

Combining RDF Vocabularies for Expert Finding



presented by Axel Polleres

DERI, National University of Ireland, Galway

Joint work with the **ExpertFinder Initiative**, particularly co-authors:
Boanerges Aleman-Meza, Uldis Bojars, Harold Boley, John G. Breslin,
Malgorzata Mochol, Lyndon JB Nixon, and Anna V. Zhdanova

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

Overview



- Motivation - Finding experts on the Semantic Web?
- Critical Success factors:
 - Common format
 - Critical Mass
 - Enabling Technologies (particularly Rules!)
- Practical Use Cases
- The ExpertFinder Vocabulary Framework
- Related work/what's next?

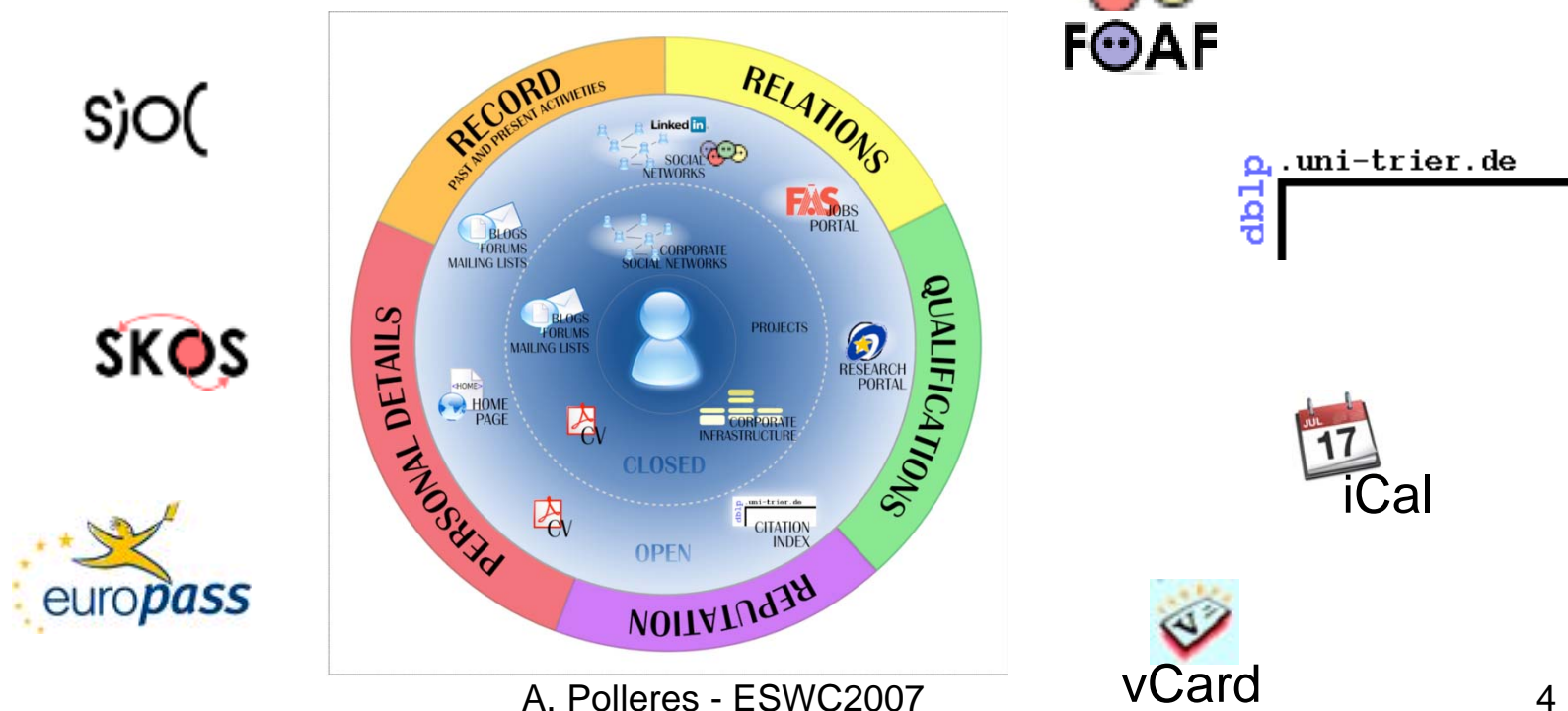


Motivation 1/2

- Goal of this paper:
 - Present initial ideas of the ExpertFinder Initiative
<http://rdfweb.org/topic/ExpertFinder>
(vision paper/application paper)
 - Sanity check of current SW ingredients to realize the vision of “finding experts online”
(position paper/survey paper)
 - Some technical details in this presentation... ;-)

Motivation 2/2

- Describing and Finding Expertise/Skills on the Web:
 - Lots of data scattered all over the Web, but already there!
 - Emerging RDF(S) formats being REALLY used, but:
 - Overlap
 - Not necessarily complete



Overview



- Motivation - Finding experts on the Semantic Web?
- Critical Success factors:
 - Common formats
 - Critical Mass
