

Combining RDF Vocabularies for Expert Finding

– prefinal draft – *

Boanerges Aleman-Meza¹, Uldis Bojars², Harold Boley³, John G. Breslin²,
Malgorzata Mochol⁴, Lyndon JB Nixon⁴, Axel Polleres^{2,5}, and Anna V. Zhdanova⁵

¹ LSDIS Lab, University of Georgia, USA, ² DERI, National University of Ireland, Galway
³ University of New Brunswick and National Research Council, Canada, ⁴ Free University of
Berlin, Germany, ⁵ Universidad Rey Juan Carlos, Madrid, Spain, ⁶ University of Surrey, UK

Abstract. This paper presents a framework for the reuse and extension of existing, established vocabularies in the Semantic Web. Driven by the primary application of expert finding, we will explore the reuse of vocabularies that have attracted a considerable user community already (FOAF, SIOC, etc.) or are derived from de facto standards used in tools or industrial practice (such as vCard, iCal and Dublin Core). This focus guarantees direct applicability and low entry barriers, unlike when devising a new ontology from scratch. The Web is already populated with several vocabularies which complement each other (but also have considerable overlap) in that they cover a wide range of necessary features to adequately describe the expert finding domain. Little effort has been made so far to identify and compare existing approaches, and to devise best practices on how to use and extend various vocabularies conjointly. It is the goal of the recently started ExpertFinder initiative to fill this gap. In this paper we present the ExpertFinder framework for reuse and extension of existing vocabularies in the Semantic Web. We provide a practical analysis of overlaps and options for combined use and extensions of several existing vocabularies, as well as a proposal for applying rules and other enabling technologies to the expert finding task.

1 Introduction

The Semantic Web has arrived! A growing number of people and institutions provide metadata on their personal or institutional Webpages in vocabularies based on RDF. Microformats provide another way to embed metadata directly within XHTML documents. The GRDDL working group¹, recently founded by the W3C, provides ways to derive semantically richer RDF from the structured data of microformats. Additionally, the Semantic Web Best Practices and Deployment Working Group² is working towards providing guidelines on how to publish and syntactically combine RDF/XML or OWL data and ontologies. Furthermore, browser extensions such as Semantic Radar³ allow

* This work has been partially supported by the EU Network of Excellence Knowledge Web (FP6-507482), the Spanish MEC (projects TIC2003-9001, TIN2006-15455-CO3), the German BMBF funded project Knowledge Nets and the Science Foundation Ireland (SFI/02/CE1/I131).

¹ <http://www.w3.org/2001/sw/grddl-wg/>

² <http://www.w3.org/2001/sw/BestPractices/>

³ <http://sioc-project.org/firefox/>

the detection of RDF data on web pages, and generic RDF browsers such as Tabulator⁴ make exploration of Semantic Web data easier. While the syntactical issues are being solved by others, we will take a closer look here at the actual vocabularies and ontologies that can be used for capturing metadata about persons and organisations for publishing on the Semantic Web. We are proposing a frequently-addressed, but still challenging, key application for the take-up of Semantic Web technologies: automating the task of finding experts (individuals, teams, and organisations), which is a daunting manual effort at the moment. Our assumption is that when persons, institutions, projects, and events are described in Web pages using agreed-upon machine readable formats, the automatic location of experts/expertise in a particular area or for a particular task will become feasible. To achieve this goal, and similarly for other applications of Semantic Web search, we identify three critical success factors:

- *Common machine readable formats* (syntax and semantics) supported by a
- *critical mass of users* (low entry barrier, tool support, reuse) as well as
- *enabling technologies* in place to solve practical use cases.

To this end, the following contributions are made in the remainder of this paper: In Sec. 2 we separately discuss each of these critical success factors. Sec. 3 describes the ExpertFinder initiative and we identify specific use cases which shall be covered within our chosen application domain. Sec. 4 contains the core of this paper, namely, we identify several quasi-standard ontologies relevant to expert finding, their relations to our domain, and propose how to combine them in the ExpertFinder Vocabulary Framework. We discuss related initiatives and approaches in Sec. 5 and conclude in Sec. 6.

2 Critical Success Factors

Common machine readable formats As mentioned before, we will consider syntactical issues to be solved for the moment and assume that people know how to publish semantic annotations with their webpages and that there is proper tool support. We therefore focus on semantic aspects, by which we mean common existing ontologies and RDF vocabularies that can be used. The vocabularies that need to be considered comprise areas such as descriptions of personal and institutional data (including curriculum vitae and addresses), actual ontologies for modeling areas of knowledge/expertise, business sectors and communities, events, and publications. In section 4 we provide an analysis of existing vocabularies, ontologies and business standards for each of these fields.

Critical mass of users Independent of whichever vocabulary we finally decide on, in order to really enable expert finding on a (Semantic) Web scale we have to either (i) convince a critical mass of users and content publishers to support the chosen vocabulary, (ii) translate/import existing content automatically into the chosen vocabulary, or (iii) provide mechanisms within the chosen vocabulary that can facilitate reuse of the vocabularies already utilised on the Web (e.g. using `owl:equivalentClass` and `owl:equivalentProperty`). As for (i), creating a new ontology from scratch and disseminating it within a closed community is difficult. Within the wider Web community we can expect that enforcing the take-up of a single vocabulary is not only unlikely,

