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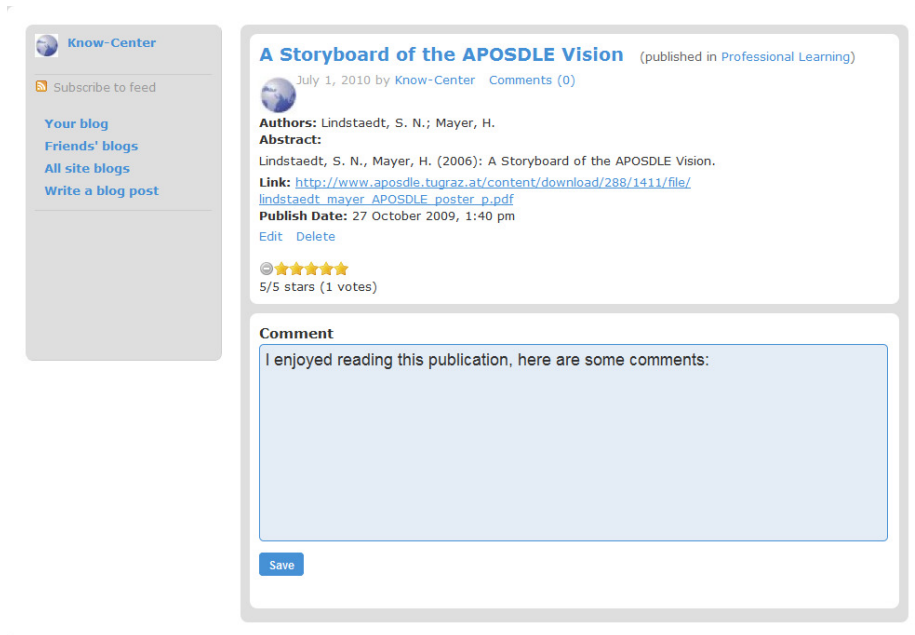
Fig. 1. Rating and commenting features of the Publication Feed Integration Widget

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.
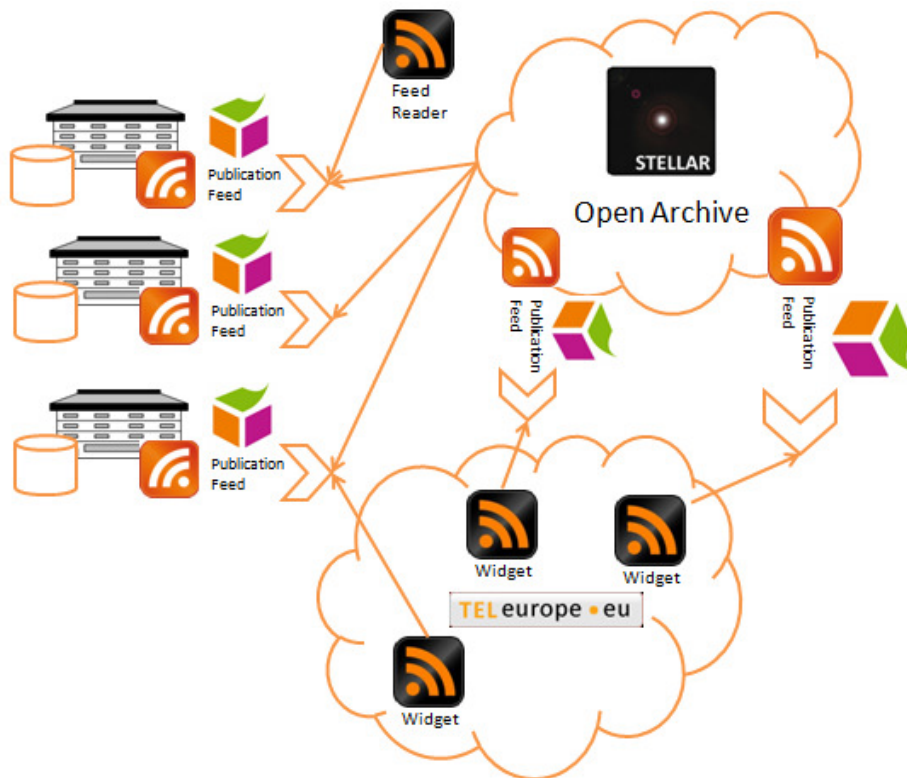

Fig. 2. Overview of the integration of the ecosystem in STELLAR

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

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/