 - Enabling Technologies (particularly Rules!)
- Practical Use Cases
- The ExpertFinder Vocabulary Framework
- Related work/what's next?



Common machine readable formats

- **Common** is the keyword!
- Ontologies can only partly be “engineered”
- An ontology without agreement/community and tool support is rather a “data model”
- Existing/adopted formats have their user communities!
 - Reuse is essential
 - Won’t change fundamentally, but develop
- You don’t want to develop a new ontology from scratch, but synthesize possibly overlapping formats into a **framework of co-existing vocabularies**

From common vocabularies to critical mass 1/2



- **Existing de facto standards** (for data exchange):
 - vCard (supported by applications)
 - EuroPass (pushed by the EU)
- RDF vocabularies with **growing user communities** and **tool support**:
 - e.g. FOAF, SIOC
 - User community wasn't built in a day! Formats simple, but lots of efforts in “spreading the good news”, tool support, social factors (“geek factor”)



How to achieve critical mass

- There is not “the” right ontology
- You don’t want to develop a new ontology from scratch but modestly **extend existing vocabularies** and **synthesize** possibly **overlapping RDF formats!**

→ Formal definition of overlaps

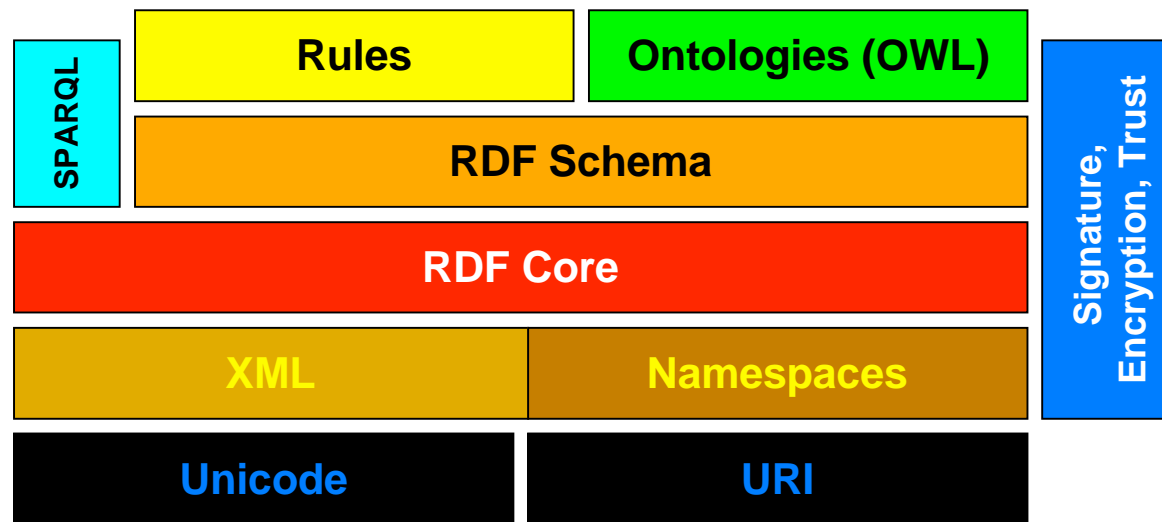
→ Best practices:

- which vocabs to use where
- how to publish (e.g. GRDDL? RDFa? RDF/XML?)