⁴ <http://www.w3.org/2005/ajar/tab>

but probably even impossible. Individual users and organisations will choose portions of ontologies, add extensions for their own purposes and will use different URIs to describe the same or similar concepts. In the context of (ii), text retrieval methods or wrapper technologies facilitated by approaches like PiggyBank [13] have become more stable and successful over the last few years. However, they still do not offer 100% precision or recall depending on how structured the underlying data source is, nor solve the problem of the right ontology/vocabulary to use for the generated annotations. Hence, there is nothing that can be called *the* right ontology for our domain and we will therefore focus on (iii) and try to reuse and effectively combine all existing and actually *used* formats. For this purpose, we can analyse and formally define their overlaps, and provide best practices on how to apply them together. This is the approach we focus on in the ExpertFinder initiative. A practical side effect of reusing and extending existing vocabularies is that we can rely on existing tools. For example, iCal [8], vCard [22], or BibTeX [18] provide vocabularies that are supported by tools (such as calendars, address book software) and online citation indexes, and are used already as de facto standards for data exchange.

Enabling technologies The previous two success factors were concerned with how to get the necessary metadata on the Web. In order to solve practical use cases, such as the ones listed in Sec. 3, we have to consider several additional technologies. For example, recommendation algorithms, rules, strategies, collaborative filtering techniques, statistical methods, etc. Such methods will help to rate the value of metadata, but also allow support in annotation tools and search engines in order to find related ontological terms. Moreover, security and trust mechanisms allow restricting access to certain metadata by encryption or authentication and are gaining importance as the Web gets populated with personal information which should not be public. As a final key enabling technology, rules will play a crucial role in several respects. First, rules (together with expressive ontology languages) allow us to formally define the exact relationships between the existing vocabularies. For example, some vocabularies (e.g., FOAF, SIOC) already formalise their structure to some degree in OWL but usually do not incorporate many more features beyond simple taxonomies expressible in RDFS alone. When defining the exact relations between overlapping vocabularies we expect to even need expressive features beyond OWL [9]. As there is no standard yet for defining such mappings rules, one could imagine SPARQL CONSTRUCT⁵ along with built-in predicates. For instance imagine a mapping from `vCard:homeTel` to `foaf:phone` where the former is a datatype property and the latter an object property. Basically, a mapping needs a conversion function, generating a URI from the source RDF literal value:

```
CONSTRUCT { ?X foaf:phone ?T . } WHERE { ?X vCard:tel ?T1 .  
  FILTER (fn:str(?T) = fn:concat("tel:",fn:encode-for-uri(?T1))) }
```

Secondly, rules, published together with RDF metadata, can serve to “link” or define implicit metadata [25]. This would enable us to link to metadata published elsewhere involving (possibly negative) dependencies [19]. Sample rules could be for instance:

⁵ Lacking a standard language to define complex mappings as this one, we admittedly “abuse” SPARQL FILTER expressions here for data conversion with XPath 2.0 Functions, which is likely not (yet) supported by current engines.

<http://www.w3.org/People/Berners-Lee/card#i> *is an expert in* http://en.wikipedia.org/wiki/Semantic_Web⁶.

All persons listed at http://www.rdfweb.org/topic/ExpertFinder_2fMembers *but not those working for companies listed at* <http://www.myCompetitors/> *are my friends.*

http://www.w3.org/People/Berners-Lee/card#i is author of all publications listed at <http://dblp.uni-trier.de/Berners=Lee:Tim.html>

Syntactically, one possibility is again to adopt CONSTRUCT queries from SPARQL as a view/link definition language [20], but also here a dedicated standard is still missing. We expect that W3C efforts like the RDF Data Access (DAWG)⁷ and Rule Interchange Format (RIF)⁸ working groups will soon provide adequate solutions in this direction.

3 Practical Use Cases from the ExpertFinder Initiative

ExpertFinder⁹ is an international collaborative initiative with the aim of devising vocabulary and rule extensions (e.g. FOAF and SIOC) and best practices and recommendations towards standardisation in order to annotate personal homepages, pages of institutions, conferences, publication indexes, etc. with adequate metadata to enable computer agents to find experts on particular topics. The initiative was founded in 2006 with the goal to align related research efforts and to tackle precisely those critical factors outlined in Sec. 2. Among others, the following use cases were identified as potential early adopters of properly aligned Semantic Web data and vocabularies.

Automatic generation of institutional and personal webpages from metadata and rules RDF metadata itself is an excellent source for content management systems separating content from any layout related issues, but in a more flexible manner than current solutions. Members of institutions could be allowed to provide their own metadata as extended FOAF files, but, if missing, the institution itself could specify standard policies for generating implicit metadata by means of default rules. Such rules could allow to aggregate metadata from some 3rd party sources. For instance, imagine that your office colleague is too lazy to generate his own homepage/metadata file. No problem: basic data can be aggregated from metadata available at the university personnel database, a default publication list can be generated by the metadata extracted from DBLP, and so on, by means of rules such as the ones in Sec. 2 or other techniques. By relying on common metadata formats natively, exporting, querying and combining information aggregated from different sources into annotated pages becomes trivial. Preliminary examples of this use of FOAF and other metadata is already proposed in [10].

Human Resource Management People searching for a job could publish their CV and profiles as metadata on the Web or employees in a company could make their skill set and experiences available on the intranet in agreed metadata formats. Job agencies can deploy agents which they feed with their preferred profile crawling the web to identify suitable candidates for a given vacancy, or vice versa job vacancies could be published

⁶ Linking to Wikipedia terms is one of many possibilities to give areas of expertise an identifier

⁷ <http://www.w3.org/2001/sw/DataAccess/>

⁸ <http://www.w3.org/2005/rules/>

⁹ <http://www.rdfweb.org/topic/ExpertFinder>

in the same formats. Team building within companies can be partly automated by selecting the right set of employees to successfully complete a given project through semantic matching and rules. ExpertFinder shall enable such scenarios and decentralise the process of expert and job finding, as opposed to current central recruitment or corporate portals, just as FOAF itself was aimed to decentralise social networks.

