OpenResearch: Collaborative Management of Scholarly Communication Metadata

Sahar Vahdati[1], Natanael Arndt[2], Sören Auer[1,3], and Christoph Lange[1,3]

[1] Enterprise Information Systems (EIS), University of Bonn, Germany
{vahdati, auer}@cs.uni-bonn.de
[2] AKSW Group, University of Leipzig, Germany
arndt@informatik.uni-leipzig.de
[3] Fraunhofer, Intelligent Analysis and Information Systems (IAIS), Germany
math.semantic.web@gmail.com

Abstract Scholars often need to search for matching, high-profile scientific events to publish their research results. Information about topical focus and quality of events is not made sufficiently explicit in the existing communication channels where events are announced. Therefore, scholars have to spend a lot of time on reading and assessing calls for papers but might still not find the right event. Additionally, events might be overlooked because of the large number of events announced every day. We introduce OpenResearch, a crowd sourcing platform that supports researchers in collecting, organizing, sharing and disseminating information about scientific events in a structured way. It enables quality-related queries over a multidisciplinary collection of events according to a broad range of criteria such as acceptance rate, sustainability of event series, and reputation of people and organizations. Events are represented in different views using map extensions, calendar and time-line visualizations. We have systematically evaluated the timeliness, usability and performance of OpenResearch.

Keywords: Scientific Events, Collaborative Knowledge Acquisition, Semantic Publishing, Semantic Wikis, Linked Data

1 Introduction

There is currently an era of departure to investigating how scholarly work and communication can be taken to the digital world. Much attention is devoted to new forms of publishing (e.g. semantic papers, micro-publications), open access, and free availability of publication metadata. Still, a large number of scholarly communication processes and artifacts (other than publications) are not currently well supported. This includes in particular information about events (conferences, workshops), projects, tools, funding calls etc. In particular for young researchers and interdisciplinary work it is of paramount importance to be able to easily identify venues, actors and organizations in a certain field and to assess their quality.

Research results are published as scientific papers in journals and events such as conferences, workshops etc. Each component of this communication needs to be open and easily accessible. Besides conducting their actual research, scholars often need to search for scientific events to submit their research results to, for projects relevant to their research, for potential project partners and related research schools, for funding possibilities that support their particular research agenda, or for available tools supporting their research methodology. For lack of better support, scholars rely a lot on individual experience, recommendations from colleagues and informal community wisdom, they do simple Web searches or subscribe to mailing lists and are stuck with simplistic rankings such as calls for papers (CfPs) sorted by deadline. Domain specific mailing lists are a medium often used by conference and workshop organizers for posting initial, second, final calls for papers, as well as deadline extensions. But this situation leads to discussions on whether to allow calls for papers on the lists or threat them as spam[4] It is especially hard for subscribers to filter those calls according to their individual interests, or maybe explicitly subscribe to important information, such as deadline extensions or subsequent calls, on a specific event or an event series.

On the other hand, the quality of scientific events is directly connected to the research impact and the rankings of the scientific *papers* published by them. For example, the *Research Excellence Framework* (REF) for assessing the quality of research in UK higher education institutions, classifies publications by the venues they are published in. This facilitates assessing every researcher's impact based on the number of publications in conferences and journals. Providing such information to researchers supports them with a broader range of options and a comprehensive list of criteria while they are searching for events to submit their research contributions. To provide comprehensive information about scientific venues, projects, results etc., we present `OpenResearch.org`. OpenResearch is a platform for automating and crowd-sourcing the collection and integration of semantically structured metadata about scholarly communication. In particular, with regard to events, OpenResearch ...

1. reduces the effort for researchers to find 'suitable' events (according to different metrics) to present their research results,
2. supports event organizers in visibly promoting their event,
3. establishes a comprehensive ranking of events by quality,
4. provides a cross-domain service recommending suitable submission targets to authors, and
5. supports easy and flexible data exploration using Linked Data technology: a structured dataset of conferences facilitates selection regarding fields of interest or quality of events.

OpenResearch empowers researchers of any field to collect, organize, share and disseminate information about scientific events, projects, organizations, funding sources and available tools. It enables the community to define views as queries

[4] Note a recent survey on calls for papers on the W3C mailing lists: `https://lists.w3.org/Archives/Public/semantic-web/2016Mar/0108.html`.

over the collected data; assuming sufficient data, such queries can enable rankings by relevance or quality. Driven by Semantic MediaWiki (SMW), OpenResearch provides a user interface for creating and editing semantically structured event profiles, tool and project descriptions, etc. in a collaborative wiki way. OpenResearch is part of a greater research and development agenda for enabling true open access to all types of scholarly communication metadata (beyond bibliographic ones) not just from a legal but also from a technical perspective. The work on OpenResearch is aligned with *OpenAIRE*, the Open Access Infrastructure for Research in Europe.

