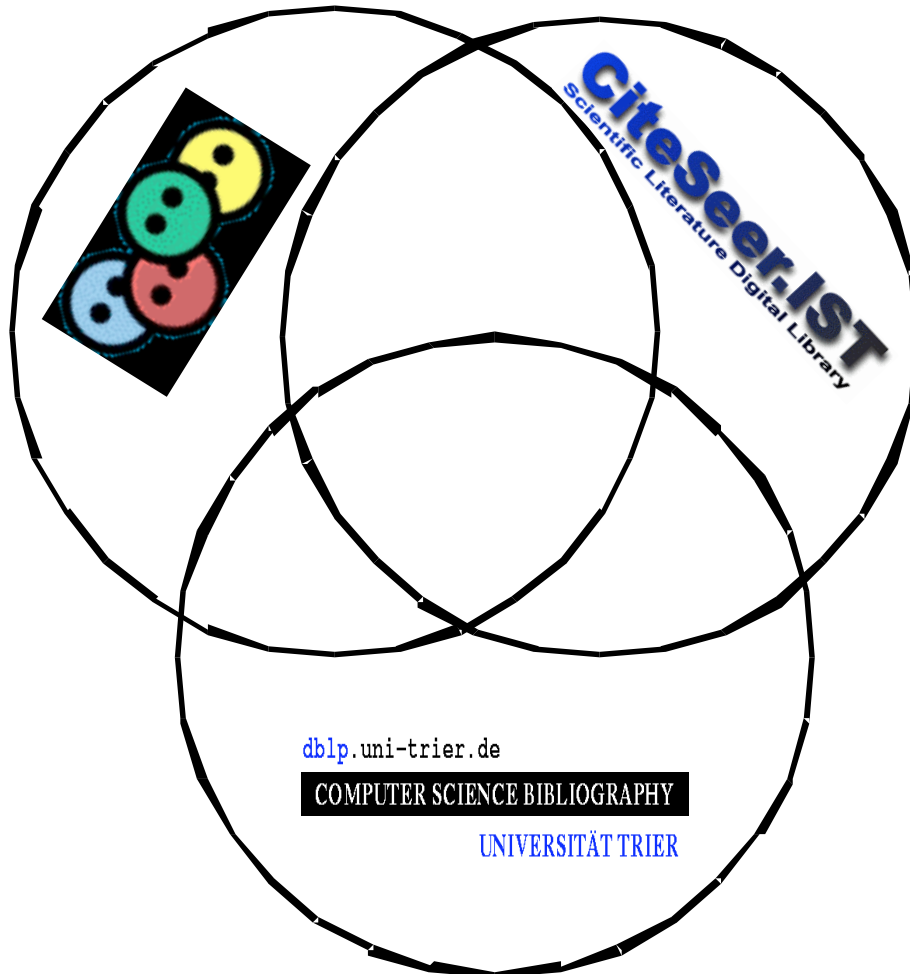


ExpertFinder Corpus 2007

Aidan Hogan

Joint work with Andreas Harth



Aim: To derive a dataset useful for ExpertFinder systems from current structured data

Dataset derived from

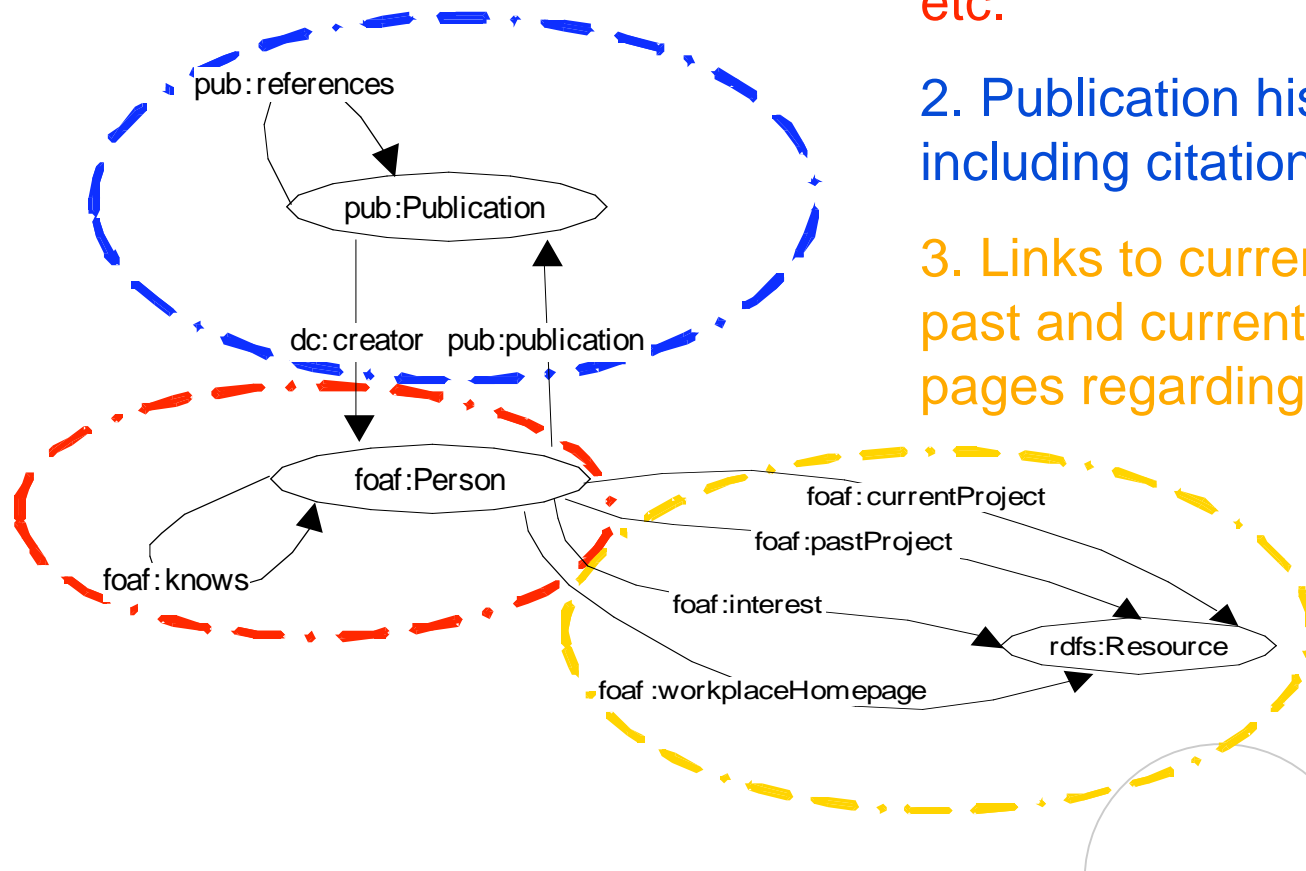
- FOAF crawled from Web
- Computer Science publication data converted from CiteSeer and DBLP to RDF using `pub` schema

