UNIVERSITY OF Southampton
School of Electronics and Computer Science

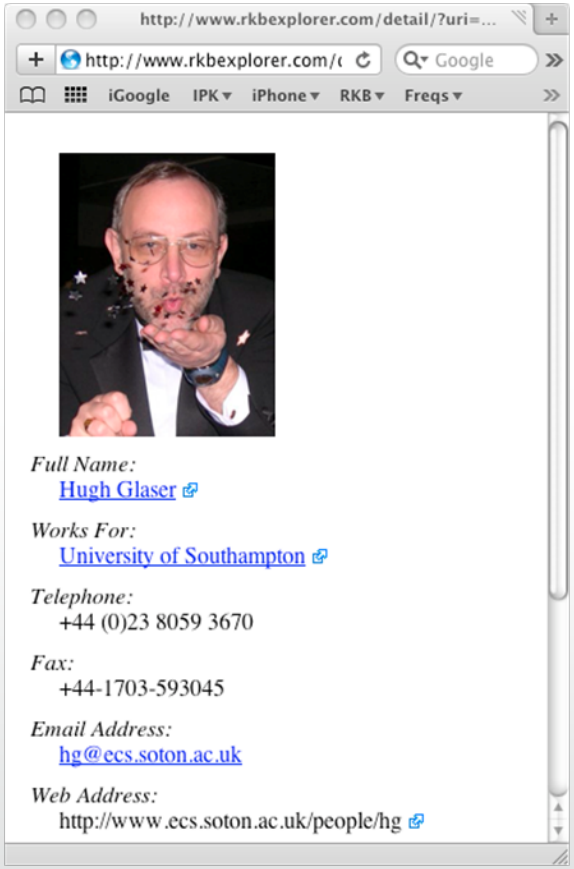
RKBExplorer: Repositories, Linked Data and Research Support

At
Eprints User Group, Open Repositories 2009

Hugh Glaser, Ian Millard & Les Carr

UNIVERSITY OF Southampton
School of Electronics and Computer Science

This is me



UNIVERSITY OF Southampton
School of Electronics and Computer Science

Background

- Rich resource of Repositories
- Other rich resources (NSF, CORDIS, ResearchIndex, Wikipedia)
- Where is the added value?

UNIVERSITY OF Southampton
School of Electronics and Computer Science

Linked Data

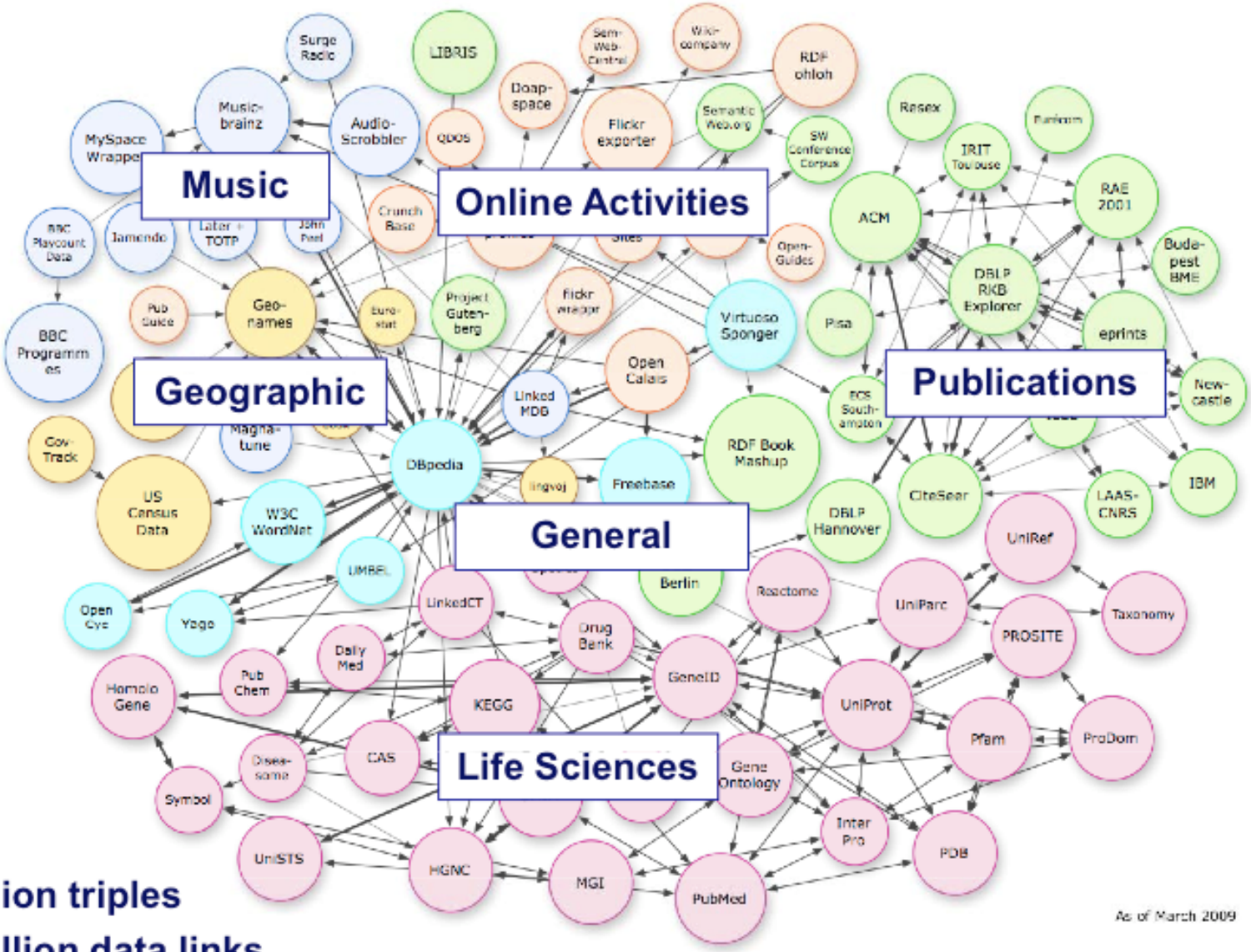
- Designed for just this
- “the Semantic Web done right, and the Web done right”
  - Tim Berners-Lee
- An Infrastructure for Naming things and Linking things
- OR can/should join