Public Semantic Research Portals The ExpertFinder idea is fruitfully applicable to R+D community portals such as EU's successful CORDIS ¹⁰ which enable institutions to find and contact each other for joint research projects. Semantic enrichment of such portals may enable refined search down to the level of individual researchers, or allow decentralised publication by the institutions themselves. In the resulting scenario additional requirements on assessing trustworthy information arise: Instead of providing central portals, public bodies like the EU could assess/certify published content.

Semantic Reviewer Selection In the academic realm, finding good reviewers remains a daunting task. However, many publications already provide pre-classified keywords such as ACM categories. Now, using citation indexes, committees of previous conferences etc. published in the agreed metadata format, one could define in a declarative rule language (possibly with priorities) some selection criteria to find appropriate expert reviewers, or adapt the selection criteria of previous workshops, if it is published by the organisers. Using common agreed vocabularies for categories, publications etc. mock-up examples using a combination of declarative rules and OWL such as presented at [9] could become a practical reality.

Trust and security for privacy-relevant meta-data In all of the above scenarios it is desirable that (parts of) metadata can be protected, for instance by provision of time-restricted keys for decryption during a process of rule based negotiation (cf. [5] for pointers). Rules, similar to the ones mentioned above, can guide such negotiation. For example, the person who wants my phone number needs to invoke a service to get my phone number where all persons I know are registered. In the simplest case the service could work per email and check the sha1-sum of people in my FOAF file and send a mail back to that address with a (temporarily valid) decryption key for an encrypted telephone number also provided in my FOAF file. For different versions of this scenario, with different credentials, more involved negotiation processes are imaginable.

4 The ExpertFinder Vocabulary Framework

Instead of proposing a new ontology for tackling the challenges of semantic expert finding we rather suggest a framework of existing vocabularies which shall be fruitfully combined. As shown in Fig. 1 we identify the following “components” to describe experts (persons, organisations or communities):

- *general descriptions* of persons, communities and organisations,
- *relations* between persons, communities and organisations,
- *educational aspects*,
- past and present *activities and projects*, and
- *skills*

¹⁰ <http://cordis.europa.eu/en/home.html>

Additional fields not uniquely connected to particular persons or organisations we want to cover are *events* and *publications*, opinions and ratings, endorsements as well as recommendations and references. Our goal is to pick some of the most widely used vocabularies in these areas, check how far they are formalised, identify what overlaps exist between these formats and how they can be reused and combined¹¹.

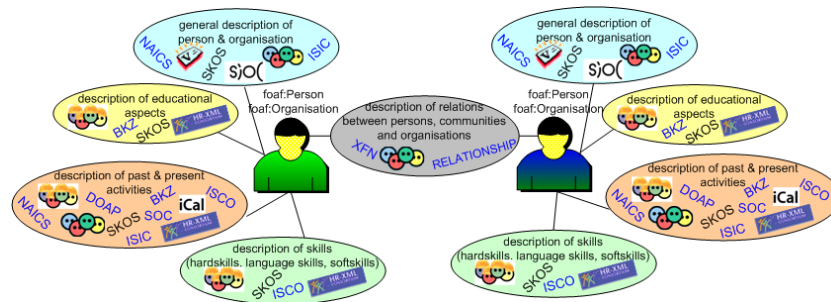


Fig. 1. How to describe an expert?

4.1 Starting points: FOAF, SIOC & SKOS

The Friend of a Friend (FOAF) [7] and Semantically-Interlinked Online Communities (SIOC) [6] ontologies mark the starting points of our work as, on one hand, they already cover the description of much of the "components" mentioned above, and, on the other hand, they are being adopted by a steadily increasing user community.

The FOAF ontology was developed to create machine readable information/metadata for people, groups, organisations and other related concepts - basically, to describe people, what they do and how they interact with each other. One of the most used properties of the FOAF ontology is the "knows" property: a simple way to create social networks through the addition of knows relationships for each individual that a person knows. Aggregations of FOAF data from many individual homepages are creating distributed social networks; this can in turn be connected to FOAF data from larger online social networking sites such as LiveJournal¹² or Tribe.

In terms of definitions of expertise the FOAF ontology has a number of relevant properties, e.g. (i) the `foaf:interest` property defines topics of interest to a person, and can be used directly to find those with an interest a particular domain (e.g. `foaf:interest` has been used to match music preferences¹³), (ii) people can create `foaf:publications` or other `foaf:Documents` (via `foaf:made/maker`) which may have an associated `foaf:topic` or `foaf:primaryTopic` that can again be used to determine a person's domains of interest, and (iii) `foaf:currentProject/pastProject` gives information on "some collaborative or individual undertaking" that a person may be (or have been) involved in.