The remainder of this paper is organized as follows: Section 2 states the problem that OpenResearch intends to address. Section 3 presents the state of the art of existing services addressing the same problem. Section 4 establishes requirements for a system that can address the problem in a comprehensive way. Section 5 explains the approach and architecture of the OpenResearch platform. Section 6 presents the services that OpenResearch provides to its end users today. Section 7 discusses how we have assessed the time-lines, usability and performance of OpenResearch. Section 8 concludes and outlines future work.

2 Problem Statement

Challenge 1: Communication. Research communities use different communication channels to distribute event announcements and CfPs. Announcing CfPs through different mailing lists is the traditional but still most popular way of disseminating information about an event. Exploring the calls for papers posted on mailing lists of the Semantic Web community shows that 500 to 700 event announcements have been posted every year between 2006 and 2016 (approx. 15-30% of the overall traffic). This shows that a large and widely spread amount of unstructured data about scientific events is increasingly being published via communication channels not specifically designed for this purpose. Due to the interdisciplinary nature of research, event organizers easily overlook relevant channels to announce their event. In addition, browsing through the CfPs in several channels to identify events that might be of interest is a time and effort consuming task.

Challenge 2: Structure. There are structural differences across events, for example, events with many co-located events or sub-events, or new events emerged from multiple smaller ones. One example for the latter is the Conference on Intelligent Computer Mathematics (CICM), which results from the convergence of four conferences that used to be separate but now are tracks of a single conference.[5] Scholars who want to find out whether an event matches their research interests therefore have to understand its structure; if they cannot find the desired information for the super-event, they will have to study the sub-events.

Challenge 3: Series. Most scientific events occur in series, whose individual editions took place in different locations with narrow topical changes. Researchers often need to explore several resources to obtain an overview of the previous

[5] http://www.cicm-conference.org/

editions of an event series to be able to estimate the quality of the next upcoming event in this series.

Challenge 4: Addressing Different Stakeholders. Event organizers aim to attract as many submitters as possible to their events. Publishers want to know whether they should accept a particular event's proceedings in their renowned proceedings series. Potential PC members want to decide whether it is worth spending time in the reviewing process of an event. Similarly, sponsors and invited speakers need to decide whether a certain event is worth sponsoring or attending. Researchers receiving CfP emails have to distinguish whether the event is appropriate for presenting their work. Researchers searching for events through various communication channels assess events based on criteria such as thematic relevance, feasibility of the deadline, close location, low registration fee etc. The organizers of smaller events who plan to organize their event as a sub-event of a bigger event have to decide whether this is the *right* venue to co-locate with. These examples prove the importance of filtering events by topic and quality from the point of view of different stakeholders. Currently, the space of information around scientific events is organized in a cumbersome way, thus preventing events' stakeholders from making informed decisions, and preventing a competition of events around quality, economy and efficiency.

Strategies. Event organizers employ a number of strategies to cope with the challenges of advertising their event and engaging with the potential audience. They use multiple channels (mailing lists, social networks, homepages) to distribute CfPs. Some organizers plan deadline extensions in advance, as a strategy to attract more submissions. Some communities employ databases on top of mailing lists for announcing scientific events e.g., researchers in information systems and databases use the DBWorld database (cf. Section 3). The strategies mentioned so far target authors of submissions, whereas event organizers also have to find sponsors, high-profile program committee members and keynote speakers. This is currently done by contacting researchers or companies that the organizers know already. An approach for a centralized and holistic infrastructure for managing the information about scientific events was missing so far.

3 Related Work

CfP classification and annotation: *CFP Manager* [4] is an information extraction tool specific to the domain of computer science; it extracts metadata of events from an unstructured text representation of CfPs. Because of the different representations and terminologies of CfPs across research communities, this approach requires domain specific implementations. The extracted data is limited to the keywords used in the content of CfPs. In addition, CFP Manager does not support data curation workflows involving multiple stakeholders. Hurtado Martin et al. proposed an approach based on user profiles, which takes a scholar's recent publication list and recommends related CfPs using content analysis [3]. Xia et al. presented a classification method to filter CfPs by social tagging [10]. Wang et al. proposed another approach to classify CfPs by implementing three

different methods but focus on comparing the classification methods rather than services to improve scientific communication [9].