UNIVERSITY OF Southampton
School of Electronics and Computer Science

RKB and RKBExplorer

- RKB (ReSIST Knowledge Base) and RKBExplorer
- Knowledge-enabled infrastructure for cooperation in research into resilient systems
- Came out of CS AKTiveSpace
  - (Semantic Web Challenge winner 2003)
- Reasonably mature system and ongoing development

LOD Datasets on the Web: March 2009

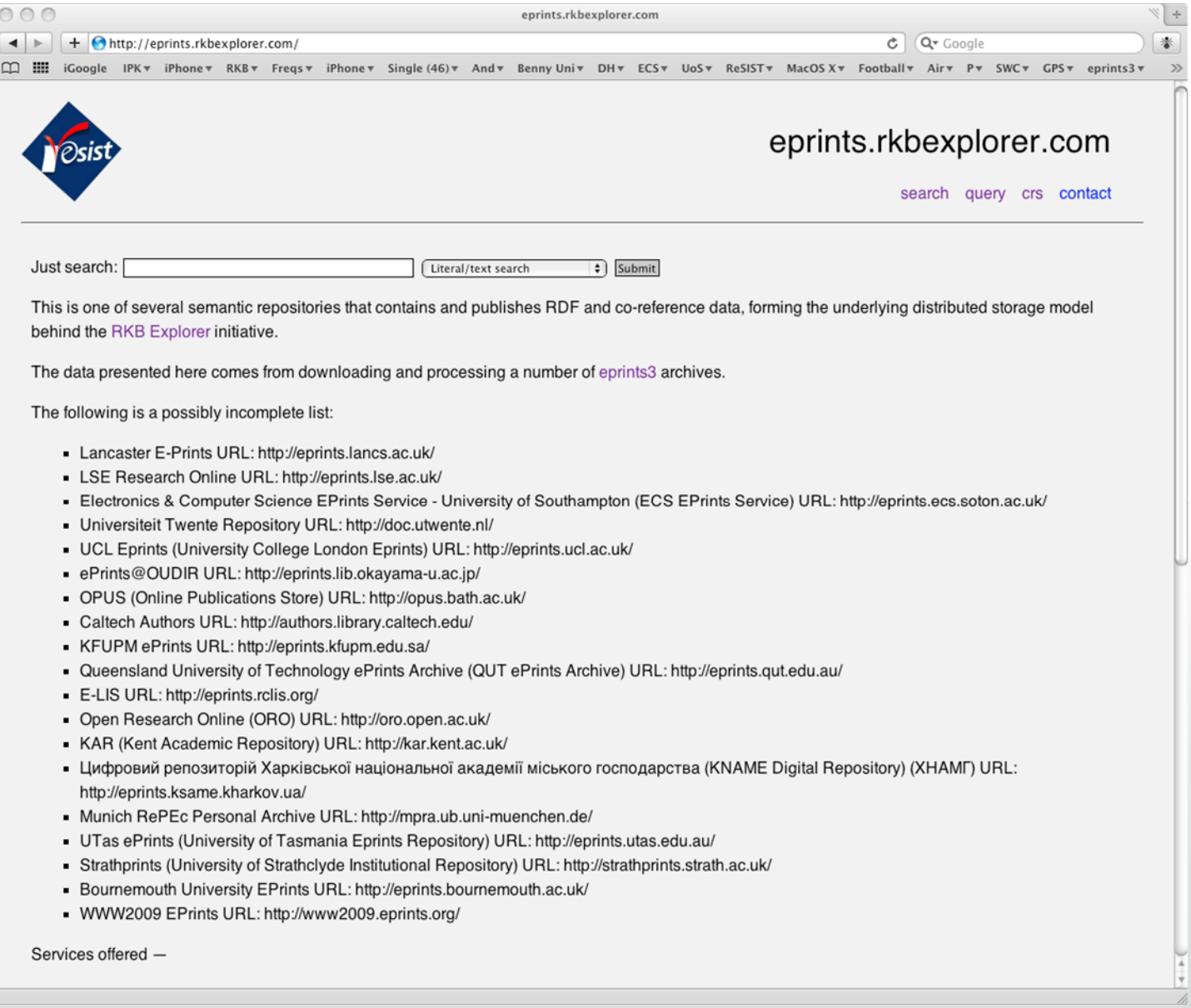


4.5 billion triples

180 million data links

Harvest some ePrints repositories

UNIVERSITY OF Southampton
School of Electronics and Computer Science



UNIVERSITY OF Southampton
School of Electronics and Computer Science

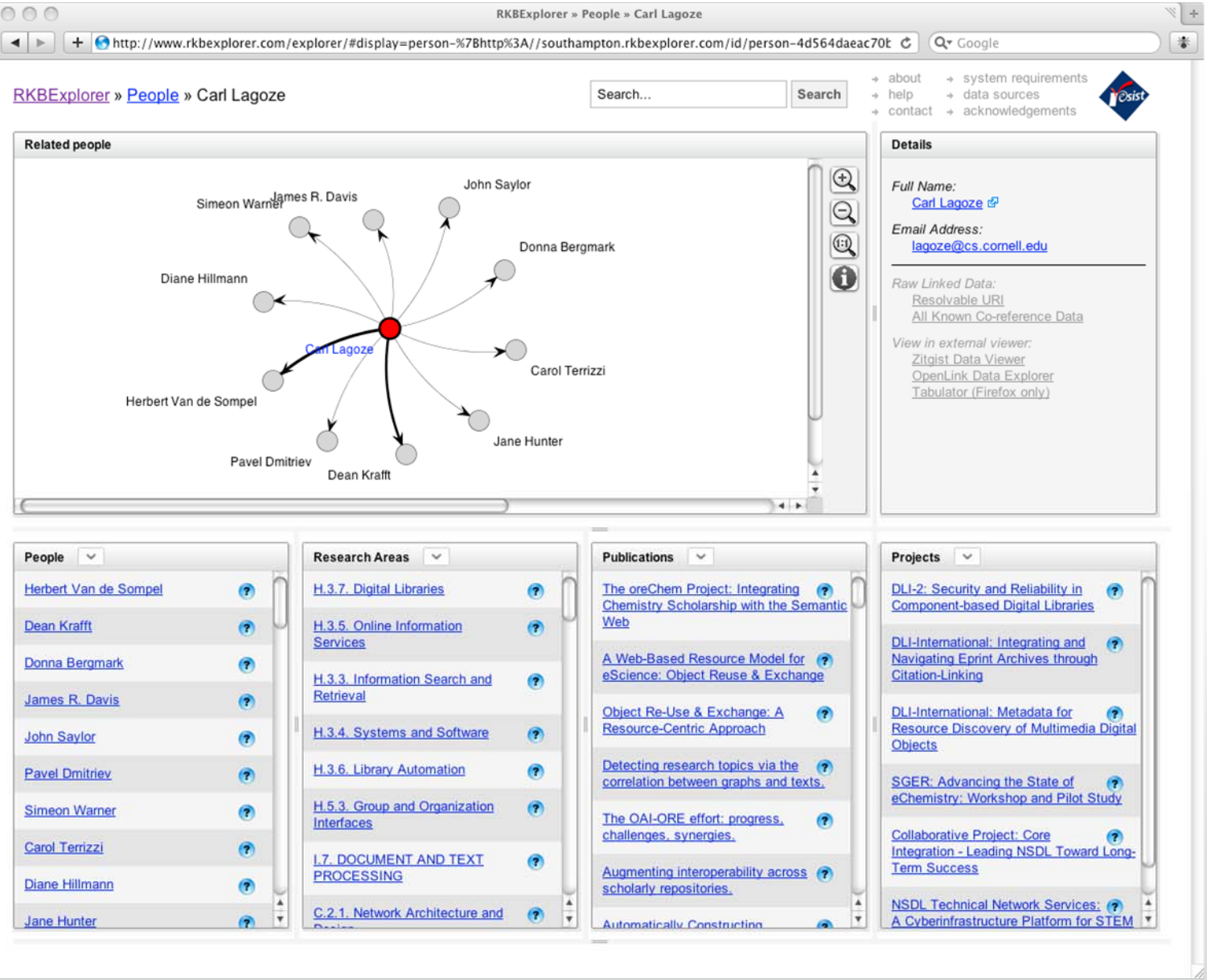
One of Many KBs, including other ePrints

acm.rkbexplorer.com	italy.rkbexplorer.com
budapest.rkbexplorer.com	kaunas.rkbexplorer.com
citeseer.rkbexplorer.com	**kisti.rkbexplorer.com**
cordis.rkbexplorer.com	laas.rkbexplorer.com
courseware.rkbexplorer.com	lisbon.rkbexplorer.com
darmstadt.rkbexplorer.com	newcastle.rkbexplorer.com
dblp.rkbexplorer.com	**nsf.rkbexplorer.com**
dbpedia.org	pisa.rkbexplorer.com
deepblue.rkbexplorer.com	rae2001.rkbexplorer.com
deploy.rkbexplorer.com	**resex.rkbexplorer.com**
epsrc.rkbexplorer.com	roma.rkbexplorer.com
eurecom.rkbexplorer.com	southampton.rkbexplorer.com
ft.rkbexplorer.com	ulm.rkbexplorer.com
ibm.rkbexplorer.com	unlocode.rkbexplorer.com
ieee.rkbexplorer.com	wiki.rkbexplorer.com
irit.rkbexplorer.com	xxx.yyy.zzz

Range from a few 100 to more than 10,000,000 “facts”

