community. Participants recommended a community facilitator to keep the flow of information going and the community members active in participating.

The third conclusion was that the tool or service needs to fulfil collaboration and communication functions and should be user-friendly in the sense that it is easy to use. The doctoral candidates already use a number of tools for these purposes as well as for research and the organisation of their projects, they did not see the pressing need for a "new" tool or platform.

Table 3, below, summarises the participants' reported use of Web 2.0 tools for communication, collaboration, research instruments and organisation.

Table 3. Tools used by case-study participants

Tool	Communication	Collaboration	Research	Organization
E-mail	x	x		x
Google Docs		x		
Google Talk	x			
Google Scholar			x	
Google Analytics			x	
Google Forms			x	
Google Sites	x	x		x
Google Wave	x	x	x	x
BSCW	x	x		x
Dropbox		x		

Mendeley		x	x	
Group Wikis	x	x	x	x
FlashMeeting	x	x		
Skype	x			
MSN Messenger	x			
Doodle				x
Gigapedia			x	
Library			x	

3 Conclusions

The results show that Research 2.0 in a doctoral community takes place on many different levels and involves quite a few issues that need to be taken into account. For one, Ph.D. candidates spend time working alone, independently on their thesis and would value support on a very personal, face-to-face level from senior researchers. Further, doctoral candidates appreciate a community of peers they can discuss problems with, share results and remain up-to-date on what is happening in their field of research. They would like to have tools that make it easier for them to gather information on relevant researchers to their topic, important events, possibilities for scholarships and internships as well as collaboration tools like a semantic wiki to collaborate and share findings. In addition, doctoral candidates find an awareness support system useful that allows them to see how is doing what in the TEL community with whom.

In terms of creating a sustainable doctoral community of practice within the field of TEL we could distinguish two main findings: we have a large, fuzzy community of TEL researchers and Ph.D. candidates. Bringing them together in one virtual doctoral community of practice and having them all collaborate and communicate seems unlikely. However, this large community is in need of a virtual space that collects information, makes it available to others and has mechanisms to share that information to increase awareness of the community and bring it closer together. This type of umbrella-platform can enable the smaller communities within the field of TEL gather under the same roof, form and proliferate and share information within the smaller communities as well as the larger TEL community.

Our second conclusion is that there seems to be little need to develop a super-tool to fulfil the needs of Ph.D. students to work and collaborate in their community. What we could see is that doctoral candidates use tools for collaborating, communicating, conducting research and organizing their work flow processes and information. There is little need for yet another tool according to the workshop participants. In addition, the participants noted that preferences as well as needs differ, so tool choice should be left up to the Ph.D. candidates. Rather, there is a need for guidelines on existing tools and their use for research.

In summary, we can say that the findings from the workshop we conducted lead to the conclusion that Ph.D. candidates working within the field of TEL feel they have sufficient Web 2.0 tools at their disposal but would appreciate more support in terms of their use as well as finding and filtering information relevant to their research.

References

1. Reinhardt, W., Moi, M., Varlemann, T.: Artefact-Actor-Networks as tie between social networks and artefact networks. In Proceedings of the 5th International Conference on Collaborative Computing (CollaborateCom'09) (2009)
2. Ochoa, X., Mendez, G., Duval, E.: Who we are: Analysis of 10 years of the ED-Media Conference. In: Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications (ED-Media 2009), pp. 189-200 (2009)
3. Fisichella, M., Herder, E., Marenzi, I., Nejdl, W.: Who are you working with? Visualizing TEL Research Communities. In: Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications (ED-Media 2010), pp. 2010 (2010)
4. Wenger, E.: Communities of Practice: Learning, Meaning, and Identity. Cambridge University Press. New York (1998)
5. Wenger, E., McDermott, R., Snyder, W.: Cultivating Communities of Practice: A Guide to Managing Knowledge. Harvard Business School Press, Boston (2001)
6. Wenger, E., White, N., Smith, J.D.: Digital Habitats: Stewarding Technology for Communities. CPSquare, Portland (2009)

Acknowledgement

This case study was carried out within the STELLAR Network of Excellence under the European Seventh Framework Programme (FP7).