Websites: *Google Scholar Metrics (GSM)*[6] provides ranked lists of conferences and journals by scientific field based on a 5-year impact analysis over the Google Scholar citation data. 20 top-ranked conferences and journals are shown for each (sub-)field. The ranking is based on the two metrics h5-index[7] and h5-median[8]. GSM's ranking method only considers the number of citations, whereas we intend to offer a multi-disciplinary service with a flexible search mechanism based on several quality metrics. *DBLP*[9], one of the most widely known bibliographic databases in computer science, provides information mainly about publications but also considers related entities such as authors, editors, conference proceedings and journals. Events, deadlines and subjects are out of DBLP's scope. DBLP allows event organizers to upload XML data with bibliographic data for ingestion. The dataset of DBLP is available as an RDF dump [10] *DBWorld*[11] collects data about upcoming events and other announcements in the field of databases and information systems. Each record comprises event title, deadline, event homepage and the full-text description. *WikiCFP*[12] is a popular service for publishing CfPs. Like DBWorld, WikiCFP only supports a limited set of structured event metadata (title, dates, deadlines), which results in limited search and exploration functionality. WikiCFP employs crawlers to track high-profile conferences. Although WikiCFP claims to be a semantic wiki, there is no collaborative authoring, versioning, minimal structure and the data is not downloadable as RDF or accessible via a SPARQL endpoint. *Cfplist*[13] works similar to WikiCFP but focuses on social science related subjects. Data is contributed by the community using an online form.*SemanticScholar*[14] offers a keyword-based search facility that shows metadata about publications, authors. It uses artificial intelligence methods in the back-end and retrieves results based on highly relevant hits with possibility of filtering.

Datasets: *ScholarlyData*[15] provides RDF dumps for scientific events. Conference-Ontology, a new data model developed for ScholarlyData, improves over already existing ontologies about scientific events such as the Semantic Web Dog Food (SWDF) ontology. *Springer LOD*[16] is a portal publishing conference metadata collected from the traditional publishing process of Springer as Linked Open

[6] https://scholar.google.com/intl/en/scholar/metrics.html
[7] h5-index is the h-index for articles published in the last 5 complete years.
[8] 5-median is the median number of citations for those articles in the h5-index.
[9] http://dblp.uni-trier.de/
[10] http://dblp.l3s.de/d2r/
[11] https://research.cs.wisc.edu/dbworld/
[12] http://www.wikicfp.com/
[13] https://www.cfplist.com/
[14] https://www.semanticscholar.org
[15] http://www.scholarlydata.org/dumps/
[16] http://lod.springer.com/

Service	Entities	Event Series	Sub-events	Quality Criteria	Community Contribution	Advanced Search	LOD
CFP Manager	Events	✗	✗	✗	✗	✗	✗
GSM	Conferences, Journals	✗	✗	✗	✗	✗	✗
DBLP	Publications, Person	✗	✗	✗	✗	✗	✗
DBWorld	Events	✓	✓	✗	✗	✗	✗
WikiCFP	Events	✗	✗	✗	✓	✗	✗
Cfplist	Events	✗	✗	✗	✓	✓	✗
CiteSeer	Publications, Events	✗	✗	✗	✗	✗	✓
ScholarlyData	Events	✗	✗	✗	✗	✗	✓
Springer LOD	Events	✗	✗	✗	✗	✗	✓
Conference.city	Events	✗	✗	✗	✗	✗	✗
PapersInvited	Events	✗	✗	✗	✗	✗	✗
SemanticScholar	Publication, Person	✗	✗	✗	✗	✗	✗

Table 1: Comparison of existing services.

Data. All these conferences are related to Computer Science. The data is available through a SPARQL endpoint, which makes it possible to search or browse the data. A graph visualization of the results is also available. For each conference, there is information about its acronym, location and time, and a link to the conferences series. The aim of this service is to enrich Springer's own metadata and link them to related datasets in the LOD Cloud.

Other services: *Conference.city*[17] is a new service initialized in 2016 that lists upcoming conferences by location. For each conference, title, date, deadline, location and number of views of its conference.city page are shown. Based on the location of the conference, Google plug-ins are used to recommend flights, accommodation and restaurants. The service collects data mainly from event homepages and from mailing lists. In addition, it allows users to add a conference using a form. *PapersInvited*[18] focuses on collecting CfPs from event organizers and attracting potential participants who already have access to the ProQuest service[19]. ProQuest acts as a hub between repositories holding rich and diverse scholarly data.The collected data is not made available to the public.

Conclusion: The comparison of currently available services in Table 1 shows that collaborative management of scholarly communication metadata in particular for events is not yet sufficiently supported.