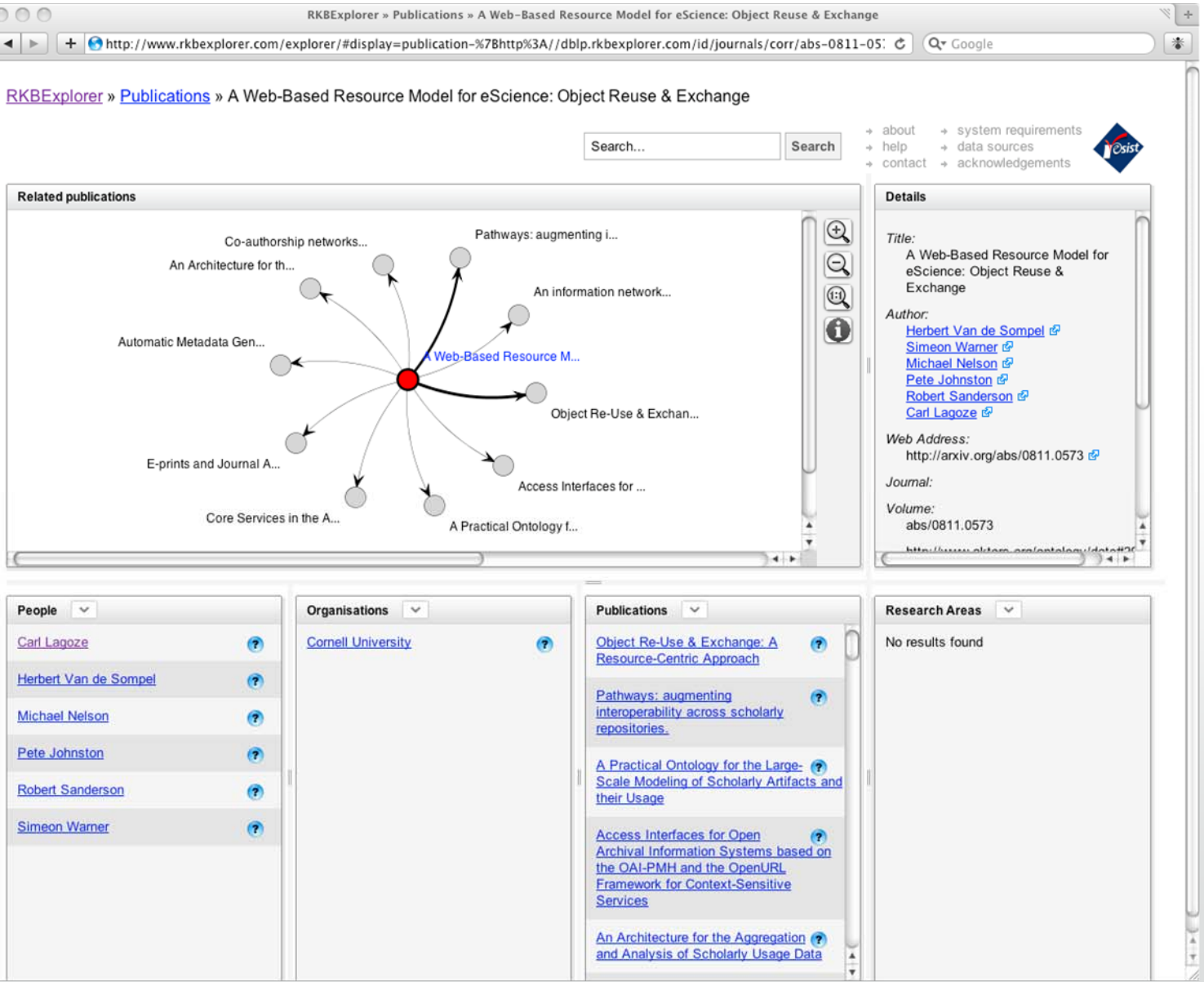
Now Look at an Author

UNIVERSITY OF Southampton
School of Electronics and Computer Science



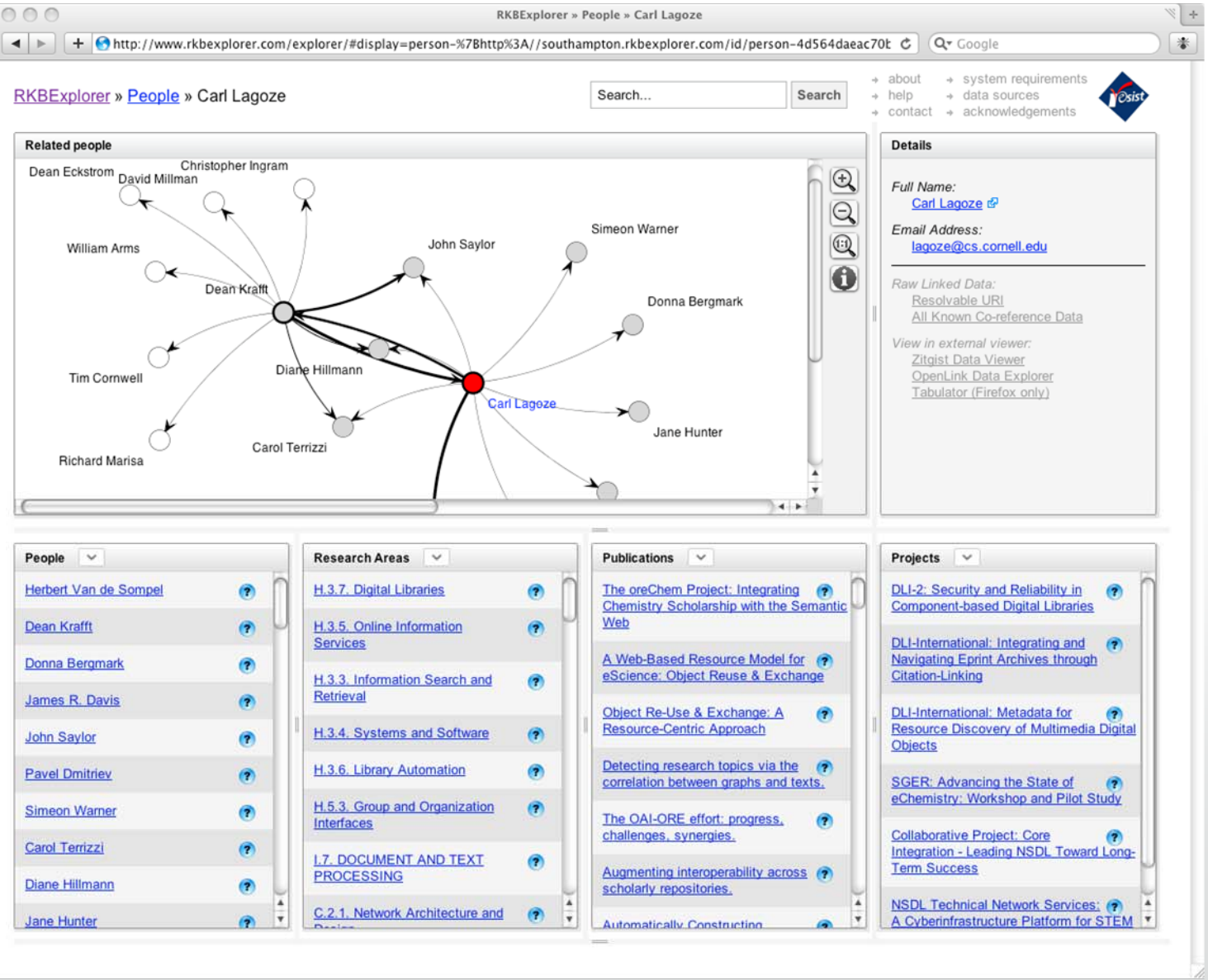
Or a Paper

UNIVERSITY OF Southampton
School of Electronics and Computer Science