¹¹ We do not aim at providing an exhaustive list of all ontologies developed in all related areas

¹² <http://rdfweb.org/topic/LiveJournal>

¹³ <http://foafing-the-music.iaa.upf.edu/>

There have been a number of extensions or modules for the FOAF ontology that are of interest to the expert finding scenarios previously mentioned. FOAFRealm [15] is a user profile management system based on FOAF that provides authentication, access control and social networking features such as “semantic social collaborative filtering”. The system allows users to share and annotate their personal taxonomies across a social network using WordNet, DDC¹⁴ and DMoz¹⁵ as base classifications. When implemented in document exchange systems such as JeromeDL¹⁶, a semantic digital library, users can classify their documents or bookmarks and allow others to access these resources using FOAFRealm’s ACL-based social networking functionality. Each user’s collection is assigned an expertise value that reflects the quality of the information that they provide; this value is calculated based on a PageRank calculation of their social network. Users are then also aware of the expertise level of others on given topics.

The SIOC project¹⁷, founded by one of the authors, aims to provide a framework for the connection and interchange of information from internet-based discussions and community portals. Such communities are primarily made up of users, the posts that they create, and the discussion forums that they subscribe to across a multitude of sites and discussion platforms. The basis for SIOC is the SIOC ontology, an RDF-based schema which describes the main concepts found in online communities [6]. While there are many classes and properties in SIOC, the main notion is that `sIOC:Users` create `sIOC:Posts` that are contained in `sIOC:Forums` that are hosted on `sIOC:Sites`. With respect to finding experts in a social network, the main SIOC property of interest is `sIOC:topic` defining a resource that a particular discussion post is related to; by aggregating all the `sIOC:topics` that are associated with a particular user’s posts across a number of sites, a picture emerges as to where their topics of interest and related expertises lie. `sIOC:Forums` or sites may also have associated `sIOC:topics`, and a user with an interest in a particular topic may be a `sIOC:subscriber_of` a certain discussion channel.

The Simple Knowledge Organisation System (SKOS) [16] completes the base we want to build on. It allows to describe general terms and concepts and define many useful properties of such terms such as declaring whether a concept is broader/narrower than another, preferred and alternative labels in multiple languages for terms, as well as related terms. SKOS facilitates sharing and representing terminologies that may not extensively require the expressive power of other languages such as OWL and where a strict hierarchy such as definable by `rdfs:subclass` cannot be imposed. In the context of ExpertFinder, we can view SKOS as the basis to define and relate skills, areas of expertise/interest (via the `foaf:interest` property) or topics people discuss in online communities described by SIOC.

The SIOC ontology developers have worked with the authors of FOAF and SKOS to align concepts and avoid any unnecessary duplication or term conflicts. The concept of `sIOC:User` has been defined to be a sub-type of `foaf:onlineAccount`, so that existing properties from FOAF can be reused and so that new properties for users

¹⁴ <http://www.oclc.org/dewey/>

¹⁵ <http://dmoz.de/>

¹⁶ <http://www.jeromedl.org/>

¹⁷ <http://sIOC-project.org/>

can be defined in SIOC without directly impacting on the FOAF ontology. As shown in Fig. 2, a `foaf:Person` can own many `sio:User` profiles (via the `foaf:holdsOnlineAccount` relationship). Similarly, content that a `sio:User` creates on a particular Forum (e.g., a Weblog, Mailing List, Bulletin Board) can be linked using `sio:topic` to a `skos:Concept` (e.g., in Fig. 2 one post is talking about clouds and another post is referring to a narrower concept, that of rain clouds). Using SKOS to define topics under discussion and of interest combined with additional rule extensions, which we plan as a next step in the ExpertFinder framework, facilitates flexible definitions of relationships between the various skills formalised using SKOS concepts.

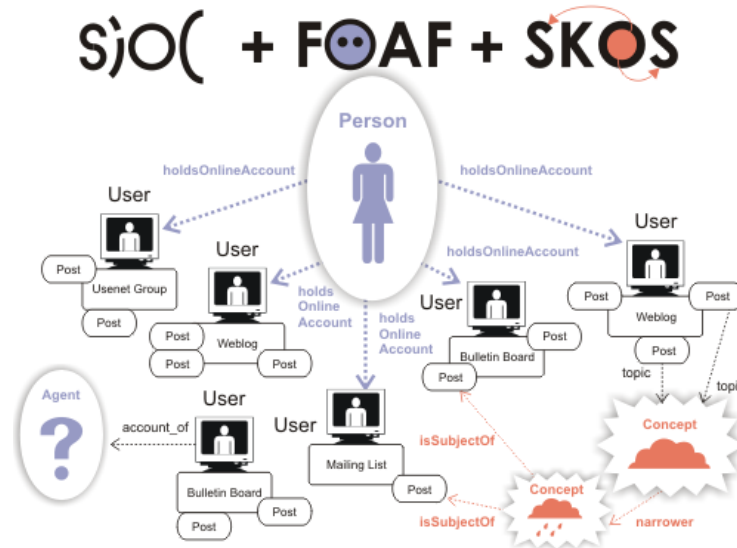


Fig. 2. Connections between SIOC, FOAF and SKOS

4.2 ExpertFinder Framework Extensions for the Core Vocabularies

FOAF, SIOC and SKOS largely cover general descriptions of, as well as relations between persons, communities and organisations. However, still some pieces are missing in order to obtain a complete picture of Fig. 1:

- FOAF misses detailed information about address data, but complementary standards such as vCard close this gap.
- The only relation between persons in FOAF is `foaf:knows`, but as we want to support more fine-grained relations, we propose the RELATIONSHIP & XFN vocabularies to close this gap.
- Projects can be linked to persons and groups by `foaf:currentProject` and `foaf:pastProject`. However, this might again be too coarse-grained if e.g. we want to know the exact timing of a project and we want to provide guidelines on how to annotate projects. We propose the use of DOAP here.
- More detailed CV information can be provided by the DOAC vocabulary.
- In the scientific context, publications are an important measure of expertise. These can be linked by the `foaf:maker` and `foaf:publication` attributes to the

`foaf:Document` class, but details on how to describe publications are missing. A de facto standard in the scientific community in this area is BibTeX.