4 Requirements

A collaborative and partially decentralized environment is required to enable community-based scientific data curation and extension, and to tap into the 'wisdom of the crowd' for elicitation and representation of metadata associated

[17] http://conference.city/

[18] http://www.papersinvited.com/

[19] http://www.proquest.com/

to scholarly communication. In particular, such a system is aimed to address the following requirements as services, which we have derived from the challenges C1–C4 pointed out in the problem statement and from the review of related work (R):

R1 It should be easily possible to create various views on the resulting data (addressing various communities), also in a collaborative way. (C1)
R2 Fine-grained and user extensible semantic representation of the (meta)data should be supported. (C1)
R3 The resulting ontological model should capture the relationships between various types of entities (e.g. event series, sub/super events, roles in event organization, etc.). (C2, C3)
R4 Different stakeholders of scholarly communication (event organizers, PC members, developers, etc.) have to be supported adequately. (C4)
R5 The data representation and view generation mechanisms should support fine-grained analyses (e.g. about the quality of events according to various indicators). (C4)
R6 The collaborative authoring and curation interfaces should be user friendly and enable novices to participate in the data gathering and curation processes.(C4)
R7 The system architecture should support automatic as well as manual/crowd-sourced data gathering from a variety of information sources. (R)
R8 All changes should be versioned to support tracking particular users' contributions and their review by the community. (R)
R9 The collected data should be easily reusable by application and service developers. (R)

5 Approach

The core of the OpenResearch approach is to balance manual/crowd-sourced contributions and automated methods. OpenResearch uses semantic descriptions of scientific events based on a comprehensive ontology; this enables distributed data collection by embedding markup in conference websites aligned with schema.org, and links to other portals and services. Semantic MediaWiki (SMW) serves as data curation interface employing semantic forms, templates various extensions and semantic annotations in the wiki markup. In the remainder, we describe the data model of OpenResearch and its architecture.

5.1 Data Model

The vocabulary used in OpenResearch reuses existing vocabularies from related domains, since reuse increases the value of semantic data. Existing related vocabularies are the *Semantic Web Conference Ontology* (SWC)[20], the *Semantic*

[20] http://data.semanticweb.org/ns/swc/swc_2009-05-09.html

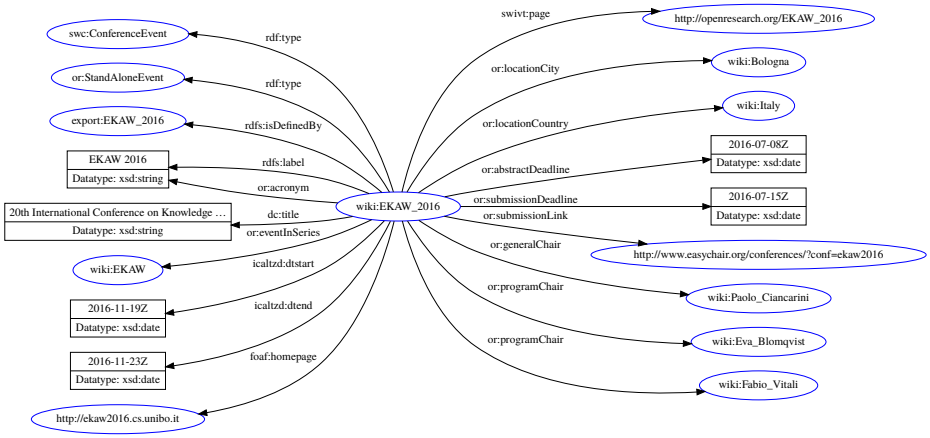


Figure 1: An exemplary usage of the OpenResearch data model showing the EKAW 2016 resource.

Web Portal Ontology (SWPO)[21], and the *Funding, Research Administration and Projects Ontology* (FRAPO)[22], as well as schema.org. The SWC, SWPO and schema.org vocabularies provide means for modeling general events and SWC and SWPO also conferences. FRAPO provides terms to express scientific projects and their relations. *Conference Linked Data* (COLINDA)[23] contains information about scientific events collected from other systems such as WikiCFP and EventSeer and published as Linked Data, and the CfP ontology[24] provides means for modeling calls. A specific ontology for CfPs has been proposed in [8].