→ **Mappings back and forth** necessary!

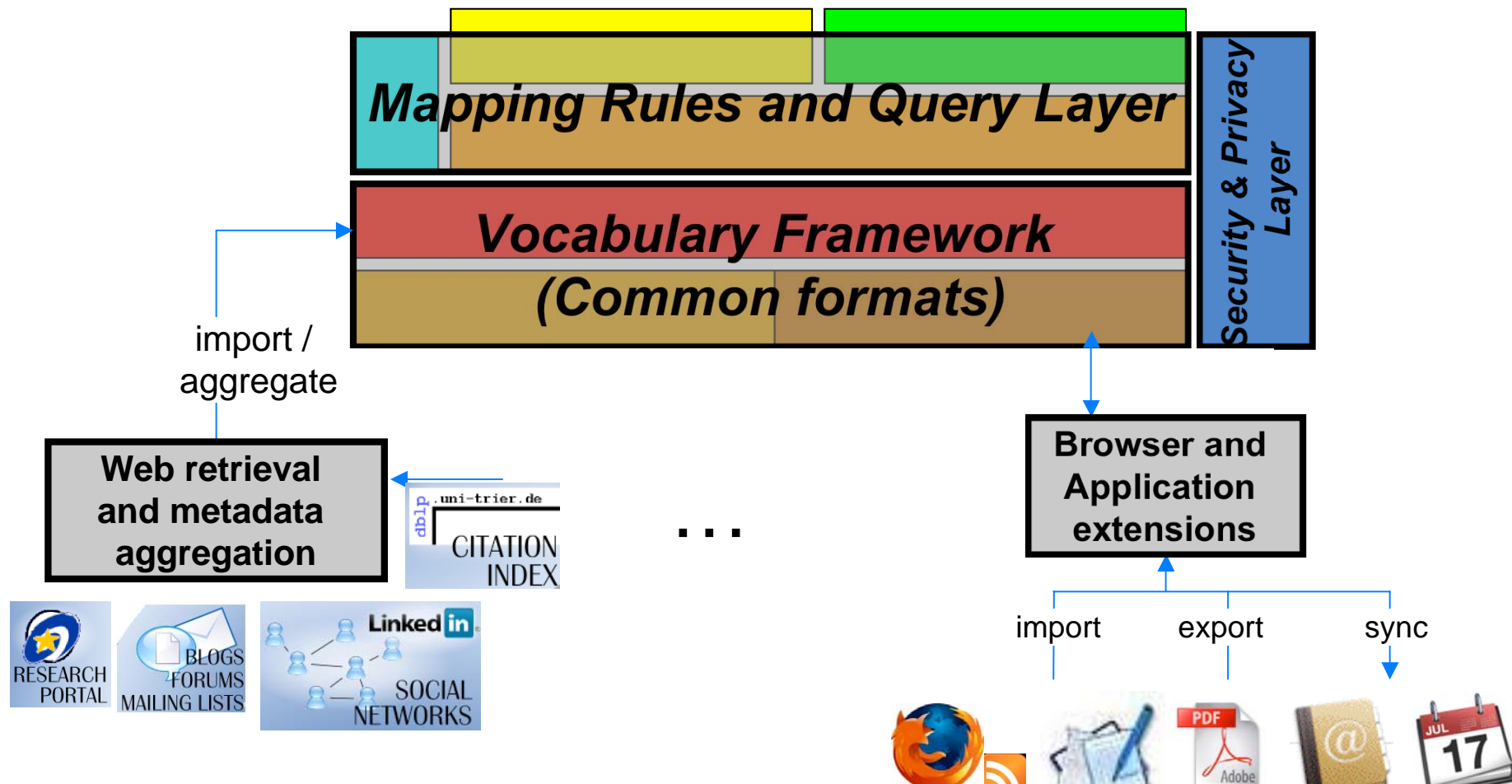
Enabling Technologies

- Where are we now?