Or a Couple of People

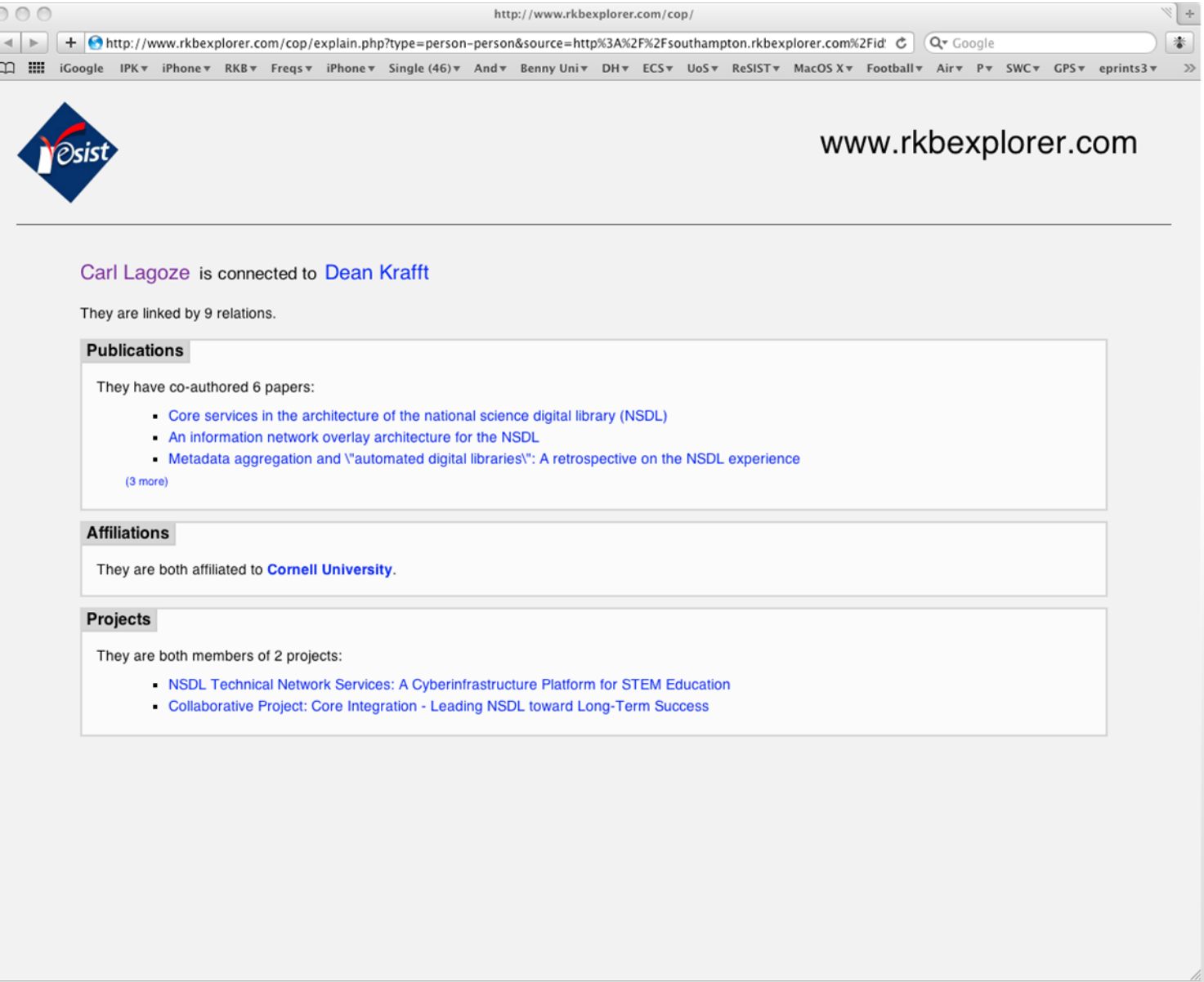
UNIVERSITY OF Southampton
School of Electronics and Computer Science



And how they are linked

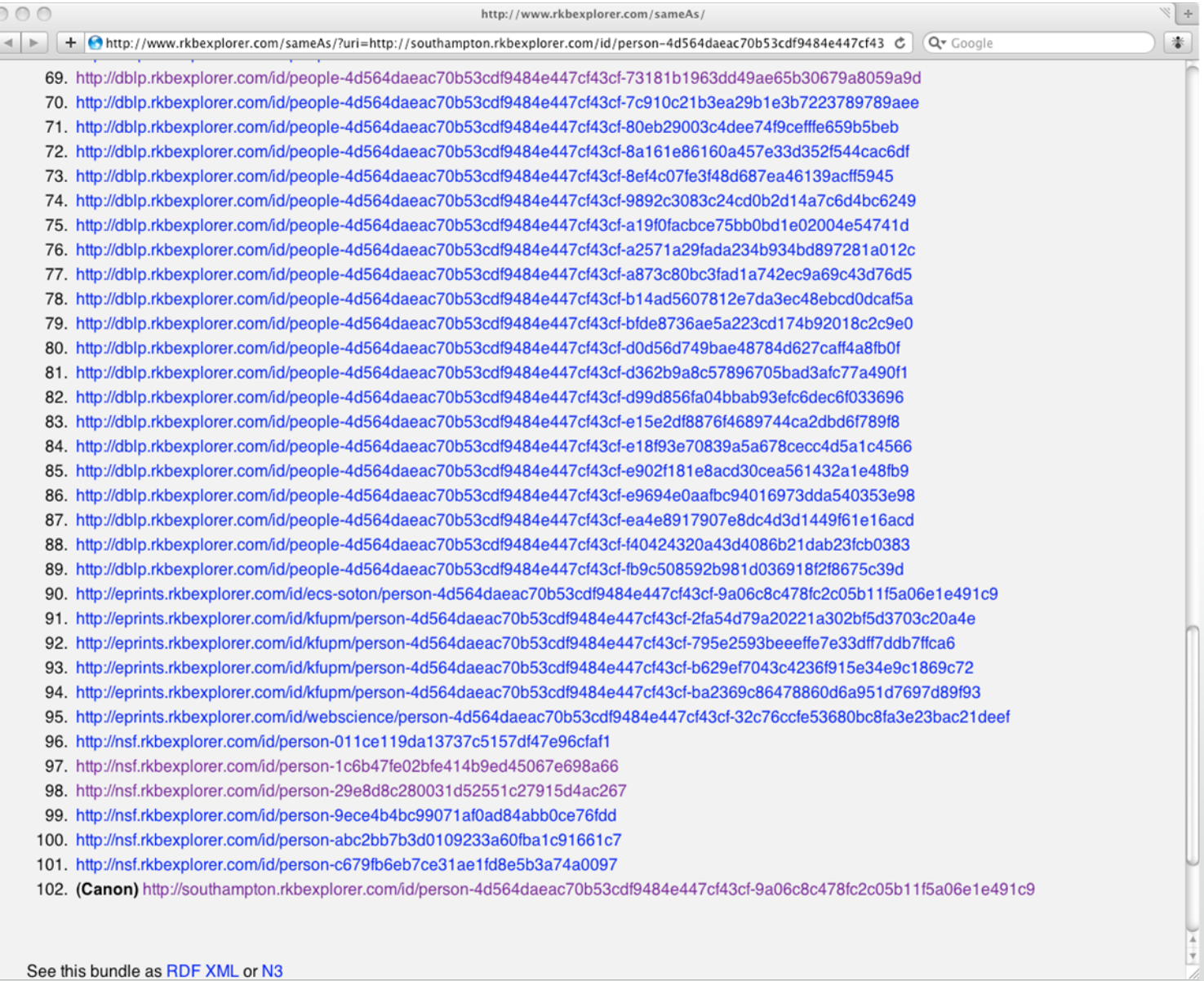
And Why they are Linked?

