UNIVERSITY OF Southampton
School of Electronics and Computer Science

Research on Linked Data and Co-reference Resolution

Dublin Core & Metadata Applications 2009
Seoul, Korea

Hugh Glaser, Ian Millard:
University of Southampton, UK
성원경, 이승우, 김평, 류범종
Won-Kyung Sung, Seungwoo Lee, Pyung Kim, Beom-Jong You:
KISTI, Korean Institute of Science and Technology Information

http://eprints.ecs.soton.ac.uk/17761/

UNIVERSITY OF Southampton
School of Electronics and Computer Science

Background

- Southampton & KISTI
  - Semantic Web research and development
  - Datasets
- Project
  - Match KISTI datasets with Southampton activities

UNIVERSITY OF Southampton
School of Electronics and Computer Science

Objectives

- Publish KISTI dataset as Linked Data
- Investigate multiple ontology interoperation
- Co-reference identification
- Integration of KISTI data into RKBExplorer

UNIVERSITY OF Southampton
School of Electronics and Computer Science

Context: KISTI

- RDF
  - Metadata for publications including texts
  - Sources include
    - Korean conferences
    - Citeseer
- Semantic Web applications

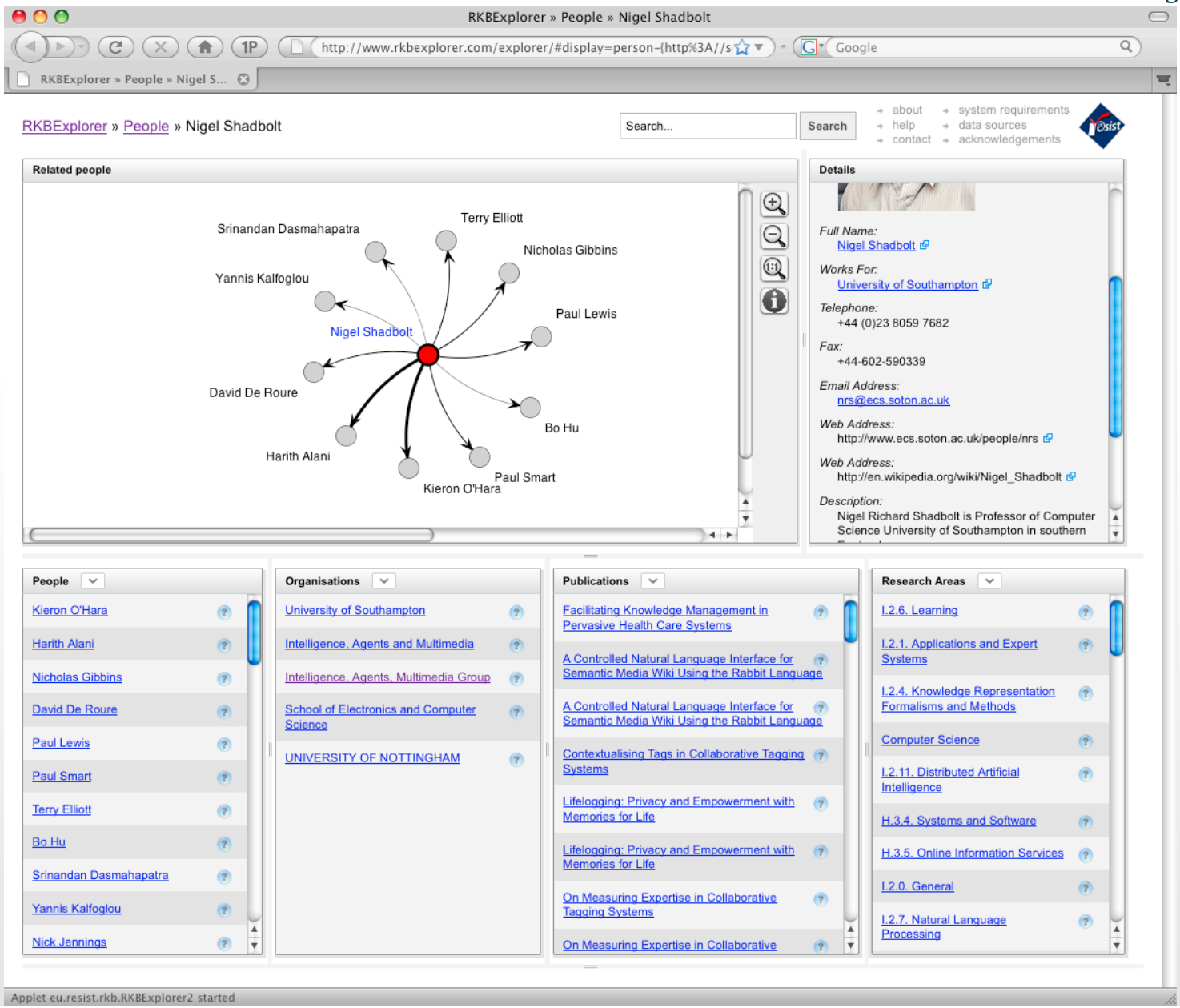
UNIVERSITY OF Southampton
School of Electronics and Computer Science