Uses of Data for ExpertFinding

1. Personal details (e.g. name, location etc.), contact information, social network etc.

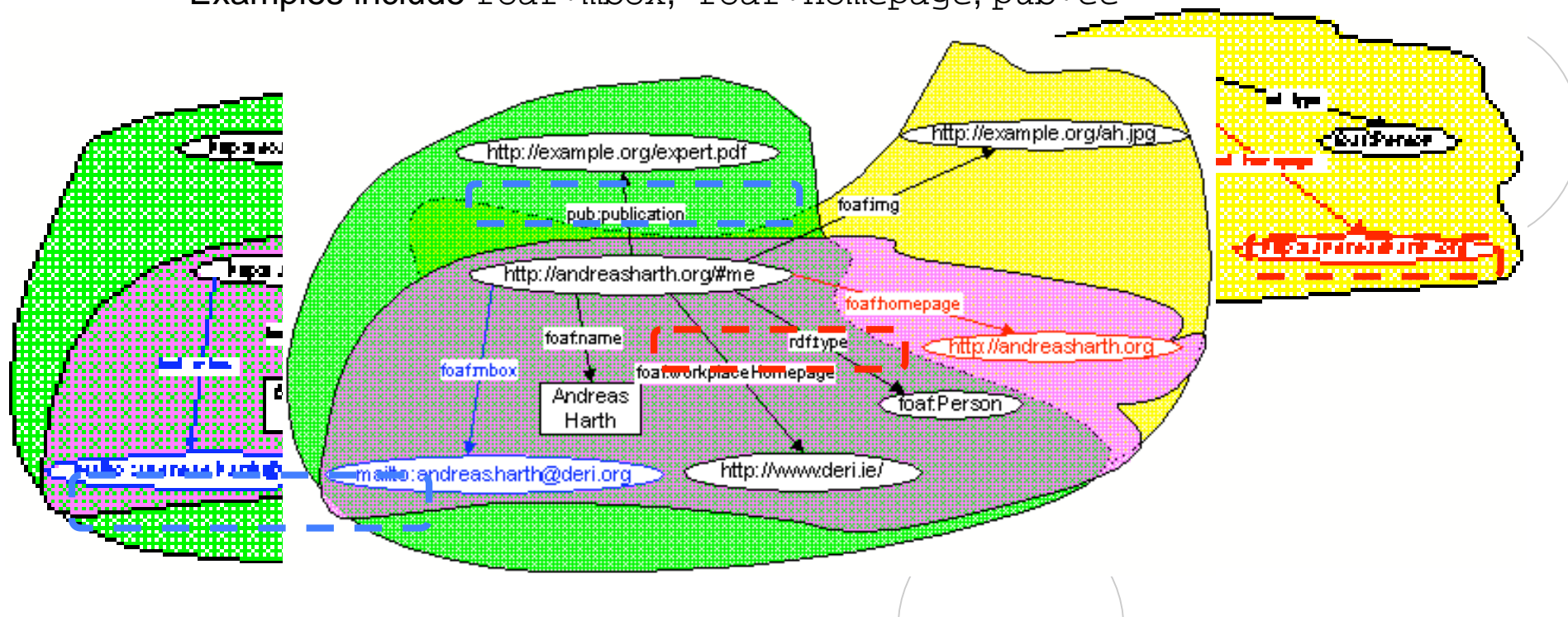
2. Publication history and details including citation information.

3. Links to current workplace homepage, past and current project pages and web pages regarding interests.



Object Consolidation

- Used to merge equivalent instances from multiple sources
- Based on analysis of inverse functional properties and their values
 - List of inverse functional properties derived from ontologies
 - Examples include `foaf:mbox`, `foaf:homepage`, `pub:ee`



Final Dataset Statistics

Characteristic	Value
Number of statements (quads/triples)	96,407,141
Number of instances	18,478,145
Number of distinct classes	224
Number of distinct predicates	719

Class	Count
foaf:Person	17,910,795
pub:Publication	715,690
pub:InProceedings	441,271
pub:Article	258,777
foaf:Document	55,582
pub:Proceedings	7,215
pub:Incollection	2,339
pub:Book	1,081

Conclusion

- Available for download tar gzipped in both RDF N-Triples and N-Quads
- <http://sw.deri.org/~aidanh/expertfinder/>