UNIVERSITY OF Southampton
School of Electronics and Computer Science



Who is Carl Lagoze?

UNIVERSITY OF Southampton
School of Electronics and Computer Science



Co-Reference

UNIVERSITY OF Southampton
School of Electronics and Computer Science

- Repositories have people, publications, etc. from other institutions who also have records there and elsewhere
- And vice versa
- Co-Reference is a Big Problem
  - Everything is a URI (not title, name, number…)
  - Identifying multiple URIs for one resource
  - Rejecting incorrectly conflated resources
  - Publishing
  - Using
- Coldstart
  - A serious problem
  - Nothing is linked to anything
  - Not even (reliably) within most repositories

UNIVERSITY OF Southampton
School of Electronics and Computer Science

Co-Reference Service (CRS)

- CRS Subsystem
  - Find co-references
  - Store them
  - Publish them
    - Essentially:
    - URI_i -> { URI_1, ..., URI_i, ..., URI_n }
  - Recommend a “Canon”
- Published by the Data Publisher
  - And possibly others
- Middleware aggregates co-references from recognised CRSes

Co-Reference Closure

University of Southampton
School of Electronics and Computer Science

Complete Co-Reference Information

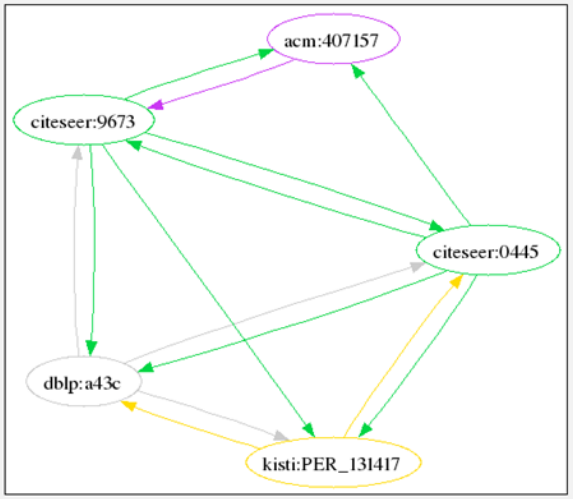
This service computes the equivalence class within the known URIs for a specified URI, by consulting all relevent CRS knowledge bases.

<http://kisti.rkbexplorer.com/id/PER_0000000000000131417> Go

Equivalent URIs...

1. **(Canon)** http://acm.rkbexplorer.com/id/person-407157
2. http://citeseer.rkbexplorer.com/id/resource-CSP179673
3. http://citeseer.rkbexplorer.com/id/resource-CSP180445
4. http://dblp.rkbexplorer.com/id/people-1ec5a600299222dd6374695ef5214f05-90d423eb148125a6e5c573dc5a15a43c
5. http://kisti.rkbexplorer.com/id/PER_0000000000000131417

The following diagram shows the interconnectivity between the CRS knowledge bases which maintain the context-dependent representation of coreference for each of the RKBExplorer domains.



Seungwoo Lee

Showing information queried from all repositories ...

Showing information queried from all repositories ...

Subject	Property	Object/Value	Source
Seungwoo Lee	akt:full-name	Seungwoo Lee [Explore]	acm-periodicals.rdf >>
Seungwoo Lee	akt:full-name	Seungwoo Lee [Explore]	acm-proceedings.rdf >>
Seungwoo Lee	akt:full-name	Seungwoo Lee [Explore]	dblp-publications-2001.rdf >>
Seungwoo Lee	akt:has-affiliation	Electrical and Computer Engineering Division, Pohang University of Science & Technology (POSTECH), Pohang, South Korea.gblee@postech.ac.kr	acm-periodicals.rdf >>
Seungwoo Lee	akt:has-affiliation	POSTECH, Pohang, Korea	acm-proceedings.rdf >>
Seungwoo Lee	kisti:engNameOfPerson	Seungwoo Lee [Explore]	datatypeproperties.ttl >>
Seungwoo Lee	rdf:type	akt:Affiliated-Person	acm-periodicals.rdf >>
Seungwoo Lee	rdf:type	akt:Affiliated-Person	acm-proceedings.rdf >>
Seungwoo Lee	rdf:type	Generic Agent	acm-periodicals.rdf >>
Seungwoo Lee	rdf:type	Generic Agent	acm-proceedings.rdf >>
Seungwoo Lee	rdf:type	Generic Agent	dblp-publications-2001.rdf >>
Seungwoo Lee	rdf:type	akt:Person	acm-periodicals.rdf >>
Seungwoo Lee	rdf:type	akt:Person	acm-proceedings.rdf >>
Seungwoo Lee	rdf:type	akt:Person	dblp-publications-2001.rdf >>
Seungwoo Lee	rdf:type	PER_char(20)"^^	datatypeproperties.ttl >>
Seungwoo Lee	rdf:type	PER_char(20)"^^	objectproperties.ttl >>
Seungwoo Lee	rdf:type	PER_char(20)"^^	resources.ttl >>