Enabling Technologies

- Where do we want to be?



Let's talk about this one:



Mapping Rules and Query Layer

- Starting points (ie standards with a W3C stamp) for mappings:
 - Map RDF/XML-to-RDF/XML via **XSLT** ... not declarative
 - **OWL/RDFS** subclassing/subproperties: not sufficient
 - **RIF** not (yet) there
 - **SPARQL** not powerful enough

Mapping Example

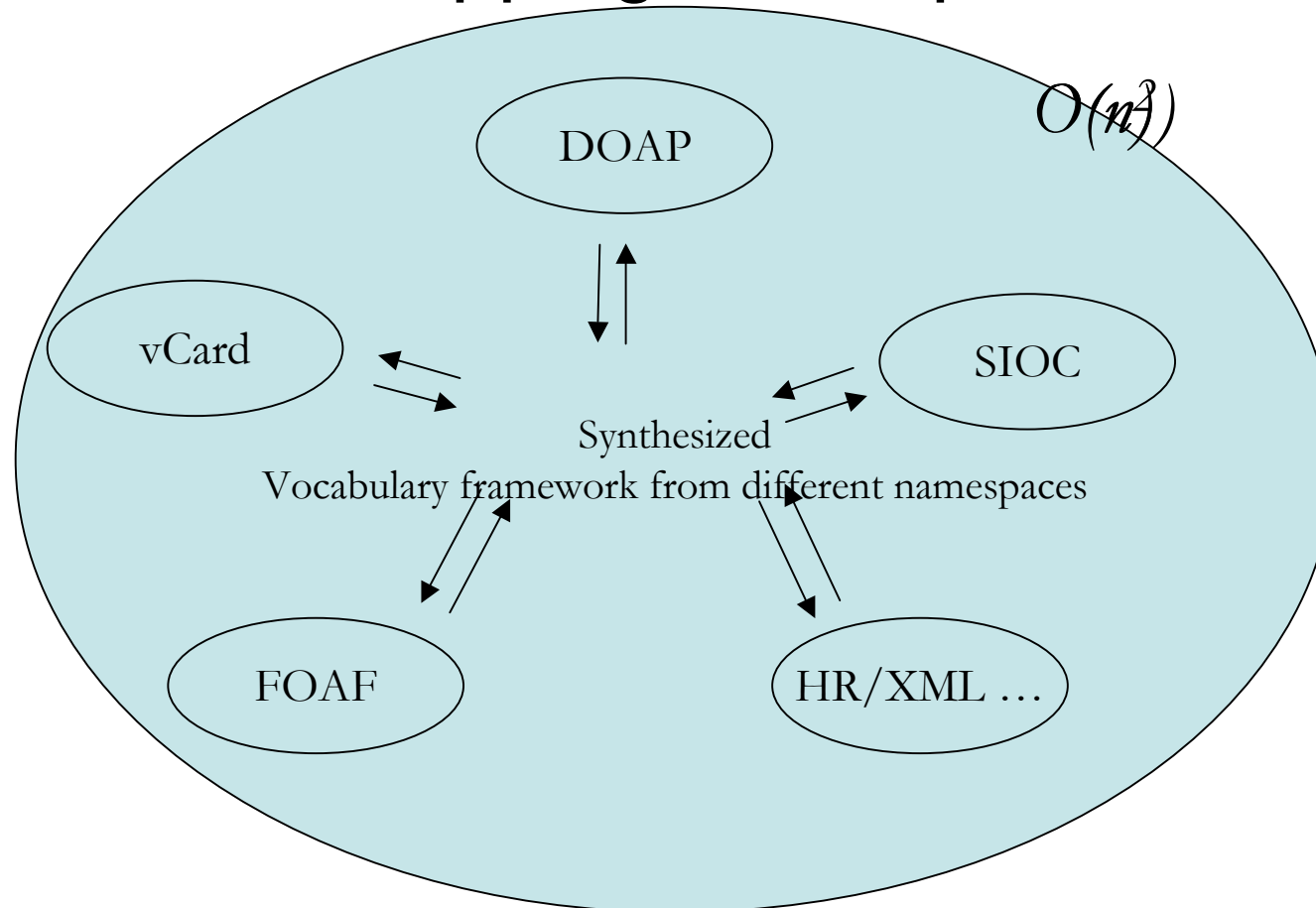
- mapping `vCard:tel` to `foaf:phone`
- conversion function generating a URI from the source RDF literal value needed...