Context: 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

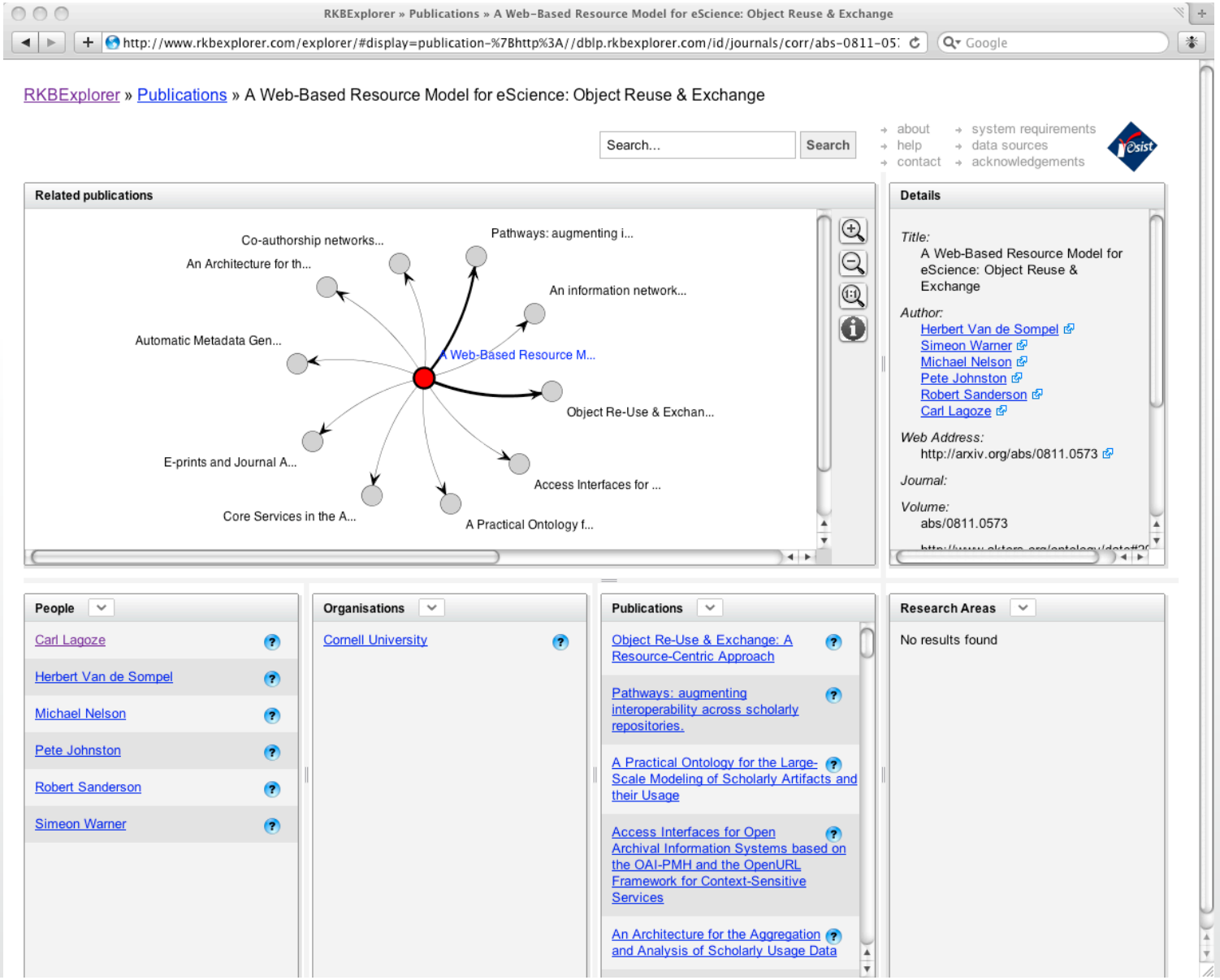
RKBExplorer.com

UNIVERSITY OF Southampton