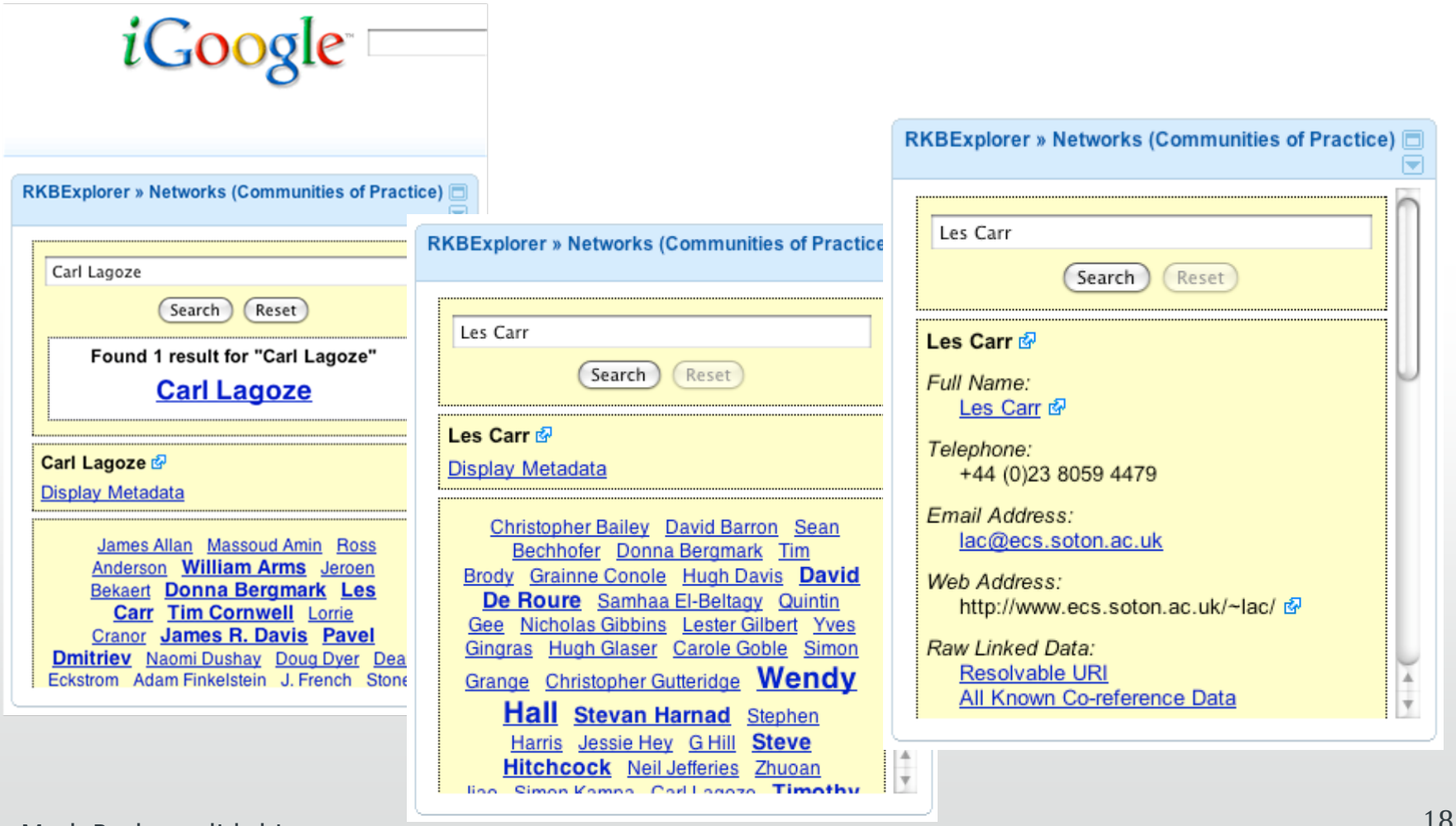
Subject	Property	Object	Source
Automatic acquisition of named entity tagged corpus from world wide web	akt:has-author	Seungwoo Lee	acm-proceedings.rdf >>
A Corpus-Based Learning Method of Compound Noun Indexing Rules for Korean	akt:has-author	Seungwoo Lee	acm-periodicals.rdf >>
SiteQ: Engineering High Performance QA System Using Lexico-Semantic Pattern Matching and Shallow NLP.	akt:has-author	Seungwoo Lee	dblp-publications-2001.rdf >>
A Corpus-Based Learning Method of Compound Noun			dblp-publications-

Open System

UNIVERSITY OF Southampton
School of Electronics and Computer Science

- RKBExplorer is only one interface
  - And not a required part
- Services:
  - Details for a paper (the right hand pane in RKBExplorer):
    - http://www.rkbexplorer.com/detail/?uri=http://southampton.rkbexplorer.com/id/eprints-12614
  - Network of people for a publication (lower pane):
    - http://www.rkbexplorer.com/network/?uri=http://southampton.rkbexplorer.com/id/eprints-12614&type=publication-person
  - …
- Other Interfaces (using the services)
  - Personal Web pages
  - iPhone
  - iGoogle Gadget