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

... which isn't (yet?) there in SPARQL

Mapping Issues

- How to avoid mapping blow-up?



- Best practices to which vocabulary/namespace use for what!
- Choose “recommended” representative for each property/class

Another Rules Example

*Mapping **Rules** and Query Layer*

- Except mappings, rules for “linking” or defining metadata are important:
→ link to metadata published elsewhere

```
CONSTRUCT { :axel foaf:knows ?Y . }  
WHERE { axel foaf:workHomepage ?H .  
        GRAPH ?H { ?Y a foaf:Person . } }
```

“View” mechanism avoids duplication/inconsistency of data.

But: How to embed such a **rule** in my foaf-description?

Compare: RIF Use case 10

http://www.w3.org/2005/rules/wg/wiki/UCR/Publishing_Rules_for_Interlinked_Metadata

Practical Use Cases

- List of Use Cases available at:
<http://rdfweb.org/topic/ExpertFinderUseCases>

Categorized with the focus in mind:

What can be solved with **existing technologies/standards** already?

- **Basic:** Should be (almost) solvable with what is there
- **intermediate:** Non-trivial requirements e.g. on scalability, integration, identification of trustworthy sources.
- **advanced:** specific requirements with respect to enabling technologies, where we are not yet there.

1 Use Case 1 (basic): Automatic generation of institution webpages and metadata with default values



Let us assume we want to design a FOAF++ enabled content management system to manage our institution's web pages and also the member homepages. All members of the institution are allowed to provide their own metadata as extended foaf files, but, if missing, the institution can also specify some standard policies by means of some default rules. Such rules e.g. allow to aggregate metadata from some 3rd party sources. For instance, imagine, your office mate is too lazy to generate his homepage/metadata-file. No problem his basic data can be aggregated from metadata available the university personnel-database, a default publication list can be generated by the meta-data extracted from DBLP, etc.

Another advantage of this scenario is the following. While an individual person can decide to present the same meta-data provided in her own FOAF-file with his individual look-and feel on her personal homepage, the institution's member page, based on essentially the same data might have a completely different look and feel and extract/merge different data. Persons working at an institution may also inherit the address of the insitution by default, while they may possibly provide their own address and thus override the more general address of the institution.

(Remark: This is a very simple scenario, but very efficient. A main point, compared with current CMS solutions is that the generation of pages such as personal homepages or institution does not rely on a central point of data, but allows for aggregation of differenc online sources in a normalized, easy to use fashion, given that publishers follow best practices in maintaining their meta-data.)

- Partly there already (e.g. RDFHomepage project), using RDF natively, makes search for experts easier!



2 Use Case 2 (intermediate/advanced): Human Resource Management

People searching for a job simply publish their CV on their homepages or skilled employees in a company make their skill set and experiences available on the Intranet in an agreed metadata format. Job agencies can deploy agents and crawlers which they feed with their preferred profile and which find automatically suitable persons for a given vacancy. Team building within companies can be empowered by automated processes selecting the right set of employees to successfully complete a given project through semantic matching and rules.

[ExpertFinder](#) shall enable such scenarios and dezentralize the process of expert and job finding, as opposed to current central recruitment or corporate portals, just as FOAF itself was aimed to decentralize social networks.