- As for describing skills and topics of interest, SKOS defines a general framework, but details of concrete classifications to use for annotation are missing.
- Finally, events and their participants are not yet describable in a sufficient manner. iCal as a de facto standard for sharing events is a natural candidate to fill this gap.

A preliminary list of mappings for overlapping concepts and attributes in the mentioned vocabularies is omitted here for lack of space, but available at http://www.rdfweb.org/topic/ExpertFinder_2fmappings

Refining Personal Data: vCard vCard is a standard for representing personal data such as business cards. Although there are various forms in which vCard data can be written, our interest is on the RDF-based representation¹⁸. Contact information, such as phone numbers or email-addresses can be expressed more fine-grained in vCard than in FOAF by means of distinguishing properties such as `vCard:homeTel` and `vCard:workTel`. Note, however, that vCard phone numbers are not directly mappable to `foaf:phone` as a subclass, as vCard uses RDF literal values whereas foaf uses URIs using the fully qualified `tel:` scheme. A workaround we propose is to adopt FOAF's representation in vCard/RDF and make the respective properties subproperties of `foaf:phone`, or otherwise define straightforward mapping rules for conversion. Such mappings can not necessarily be bidirectional, e.g., `vCard:email` may not be simply mapped to `foaf:mbox` as a `foaf:mbox` is supposed to be unique for a person, which is not necessarily the case for vCard. Affiliation information or role information in vCard can indicate knowledge areas or particular expertise aspects, which again should be linkable to SKOS concepts.

Refining Relations: RELATIONSHIP & XFN RELATIONSHIP¹⁹ and XFN²⁰ (XHTML Friends Network) are two vocabularies used for describing interpersonal relationships. Since `foaf:knows` describes relationships between people rather sketchily, these vocabularies are deployed to fill the gap and assert such relationships in more detail by defining different subproperties.

Refining Project Descriptions: Description of a Project (DOAP) DOAP²¹ is an XML/RDF vocabulary mainly conceived to describe open source projects. Its initial goals include: (i) internationalisable description of a software project and its resources; (ii) data exchange between software directories; (iii) automatic configuration for resources such as shared CVS repositories; (iv) interoperability with other popular Web metadata projects (RSS, FOAF, DC) and (v) the ability to extend the vocabulary for specialist purposes. DOAP describes the current state of a project but it does not highlight changes and updates. Nevertheless, to keep the repository up to date with releases, the CodeZoo with an Atom²² feed containing embedded DOAP can be used. Nevertheless, even if a feed to keep older versions can be used for DOAP a way to transform the information into RDF and to distinguish between current release and past releases is still needed.

¹⁸ <http://www.w3.org/TR/vcard-rdf>, <http://www.w3.org/2006/vcard/ns>

¹⁹ <http://vocab.org/relationship/>

²⁰ <http://gmpg.org/xfn/join>

²¹ <http://usefulinc.com/doap/>

²² http://www.codezoo.com/about/doap_over_atom.csp

DOAP uses `foaf:Person` to describe the corresponding contributors of each project part (e.g. project maintainer, developer). From the FOAF side, the use of `foaf:project`, `foaf:currentProject` and `foaf:pastProject` properties do not really allow to define the duration of participation in a certain project. Neither are project durations definable properly in DOAP alone. We suggest to remedy this problem by either adding new attributes for start and end (possibly subclassing iCal events) or using temporal RDF as mentioned below.

Refining CV information: Description of a Career (DOAC), Resume RDF Schema and BIO Vocabulary: DOAC²³ is a RDF metadata vocabulary to describe professional capabilities of workers gleaned for example from CVs or resumes. The metadata enhances specific description and facilitate the search to locate suitable (regarding the given position requirements) job candidates. DOAC has been designed to be compatible with the European CV (known as Europass), which can be generated from a FOAF+DOAC file. It includes information about education, work experience, publications, spoken languages and other skills that can be shared and processed by applications. As an alternative, one of the authors [4] proposes the Resume RDF schema²⁴ for extending FOAF profiles with curriculum vitae information. This schema includes terms for work and academic experience, skills, courses and certifications, publications, references, etc. Again, we propose to link to SKOS concepts to describe the respective concepts. BIO²⁵ describes biographical information about (living and dead) people and has been designed to be compatible with both RDF and non-RDF XML formats.

DOAC uses the `foaf:Person` class to general descriptions of job seekers and `foaf:Organisation` to define which schools and institutions the individual attended. Furthermore, the `foaf:pastProject` concept could be added as a subclass to the `doac:Experience` class. This would allow description of not only a job seeker's general experiences in a company but also their experiences in different projects. Next, the `doac:publication` property which establishes a connection between the `foaf:Person` and `doac:Publication` can be defined as a `foaf:publication` linking `foaf:Person` with `foaf:Document`.