The property alignment is implemented using the SMW mechanism for importing vocabularies[25]. This includes definitions of the reused vocabularies in special vocabulary pages e.g. for SWC[26], which lists all imported properties and annotates them with SMW data types for the values. Wiki categories and properties are then aligned with the vocabulary terms using special *imported from* links. For instance *Category:Conference* is aligned to *swc:ConferenceEvent* with `[[imported from::swc:ConferenceEvent]]`. For modeling the calls and roles for a conference we defined new properties in our own vocabulary[27]. Fig. 1[28] provides an example for using the data model. In contrast to the existing data model for calls and roles in the SWC ontology we are following a flat structure, which

[21] `http://sw-portal.deri.org/ontologies/swportal`

[22] `http://www.sparontologies.net/ontologies/frapo/source.html`

[23] `http://colinda.org`, offline at the time of this writing

[24] `http://sw.deri.org/2005/08/conf/cfp` referenced by SWC, offline at time of writing

[25] `https://www.semantic-mediawiki.org/wiki/Help:Import_vocabulary`

[26] `http://openresearch.org/MediaWiki:Smw_import_swc`

[27] `http://openresearch.org/vocab/`

[28] Besides the usual prefix mappings that are available at `http://prefix.cc/`, we also use `wiki: http://openresearch.org/Special:URIResolver/` and `export: http://openresearch.org/Special:ExportRDF/`

allows users, e.g., to directly attach a deadline to an event rather than creating a new instance for a call in addition to the actual event.

5.2 Architecture

Figure 2 depicts the three layers of OpenResearch's architecture: Data gathering, Data processing and Data representation.

Data Gathering and Scrapers This layer supports ingestion, semantic lifting and integration of relevant information from various sources. To populate the OpenResearch knowledge base in addition to crowd-sourcing, we gather information from different sources. Sources can be available as Linked Data already, or structured, semi-structured and unstructured. SMW itself provides two options for importing data: creation of individual pages/resources and bulk import[29] using the MediaWiki export format. Structured and semi-structured information can be imported as CSV and RDF: CSV files, prepared manually or obtained from WikiCFP via a crawler that we have implemented[30], can be transformed to the MediaWiki export format using the MediaWiki CSV Import[31] and then imported using the bulk importer; RDF datasets can be imported using the RDFIO MediaWiki extension[32].

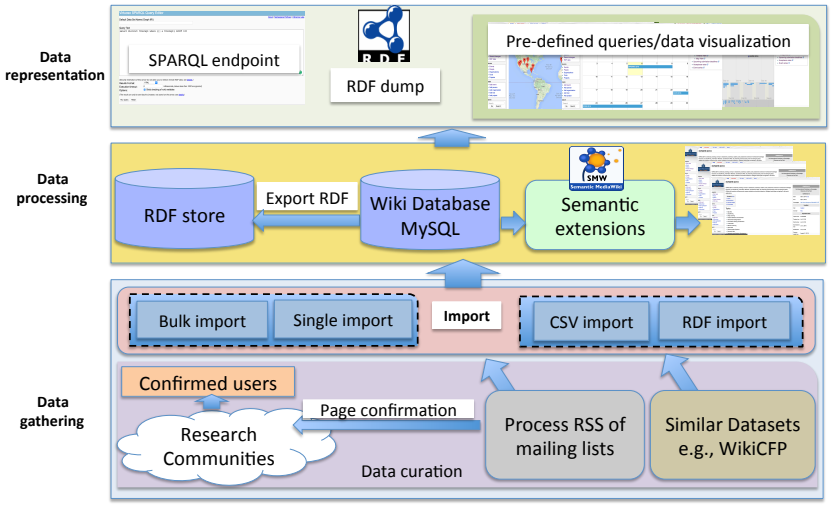


Figure 2: OpenResearch Architecture

[29] https://www.mediawiki.org/wiki/Help:Export

[30] https://github.com/EIS-Bonn/OpenResearch/tree/master/wikiCFP

[31] http://mwcsvimport.pronique.com/; usage described at http://openresearch.org/OpenResearch:HowTo

[32] https://www.mediawiki.org/wiki/Extension:RDFIO

Data processing This layer enables the storing and management of unstructured (text markup), semi-structured (annotations and infoboxes), structured data (RDF data adhering to an ontology) and schema data (the underlying ontology) Two database management systems are used in the OpenResearch architecture: one to store the schema-level information, the other to store the generated semantic triples. SMW supports multiple triple stores for storing the RDF graph, e.g., Blazegraph or Virtuoso. We use Blazegraph as it has been selected Wikimedia Foundation based on a performance and quality.[33] A MySQL relational database is used to store the templates, properties and, form names.

Data exploring This layer comprises various means for human and machine-readable consumption of the data. Several types of data representation are made possible by data exploration. CfPs are represented as individual wiki pages for each event instance, including a semantic representation of their metadata. SMW provides a full-text search facility and supports semantic queries. Queries and the visualization of their results are detailed in Section 6. Furthermore, the RDF triple store can be accessed using a SPARQL endpoint or downloadable RDF dump.