- Formats, etc partly there, but glue is missing.

3 Use Case 3 (intermediate): Review selection

For my new journal/workshop I look for reviewers. I asked the authors during my submission to provide keywords and have found a set of ACM categories which fit my CfP. Now, using citation indexes, committees of previous conferences etc. published in the agreed metadata format, I can easily define in a declarative rule language (possibly with priorities) what are my selection criteria, or adapt selection criteria of previous workshops, if published by the organizers... A mock-up example using [DLV-HEX](#) and OWL has been presented at the [Answer Set Programming for the Semantic Web Tutorial](#) at [ESWC'06](#).

The reviewer-examples in Unit 6 of this Tutorial describe the mock-up scenario. Another nice example in Unit 7 is about aligning a meeting schedules via iCal. We used google-calendar with it's iCal export at real-time during the tutorial, was fun! ;-)

- Nice, extensible scenario with variable complexity, well-understood.



4 Use Case 4 (advanced): Trust and security for privacy-relevant meta-data

It would often be desirable that metadata should be encrypted and that the keys to decrypt it shall be provided on a timely limited basis during a process of rule based negotiation. This is for instance nice if I don't want to disclose my private phone number.

I could have rules to process this: if e.g. the person who wants my phone number calls a service to get my phone number where all persons I know are registered (even the service could simply check the sha1-sum in my foaf-file), that service will send a mail back to that address with the (temporary valid) decryption key. only this way, the person can decode my private number.

Different versions of this scenario, with different credentials, more involved negotiation are imaginable.

- More a proposed solution for the lacking privacy/security layer than a use case.



5 Use Case 5 (advanced): "Semantic CORDIS"

The [ExpertFinder](#) Initiative is obviously related to community efforts such as the EU's successful [CORDIS](#), and could be positioned as a semantic enrichment of CORDIS and a refinement to the level of individual researchers, in the direction of decentralization and enrichment of the available information stored in such portals. CORDIS enables institutions to find and contact other institutions for joint research projects. Similar to Use cases 2 and 3 the idea here is to create a kind of decentralized social network for institutions. Such an endeavour however combines the requirements for the previous use cases in a more complex scenario: Trust relevant information about projects/partners shall only be disclosed to trustworthy other parties, reliability of the information provided by different parties needs to be assessed, etc.

6 Use Case 6 (advanced): Semantic Email Addressing

[Semantical Email Addressing](#) (SEA) allows emails to be sent to semantically specified recipients. Instead of subscribing to mailinglists persons can indicate interests semantically. If one wants to send a message to a certain group of people with a common interest, one can send this specifying, e.g., a query "To all people with an interest in FOAF". Currently, a prototype SEA module is built on top of [Infomaster](#).

As mentioned in the paper [Semantic Email Addressing: Sending Email to People, not Strings](#) FOAF is a possible ontology that can be used to express interest (using foaf:interest). The use case would extend a SEA module, based on FOAF interest, to involve the use of rules to specify for example connections between interests:

if interested in FOAF then interested in Semantical Technologies

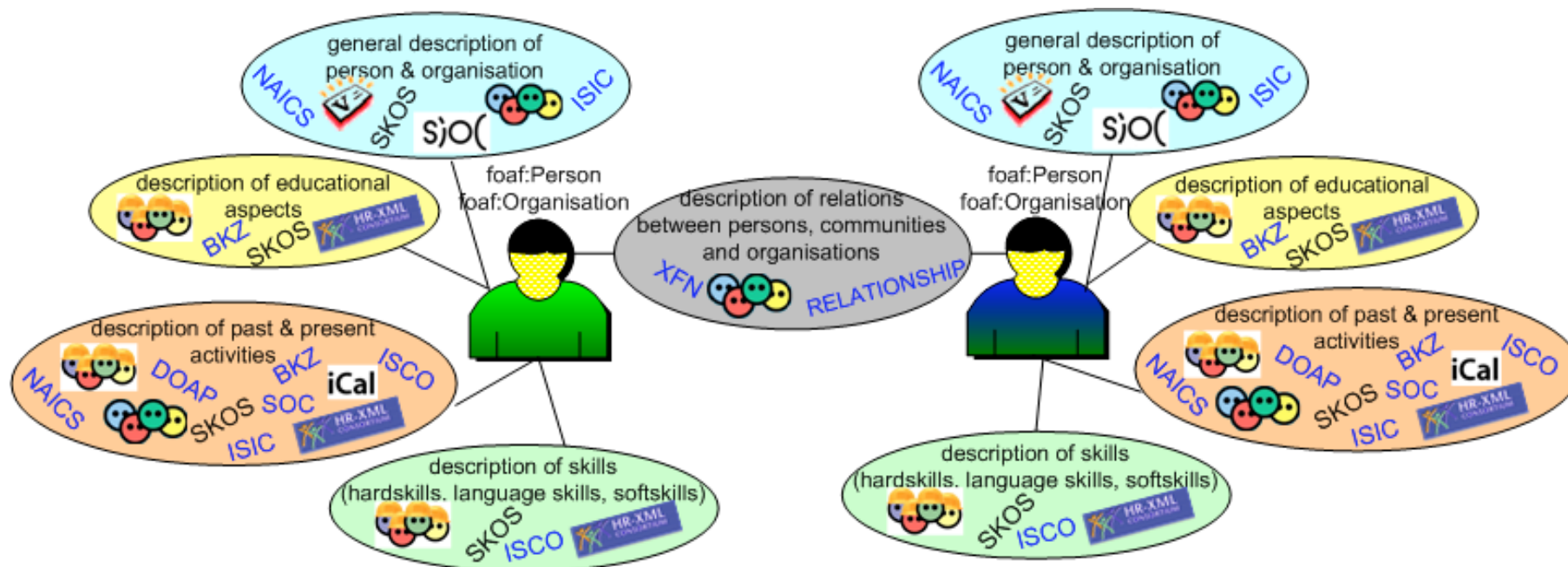
A user that indicates an interest in FOAF will, using the knowledge in the above rule, receive emails that are sent to a semantical address *all people interested in Semantical Technologies*.

Overview



- Motivation - Finding experts on the Semantic Web?
- Critical Success factors:
 - Common format
 - Critical Mass
 - Enabling Technologies (particularly Rules!)
- Practical Use Cases
- The ExpertFinder Vocabulary Framework
- Related work/what's next?

How to describe an expert?



Towards the ExpertFinder Vocabulary Framework:

- We made some first steps to identify core formats
- and suggestions how to combine them

The ExpertFinder Vocabulary Framework



- fruitfully combining of existing vocabularies
- Basis: FOAF, SIOC & SKOS
- to overcome limitations of/add missing pieces to FOAF, SIOC & SKOS use
 - refined personal data: **vCard**
 - detailed relations between persons: **RELATIONSHIP & XFN**
 - project descriptions: **DOAP**
 - CV information: **DOAC, Resume RDF Schema,...**
 - Bibliographic Descriptions: **BibTeX, DC,...**
 - Standards for Skills, e.g.
 - ACM categories for CS, Wikipedia URIs as SKOS terms
 - ...

Overview



- Motivation - Finding experts on the Semantic Web?
- Critical Success factors:
 - Common format
 - Critical Mass
 - Enabling Technologies (particularly Rules!)
- Practical Use Cases
- The ExpertFinder Vocabulary Framework
- Related work/what's next?

Self-definition & Next steps



- ExpertFinder sees itself in a role to complement efforts like FOAF and SIOC
- Recent months rather busy with
 - finishing SIOC (W3C member submission in preparation!)
 - Updates/stabilizing on FOAF
 - Set the theoretical foundations for mappings (=lightweight rules+ontology framework tailored mostly for SPARQL+RDFS)
- Provide the “glue” to plug together missing vocabularies
- Tackle use cases!
- Eventually influence standardization in this area with own standard submissions.

New ideas always welcome!

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