Refining Bibliographic Descriptions: BibTeX, DC and others BibTeX was designed by Patashnik and Lamport in 1985 as the LaTeX bibliographic format [18] and has established itself as a de facto standard format for publishing bibliographic information in several online citation indices. Several RDF versions of BibTeX exist, e.g., `bibtex2rdf`²⁶ or `bib2rdf`²⁷ where the former seems to be wider adopted and also reuses existing formats in the same spirit as ExpertFinder. This could be directly adopted or combined with more comprehensive ontologies for digital libraries such as MarcOnt²⁸. As MarcOnt also allows to import/export BibTeX, we currently suggest the RDF vocabulary supported by `bibtex2rdf` (cf. [10]). The Dublin Core (DC) initiative, started 10 years ago

²³ <http://ramonantonio.net/doac/>

²⁴ <http://purl.org/captsolo/semweb>

²⁵ <http://vocab.org/bio/0.1/>

²⁶ <http://www.l3s.de/~siberski/bibtex2rdf/>

²⁷ <http://www.cs.vu.nl/~mcaklein/bib2rdf/>

²⁸ <http://www.marcont.org/>

by librarians in order to provide a metadata standard for describing documents, may be viewed as a subset of bibtex, and is actually reused by bibtex2rdf.

Classifications & Standards for Skills and Topics In the following we describe a few selected standards and classifications of occupations, competencies and economic activities as possible schemes which could be used for defining skills and topics of interest. Some of these standards are used for example as an instrument for assembling and presenting statistics of education/training on national as well as international level, some others are developed for fostering international comparability of data in studying economic phenomena. While the national and international types of classifications are used for example by federal agencies for education and training statistics, the international standards should also facilitate international communication.

The *Occupational Classification (SOC)*²⁹ system is used by federal US statistical agencies to classify workers into over 820 occupational categories, grouped into 23 major groups, 96 minor groups, and 449 broad occupations. Each broad occupation includes detailed occupation(s) requiring similar job duties, skills, education, or experience. The *Profession Reference Number Classification (BKZ)*³⁰ is a German version SOC, detailing 5597 occupations. The *International Standard Classification of Occupations (ISCO-88)* is developed to facilitate international communication regarding occupations and occupational groups. Persons are classified by occupation through their relationship to a past, present or future job. The *International Standard Industrial Classification of All Economic Activities (ISIC)*³¹ is a standard classification of economic activities arranged to classify entities according to the type of activity they carry out. *North American Industry Classification System (NAICS)*³² provides common industry definitions for Canada, Mexico, and the US to facilitate economic analyses. Further standards to classify products and services like *eCl@ss*, *eOTD*, or the *RosettaNet Technical Dictionary*, or *UNSPSC* could also partly serve to describe skills and topics. All these are however, with few exceptions such as e.g. eCl@ssOWL [12], not (yet?) available in “ontologised” versions. Mappings of concepts into SKOS terms is an open issue on our agenda. A key issue here is to assign proper URIs usable in RDF to these concepts:

However, apart from special suitable classification systems which yet need to be “webised”, a simpler possibility to define topics which we already used in an example of Sec. 2, was simply referring to an online encyclopedia such as Wikipedia. Recent efforts towards semantically structuring wikis (cf. [14]) support such an approach. For some smaller domains such as e.g. computer science, ACM categories³³ or the WWW Conference Archive areas³⁴ already provide URI-addressable categories of topics, usable as SKOS terms, e.g. using the latter, we refine the example of Sec. 2 to:

http://www.w3.org/People/Berners-Lee/card#i_is_an_expert_in <http://wwwconf.ecs.soton.ac.uk/view/subjects/2004-C4.html>

²⁹ <http://www.bls.gov/soc/>

³⁰ www.arbeitsamt.de/hst/markt/news/BKZ_alpha.txt

³¹ <http://unstats.un.org/unsd/cr/registry/regcst.asp?Cl=17&Lg=1>

³² <http://www.census.gov/epcd/www/naics.html>

³³ <http://www.acm.org/class/1998/>

³⁴ <http://wwwconf.ecs.soton.ac.uk/view/subjects/subjects.html>

where we can further add:

<http://wwwconf.ecs.soton.ac.uk/view/subjects/2004-C4.html> is skos:narrower than http://en.wikipedia.org/wiki/Semantic_Web.

Events and temporal information iCal [8] as a de facto standard for calendar information supported by many applications is a natural starting point for ExpertFinder to refer to events. RDF formats and conversion tools for iCal are available³⁵. Still, iCal alone might not be sufficient to denote e.g. the validity duration of certain RDF information such as participation duration in a project. Exploring the use of RDF extensions by temporal information [11] to express the validity duration of triples would be an interesting option, but standardisation of such extensions does not yet seem to be in sight.

5 Related Projects, Initiatives and Approaches

Many other projects and initiatives overlap with or are relevant to the ExpertFinder initiative. As an umbrella initiative involving several organisations, some of these efforts are continued among the ExpertFinder participants and results will be exchanged both from these to ExpertFinder and vice versa, giving ExpertFinder the opportunity to already impact through its work into present activities as well as being open to these activities to impact upon the broader ExpertFinder efforts.

5.1 Related projects

Several projects in the Semantic Web realm have already created their own ontologies for describing persons, organisations and activities. For instance, the *KnowledgeWeb platform ontologies*³⁶, the *AKT portal ontology*³⁷, the *SWRC portal ontology*³⁸, and the *DERI Semantic Web Portal (SWP) working group's ontology*³⁹ cover many aspects of the expert finding domain, and could thus be arguably seen as equally valid starting points. However, so far these approaches seem to have experienced little take-up outside the projects where they have been developed and were developed from scratch rather than being based in pre-existing vocabularies or de facto standards. Only the DERI SW-portal ontology reuses FOAF, RSS and BibTeX to some extent, while such reuse is a central rationale in our approach.

Some related projects conducted by ExpertFinder initiative members have already followed this rationale, like Knowledge Nets, SemDis, FindXpRT and SIOC.