6 OpenResearch Services

On top of the basic architectural layers, OpenResearch offers services for different stakeholders of scientific communication. As a semantic wiki, it offers initial LOD services and semantic representation of metadata about events. We address the issues discussed in section 2 by establishing a set of quality metrics for scientific events and implementing them as properties. We adopt the definition of quality as *fitness for use*, which, here, means the extent to which the specification of an event satisfies its stakeholders [5, 6]. In the remainder of this section, the current services are explained in three categories: wiki pages, LOD services and queries.

Semantic Wiki Pages SMW powers OpenResearch to provide semantic representation of CfPs as one wiki page per event. In OpenResearch, specific semantic forms have been designed for each type of entities to make content creation and revision as easy as possible for users. Properties of each semantic object are populated via fields in these semantic forms. The following example shows the generated SMW wiki markup containing general information about an event. Further information about committee members, extensions and other important dates can also be provided in other parts of the form. The complete textual representation of the CfPs can also be added as content of the wiki page with embedded semantic annotations.

```
{{Event
 | Acronym = EKAW 2016
 | Title = 20th International Conference on Knowledge Engineering and Management
```

[33] https://goo.gl/NNm407

```
 | Series = EKAW                        | Type = Conference
 | Field = Knowledge Engineering
 | Start date = 2016-11-19              | End date = 2016-11-23
 | Homepage = ekaw2016.cs.unibo.it      | has Twitter = @ekaw2016
 | City = Bologna                       | Country = Italy
 | Submission deadline = 2016-07-15     | Abstract deadline = 2016-07-08
 | submission link = www.easychair.org/conferences/?conf=ekaw2016,
 | has general chair = Paolo Ciancarini,
 | has program chair = Eva Blomqvist, Fabio Vitali
}}
```

LOD Services All data created within OpenResearch is published as Linked Open Data (LOD). In the sequel, we describe ways for accessing OpenResearch LOD. Afterwards, we outlines how the LOD approach enables building further services on top by sketching two possible ways of consuming the OpenResearch LOD: interlinking with relevant datasets, and using OpenResearch LOD as external plug-in for the Fidus Writer scientific authoring platform[34].

Accessing OpenResearch LOD An updated version of the OpenResearch dataset is produced daily and available for download and query.[35]. The data is also queryable via a SPARQL endpoint[36]. In addition, the semantic representation of the metadata for each event is represented as an RDF feed in each page. The RDF feed for the EKAW 2016 resource is available at `http://openresearch.org/Special:ExportRDF/EKAW_2016`. To expose dereferenceable resources conforming with Linked Data best practices, the URI resolver provides URIs with content negotiation; e.g., for the EKAW 2016 resource the URI is `http://openresearch.org/Special:URIResolver/EKAW_2016`.

Interlinking To increase the coherence of the data, we interlink the OpenResearch LOD with other relevant datasets. We are applying the same technical framework that we are using for OpenAIRE[37] Interlinking [1]. The following use cases enabled by interlinking show how the results of connecting the linked dataset of OpenResearch with other relevant datasets enhance the services:

1. *PC members recommendation:* one of the difficult and time-consuming tasks for event organizers is to collect a group of high-profile researchers as PC members. Interlinking OpenResearch LOD with datasets including author and person information such as ORCID[38] helps in this regard.
2. *Sponsoring recommendation:* it is often a challenge especially for smaller events to find local and international sponsors. On the other hand organizations and companies who want to gain visibility and decide whether or not to sponsor an event can use OpenResearch.

[34] https://www.fiduswriter.org/
[35] https://zenodo.org/record/57899
[36] http://openresearch.org/sparql
[37] https://www.openaire.eu/
[38] http://orcid.org/

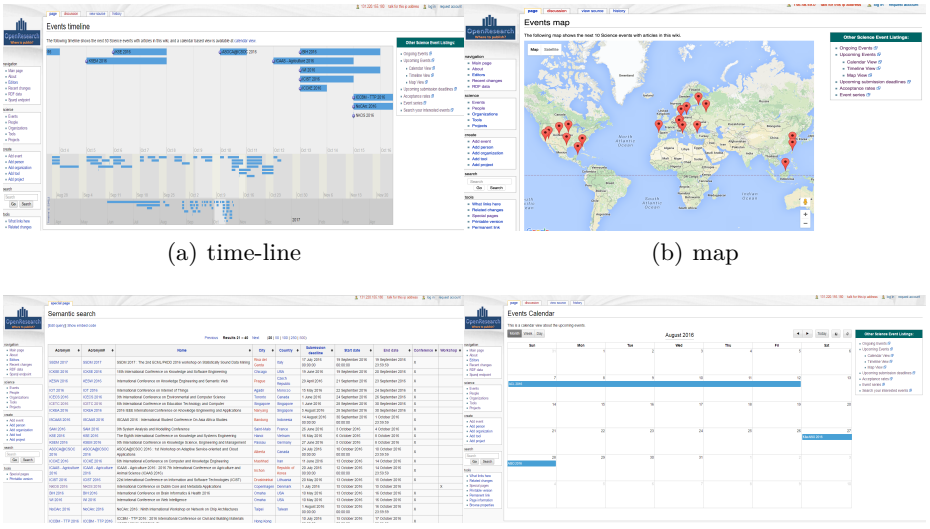
Integration with an Authoring Platforms In this section we introduce our approach to improve the workflow of authoring processing [7]. The OpenResearch LOD will be plugged into the Fidus Writer authoring platform to improve the workflow in the following use cases:

1. *Venue recommendation:* One of the critical aspects in the process of writing and publishing is to find a suitable event to submit the scientific results. The OpenResearch dataset contains data about events annotated with corresponding scientific field as *:category* and keywords. We also annotate keywords from the content of the under-production scholarly document in the OSCOSS project that could be imported to the OpenResearch search services.For example, *Find all events in the computer science field that focus on data analysis, big data, knowledge engineering, linked data.* The result of queries can be shown to the authors with a user-friendly interface and filtering metrics such as deadline and location distance.
2. *Direct link to submission pages:* The OpenResearch data contains a property named *submission link* that provides a direct link to paper submission pages of events. The submission page of the targeted event can be made accessible easily from the authoring platform.
3. *Notification services:* there are different deadlines attached to the events that should be considered by authors such as abstract deadline, submission deadline or registration deadline as well as deadline extensions. Enabling notification services in the authoring platform will support both organizers and researchers.

Queries and Visualization of Results To support the creation of various views, recommendations and ranked lists (by quality indicators), queries can be defined and executed using all defined properties and classes and the results can be embedded in wiki pages. For example, events can be ranked by acceptance rate using the corresponding properties in queries:

```
{{#ask:[[Category:Event]]
 | ?title = Name | ?Event in series = Series | ?Category | ?Acceptance rate
 | format = table | limit=10 | sort=Acceptance rate | order=desc
}}
```

It is also possible to capture the relationships between various types of entities (e.g. event series, sub/super events, roles of a person in event organization, etc.). Many popular views have been implemented in OpenResearch as pre-defined queries. Various display formats provided by SMW extensions are used to visualize the query results. Figure 3 shows a map view of the upcoming events using location-based filtering. Similarly, calendar and timeline views show upcoming submission and notification deadlines as well as the events themselves. In addition, taking, for example, participation figures into account enables new indicators for measuring the quality and relevance of research that are not just based on citation counts [2]. Based on semantically enriched indicators, predefined SPARQL queries as well as form-based search facilities will be implemented for recommendation services.



(a) time-line (b) map

(c) table (d) calendar

Figure 3: Upcoming events in different visualizations

7 Evaluation

The main objective of this work is to introduce a comprehensive approach for collaborative management of scholarly communication metadata with a special focus on events. We are for now mainly interested in collecting data, as this allows to provide more interesting analysis services.Nevertheless, we evaluated three aspects of OpenResearch including two surveys, performance measurements of the system as well as a usability analysis.

Timeliness Questionnaire: In a survey, we asked 40 researchers from different fields including Computer Science, Social Science to explain how they explore scientific events[39]. Over 75% of the participants agree that having an event recommendation service is very relevant for them. For selecting an event to participate, all participants confirmed that they consider information that is not served directly by the current communication channels. Some of these criteria are networking possibilities, review quality, high-profile organizers, keynote speakers and sponsors, low acceptance rate, having high quality co-located events, close location, citations counts for accepted papers of previous years. Participants indicated that they explore scientific events using: search engines, mailing lists, social media and personal contacts. Then, they assess the CfPs to find out whether that event satisfies their criteria. Over 85% of the participants supported the idea of using a knowledge base for this purpose.

Usability survey: We asked users to tell us about their experience wrt. the ease

[39] `https://goo.gl/L02UU5`

Objective Comparison Metrics	**Data Import**	**Complex Queries**
Time(s)	32.6	0.31
Memory (MB)	24.44	2.89
Number of pages	100	n/a
Number of queries	n/a	10

and usability of the system[40]. Overall 12 users participated in the survey; they have had several roles in scientific events (participant, PC member, event organizer and keynote speaker). 75% of the users replied they had basic knowledge about wikis in general, however, half of them did not know about SMW. 66% got familiarized easily with OpenResearch which shows its suitability for researchers of different fields. Again 66% answered that they needed less than 5 minutes to add a single event which is relatively low time wrt. the time organizers need to announce their event in several channels. The average number of single events created by individual users is 10. More than half of the participants needed less than 5 minutes for a bulk upload.The participants largely agreed that these times are reasonable.