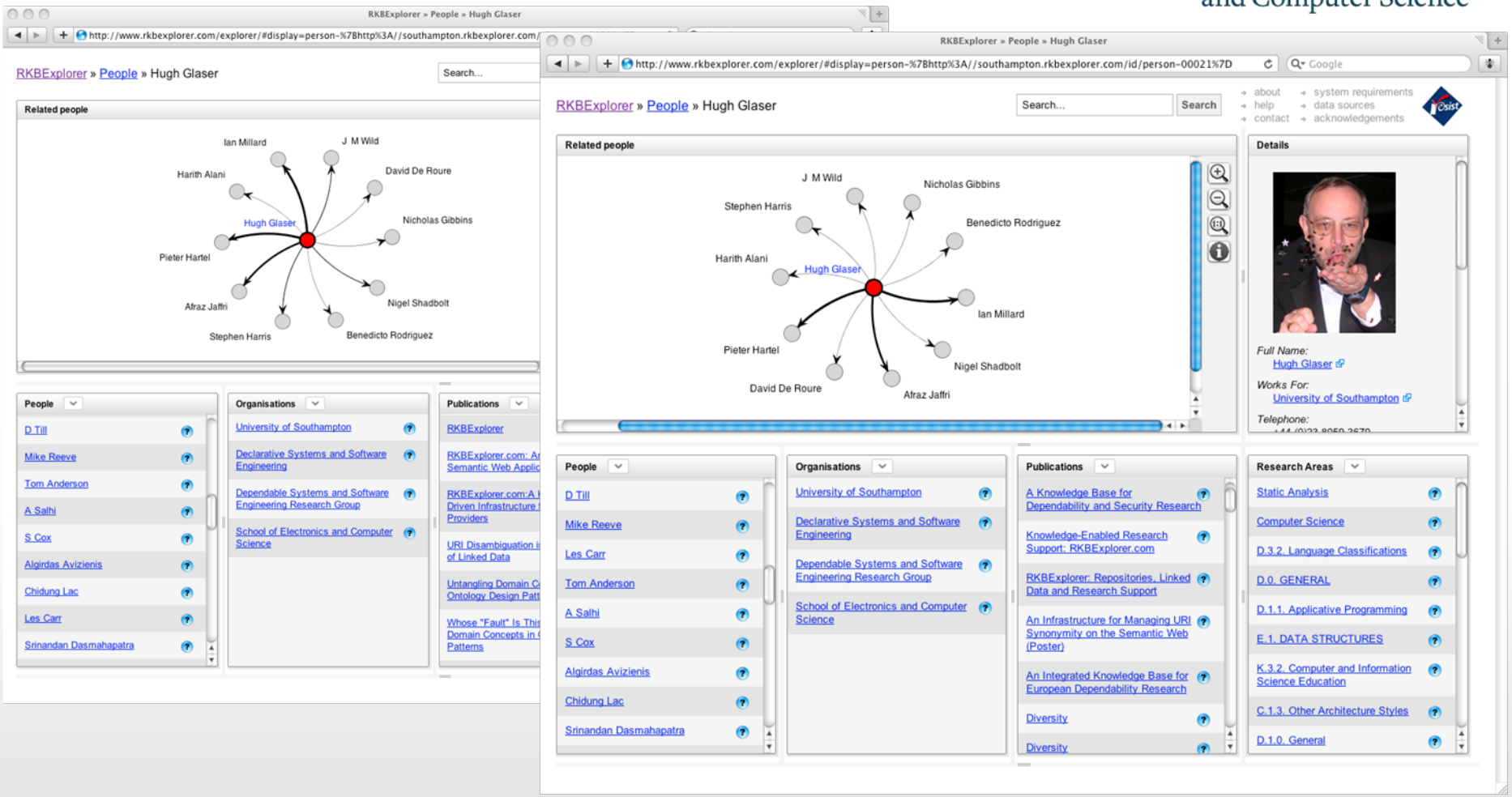
Gadget – find out about people



Mark Borkum did this

Update in Action

UNIVERSITY OF Southampton
School of Electronics and Computer Science



Before and after inserting this paper in the Southampton ePrints repository and RKB has noticed
Note the position of Les Carr in Hugh's related People

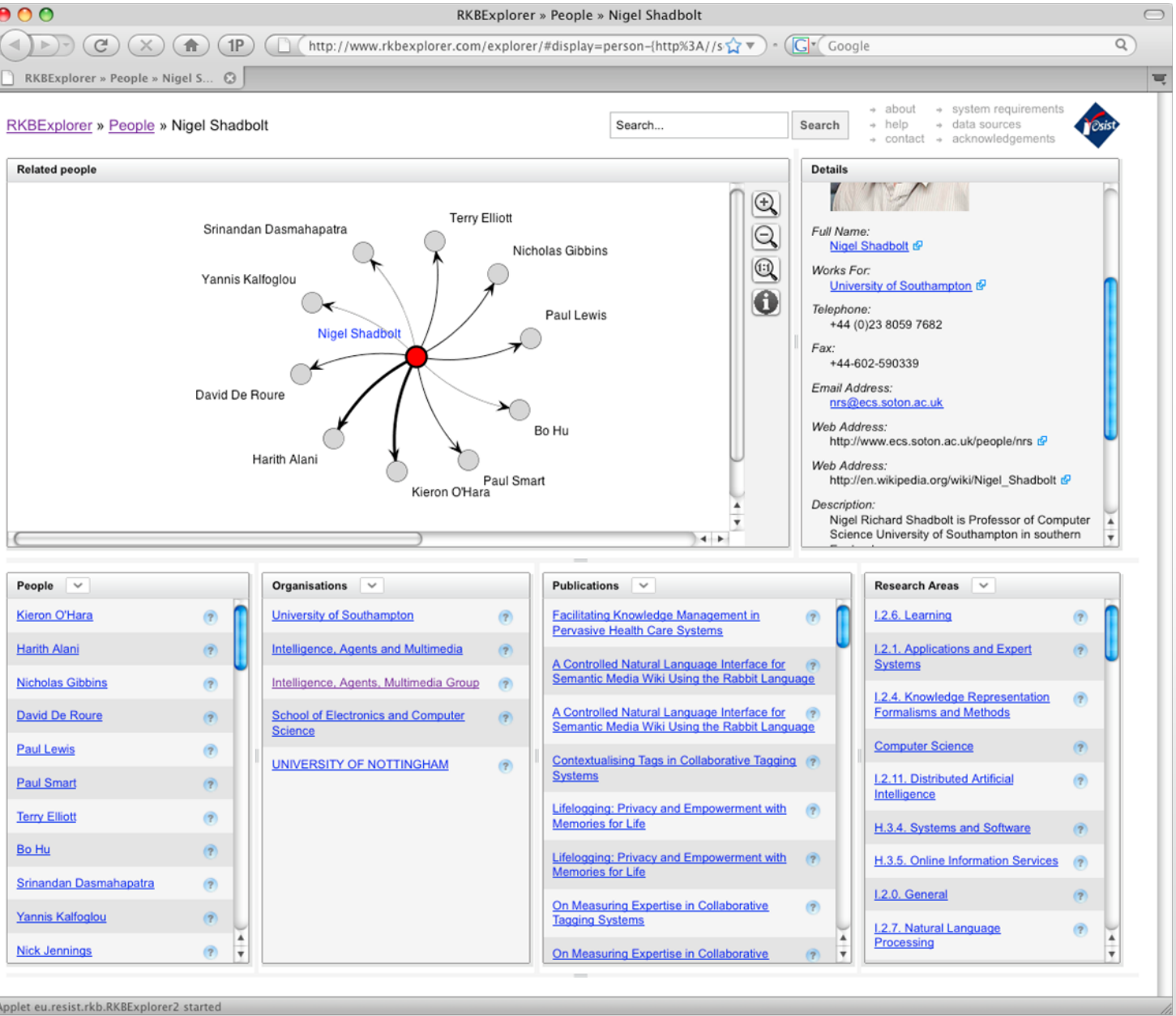
UNIVERSITY OF Southampton
School of Electronics and Computer Science

Concluding Remarks

- ePrints today, other systems tomorrow
  - Other related technologies (such as OAI-ORE)
  - Are they right for this?
- Please don't stop at the repository
- Go on and get the added value of Linked Data
- ePrints has plans to publish RDF
  - Will the schema (ontology) by expressive enough
- Worry about your co-reference
  - Do you have IDs in your respository?
  - Can you reliably identify all the papers of a single person?

http://eprints.ecs.soton.ac.uk/17373/

UNIVERSITY OF Southampton
School of Electronics and Computer Science



RKBExplorer.com

Try it!