The *Knowledge Nets*⁴⁰ project explores the potential of Semantic Web from a business and a technical viewpoint by means of pre-selected use case scenarios. For this purpose, a prototype for the e-Recruitment domain containing the online job seeking and job procurement processes has been developed [2, 17, 21]. The requirements analysis

³⁵ <http://www.w3.org/2002/12/cal/>, <http://www.kanzaki.com/courier/ical2rdf> or <http://torrez.us/ics2rdf/>

³⁶ <http://knowledgeweb.semanticweb.org/semanticportal/OWL/>

³⁷ <http://www.aktors.org/ontology/portal>

³⁸ <http://swrc.ontoware.org/ontology>

³⁹ <http://sw-portal.deri.org/ontologies/swportal>

⁴⁰ <http://wissensnetze.ag-nbi.de>

revealed the necessity of aligning with commonly used domain standards and classifications (SOC, BKZ, WZ2003, NAICS, HR-XML⁴¹ and Skill Ontology) in order to integrate job seeker profiles and job postings as well as to support common practices from industry. Reusing these standards, the HR-ontology contributes to the realisation of more powerful and flexible e-Recruitment solutions which include advanced search and presentation facilities based on knowledge about the application domain.

The *SemDis* project addresses development of query/discovery techniques for semantic relationships. For example, `dblp:co-authorship` and `foaf:knows` were used to detect possible conflicts of interest between reviewers and authors in a peer-review process [1]. An extension on such work aims at determining possible reviewers by comparing their expertise to the topics of a paper. The expertise of a person on different topics or areas is described using the SwetoDblp dataset⁴², a large populated ontology of computer science publications based on the DBLP⁴³ bibliography database.

The *FindXpRT* (Find an eXpert via Rules and Taxonomies) [25] focuses on the important aspect of rules by combining FOAF facts and RuleML [3] rules. This implemented system⁴⁴ allows users to derive FOAF data by deploying person-centric rules, either before FOAF publication or, on demand, from published (RuleML FOAF) pages.

Finally, the SIOC initiative, which forms a central part of the ExpertFinder framework is being developed by and in collaboration with initiative members with cross-fertilizing effects between the two initiatives.

5.2 Community-Driven Approaches

In this paper we focused on the specific domain of expert finding and explored “established” vocabularies in this domain. Other Web and Semantic Web application areas show the dynamics and need for alignment even more drastically: A recent trend in many popular non-academic portals is to allow communities to create their own vocabularies and tag the items/information they want to share with others with arbitrary tags from their vocabularies: The *del.icio.us* portal⁴⁵ allows communities to tag and share their bookmarks, and search others bookmarks on the basis of these tags. The *43Things* and *43Places* community portals⁴⁶ allow describing and sharing by community-created tags information about the things people do and places they travel or want to travel. *flickr*⁴⁷ allows members to share, search and tag photos, again with arbitrary tags. *GoogleBase*⁴⁸ is a community application which allows Web users to share and search arbitrary items (pictures, text, ads, web-sites) and annotate these items using arbitrary attribute-value pairs. Most popular/shared attributes and attribute values come up in the upper level of Google search interfaces and are proposed to be used for searching and browsing the available items.

⁴¹ Developed by the HR-XML Consortium, <http://www.hr-xml.org>

⁴² <http://lsdis.cs.uga.edu/projects/semdis/swetodblp/>

⁴³ <http://www.informatik.uni-trier.de/~ley/db/>

⁴⁴ <http://www.ruleml.org/usecases/foaf/JieLiMCSThesis.pdf>

⁴⁵ <http://del.icio.us>

⁴⁶ <http://www.43things.com> and <http://www.43places.com>

⁴⁷ <http://www.flickr.com>

⁴⁸ <http://base.google.com>

None of these portals is based directly on Semantic Web technologies. However, the offered functionality is reminiscent significantly of earlier academic proposals in the Semantic Web realm, e.g. the People's Portal [23]. The examples reveal a trend of the Web becoming more structured and annotated in a community-driven manner via social processes and contributions of Web users. Reuse and adoption of already existing broadly used formats as we propose here could accelerate this process on the one hand, and on the other hand extensions for existing vocabularies could be developed in a community-driven process.

A common problem in completely unguided community-driven approaches is that entities and tags are different, yet semantically similar. This tendency brings difficulties for the community members in reuse of the community-contributed knowledge contained in the system. Defining mappings and finding an agreement on a meta-level upon which tags might become superfluous/deprecated by enforcing best practices are crucial for applications in an open web environment. Existing community-driven proposals ignore this problem to a large extent, or, in the most advanced cases, users are proposed to create ad-hoc, non-reusable alignments to achieve a specific task. Minimal support for reuse, such as auto-completion of tag names or suggestion of related search terms in annotation tools and search engines is partly supported by the above-mentioned platforms but mappings are not definable themselves in a community-driven process.

Community-driven ontology management and ontology matching extends conventional ontology matching by involving end users, knowledge engineers, and developer communities in the processes of establishing, describing and reusing vocabularies and inter-ontology mappings [24]. We believe that easy to use mapping and rules languages and tools as the next logical step. However, as we mentioned already in Sec. 2, a standard format for defining these mappings is still missing.