Performance measurement: Currently, OpenResearch is running on a Debian server at the University of Bonn with 8 GB of RAM allocated. By private invitation (OpenResearch has not yet been publicly announced at a large scale), 70 users have been added during the last two months. Above 300 events have been added by the users during last two months and several bulk uploads of data are performed every week by the admins; each time 100 pages were created. The measured time for bulk import varies with the content of CfPs and reduces when events exist already in the system. The table below shows a performance measurements of OR w.r.t. the average time and memory usage for several bulk imports and complex queries running over the event query form.

8 Conclusion and Future Work

With regard to scholarly communication we are currently at a crossroad: On the one hand, there are commercial publishers and new incumbents such as social networks for researchers (e.g. ResearchGate, Academia.edu), which provide commercial services to the research community. Researchers either pay directly for these services by means of publication and access fees or indirectly (such as in the case of social networks) with their data. Either way, these commercial services strive to create a lock-in effect, which forces researchers to continue using these services without being able to migrate and choose competing services. On the other hand, there is an increasing push towards more open-access and open platforms for scholarly communication. Examples are open-access repositories such as arXiv, Zenodo, bibliographic metadata services such as DBLP and OpenAIRE, journal and conference management software and services such as Open Journal Systems and EasyChair or OpenCourseWare platforms such as SlideWiki.org. We see the work on OpenResearch presented in this article

[40] https://goo.gl/HIIeEh

as a first step towards tighter interlinking and integrating of open services for scholarly communication.

In future, we envision to intensify data flows and service integration between OpenResearch and other open scholarly services. In particular, we are planning to import information from events' web pages, mailing lists and proceedings catalogs. Crawling event's web pages and extracting, e.g., embedded structured information such as schema.org RDFa or microdata, including the *Event* class and properties such as *name*, *organizer*, *location*, *startDate*, *endDate*, *subEvent*, or *superEvent*, keeps us up to date with the organizers. Extracting information from unstructured emails is challenging, but some emails have iCalendar attachments. Further information about events and their proceedings could be scraped from semi-structured listings such as the index page of the CEUR-WS.org open access workshop proceedings. Furthermore, we plan to relate events with other entities e.g., publications, projects, datasets.

Acknowledgments. This work has been partially funded by the European Commission with a grant for the project OpenAIRE (GA no. 643410) and by a grant for the project LEDS (GA no. 03WKCG11C) from the German Federal Ministry of Education and Research (BMBF). We thank Yakun Li for his technical work on openresearch.org.

References

1. Alexiou, G. et al. OpenAIRE LOD services: Scholarly Communication Data as Linked Data. In: 2nd Workshop. SAVE-SD. LNCS. Springer, 2016.
2. Di Iorio, A. et al. Semantic Publishing Challenge. Assessing the Quality of Scientific Output by Information Extraction and Interlinking. In: SemWebEval 2015 at ESWC 2015. 548. Springer, 2015.
3. Hurtado Martin, G. et al. An Exploratory Study on Content-based Filtering of Call for Papers. In: Multidisciplinary Information Retrieval (2013).
4. Issertial, L., Tsuji, H. Information Extraction for Call for Paper. In: Int. Journal of Knowledge and Systems Science (IJKSS) 6(4) (2015).
5. Juran, J. M. Juran's Quality Control Handbook. 4th. McGraw-Hill (Tx), 1974.
6. Knight, S.-A., Burn, J. M. Developing a framework for assessing information quality on the World Wide Web. In: Informing Science: International Journal of an Emerging Transdiscipline 8(5) (2005), pp. 159–172.
7. Mayr, P., Momeni, F., Lange, C. Opening Scholarly Communication in Social Sciences: Supporting Open Peer Review with Fidus Writer. In: EA Conference. 2016.
8. Tomberg, V. et al. Towards a Comprehensive Call Ontology for Research 2.0. In: *i-KNOW*. ACM, 2011.
9. Wang, H.-D., Wu, J. Collaborative Filtering of Call for Papers. In: (7th Dec. 2015). IEEE, 2015, pp. 963–970.
10. Xia, J. et al. Optimizing Academic Conference Classification Using Social Tags. In: *CSE*. (Hong Kong, China). IEEE, 2010, pp. 289–294.