6 Conclusions and Outlook

We described the integration of efforts of members of the ExpertFinder initiative towards a common goal: combining commonly-agreed vocabularies including but not limited to describing information of people and their expertise, organisations, contact information, social and collaborative networks, etc. As members of this initiative, we have described various practical use cases for the task of expert finding which we identified as promising applications for actual take-up of Semantic Web technologies. We described three key success factors for bringing agreement and facilitating the take off of a joint vocabulary for expert finding. Based on this, we proposed the ExpertFinder vocabulary framework which stresses reuse and cautious extension of existing and established vocabularies in the Semantic Web. In this framework, we described how FOAF, SIOC and SKOS mark the starting points. We also discussed how to use these together extended with various existing vocabularies, pointing out the necessity for formal mappings between overlapping terms which we provide at the initiative's Web page. Along the way, we have given a survey and analysis of the related vocabularies and classifications which, although restricted to the particular domain of expert finding, we hope to be useful as such also for other related Semantic Web applications. Although we deem the core defined so far a useful start which can already be used to cover several of our

proposed use cases, we have to leave some extensions towards security, reputation and trust mechanisms (e.g., referencing endorsements or trust ontologies), which we only treated superficially so far, for future work.

References

1. B. Aleman-Meza, et al. Semantic Analytics on Social Networks: Experiences in Addressing the Problem of Conflict of Interest Detection. *15th Intl. WWW Conference 2006*, 2006.
2. C. Bizer, R. Heese, M. Mochol, R. Oldakowski, R. Tolksdorf, and R. Eckstein. The Impact of Semantic Web Technologies on Job Recruitment Processes. *7th Internationale Tagung Wirtschaftsinformatik 2005*, 2005.
3. H. Boley, S. Tabet, and G. Wagner. Design Rationale of RuleML: A Markup Language for Semantic Web Rules. *Semantic Web Working Symposium (SWWS'01)*, 2001.
4. U. Bojars. Extending FOAF with Resume Information. *1st Workshop on FOAF, Social Networks and the Semantic Web*, 2004.
5. P.A. Bonatti and D. Olmedilla. Semantic web policies: Where are we and what is still missing? Tutorial at *3rd European Semantic Web Conference (ESWC'06)*, 2006.
6. J.G. Breslin, A. Harth, U. Bojars, and S. Decker. Towards Semantically-Interlinked Online Communities. *2nd European Semantic Web Conference (ESWC'05)*, 2005.
7. D. Brickley and L. Miller. Friend of a Friend Vocabulary Specification. <http://xmlns.com/foaf/0.1/>, 2001.
8. F. Dawson and D. Stenerson. Internet Calendaring and Scheduling Core Object Specification (iCalendar). <http://www.ietf.org/rfc/rfc2445.txt>, 1998.
9. T. Eiter, G. Ianni, A. Polleres, and R. Schindlauer. Answer set programming for the semantic web. Tutorial at *3rd European Semantic Web Conference (ESWC'06)*, 2006.
10. G. AAstrand Grimnes, S. Schwarz, and L. Sauermann. RDFHomepage or “Finally, a use for your FOAF file”. *2nd Workshop on Scripting for the Semantic Web (SFSW '06)*, 2006.
11. C. Gutierrez, C. Hurtado, A. Vaisman. Temporal RDF. *2nd European Semantic Web Conference (ESWC'05)* 2005.
12. M. Hepp. Products and Services Ontologies: A Methodology for Deriving OWL Ontologies from Industrial Categorization Standards, *Intl. Journal on Semantic Web & Information Systems*, 2(1):72–99, 2006.
13. D. Huynh, S. Mazzocchi, and D. Karger. Piggy Bank: Experience the Semantic Web Inside Your Web Browser. *Intl. Semantic Web Conference 2005 (ISWC2005)*, 2005.
14. M. Krötzsch, Denny Vrandečić, and M. Völkel. Wikipedia and the Semantic Web - The Missing Links. *Proc. of WikiMania2005*, 2005.
15. S.R. Kruk and S. Decker. Semantic Social Collaborative Filtering with FOAFRealm. *Semantic Desktop Workshop colocated with Intl. Semantic Web Conference (ISWC2005)*, 2005.
16. A. Miles and D. Brickley (eds.). SKOS Core Vocabulary Specification, 2 November 2005. W3C Working Draft, <http://www.w3.org/TR/swbp-skos-core-spec>.
17. M. Mochol, R. Oldakowski, and R. Heese. Ontology-based Recruitment Process. *GI 2004*.
18. O. Patashnik. BIBTeXing, 1998.
19. A. Polleres, C. Feier, and A. Harth. Rules with contextually scoped negation. *3rd European Semantic Web Conference (ESWC'06)*, 2006.
20. A. Polleres. SPARQL Rules! Tech. Report, <http://www.polleres.net/publications/GIA-TR-2006-11-28.pdf>, 2006.
21. R. Tolksdorf, M. Mochol, R. Heese, R. Eckstein, R. Oldakowski, and C. Bizer. Semantic-Web-Technologien im Arbeitsvermittlungsprozess. *Wirtschaftsinformatik: Internetoekonomie*, 48(1):17–26, 2006.
22. A versit Consortium. vCard: The Electronic Business Card. <http://www.imc.org/pdi/vcardwhite.html>, 1997.

23. A.V. Zhdanova. An Approach to Ontology Construction and its Application to Community Portals, PhD thesis, 2006.
24. A.V. Zhdanova and P. Shvaiko. Community-Driven Ontology Matching. *3rd European Semantic Web Conference (ESWC'06)*, 2006.
25. J. Li, H. Boley, V.C. Bhavsar, and J. Mei. Expert Finding for eCollaboration Using FOAF with RuleML Rules. *2006 Conference on eTechnologies*. Montreal, Canada, 2006. .