School of Electronics and Computer Science



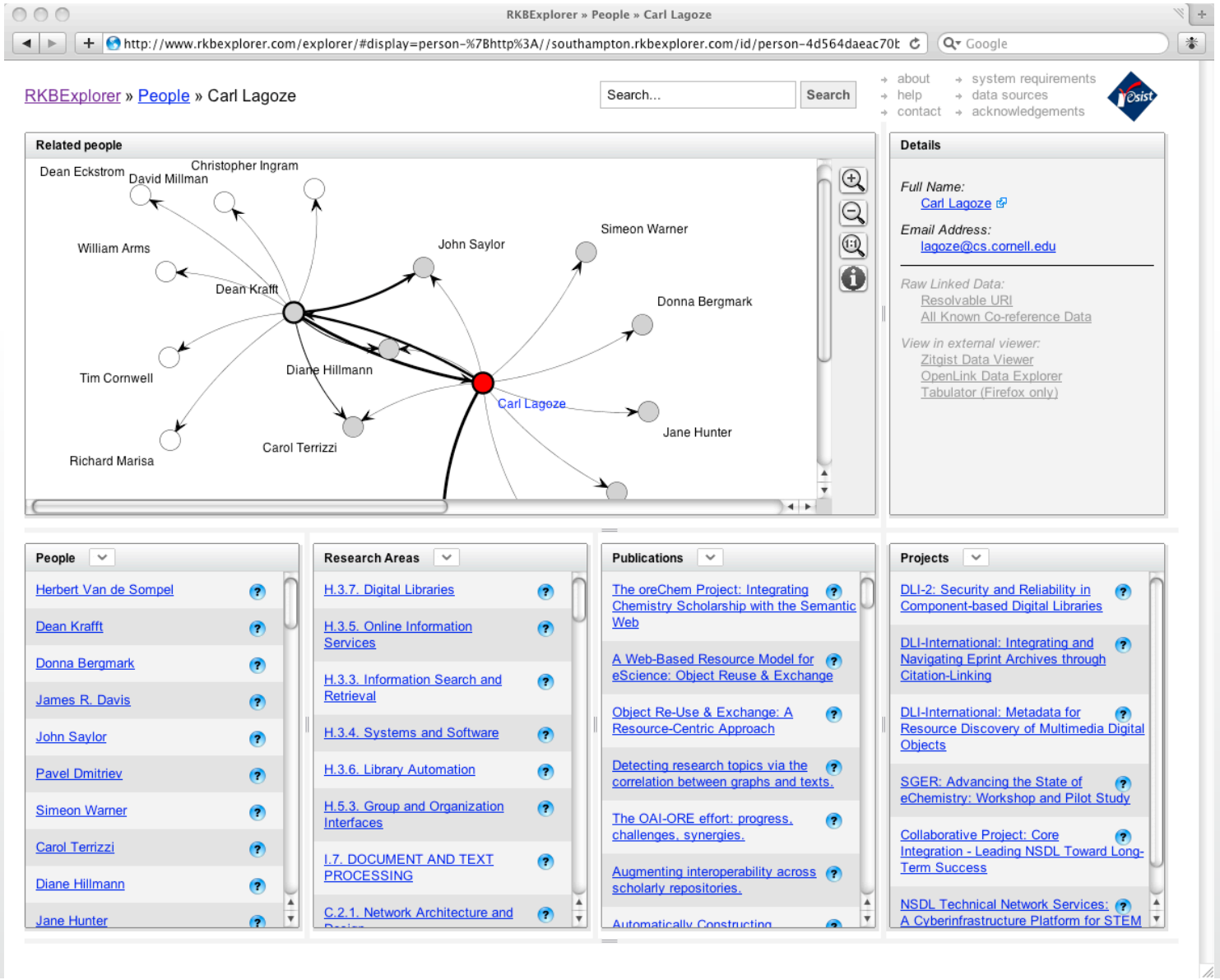
Or a Paper

UNIVERSITY OF Southampton
School of Electronics and Computer Science



Or a Couple of People

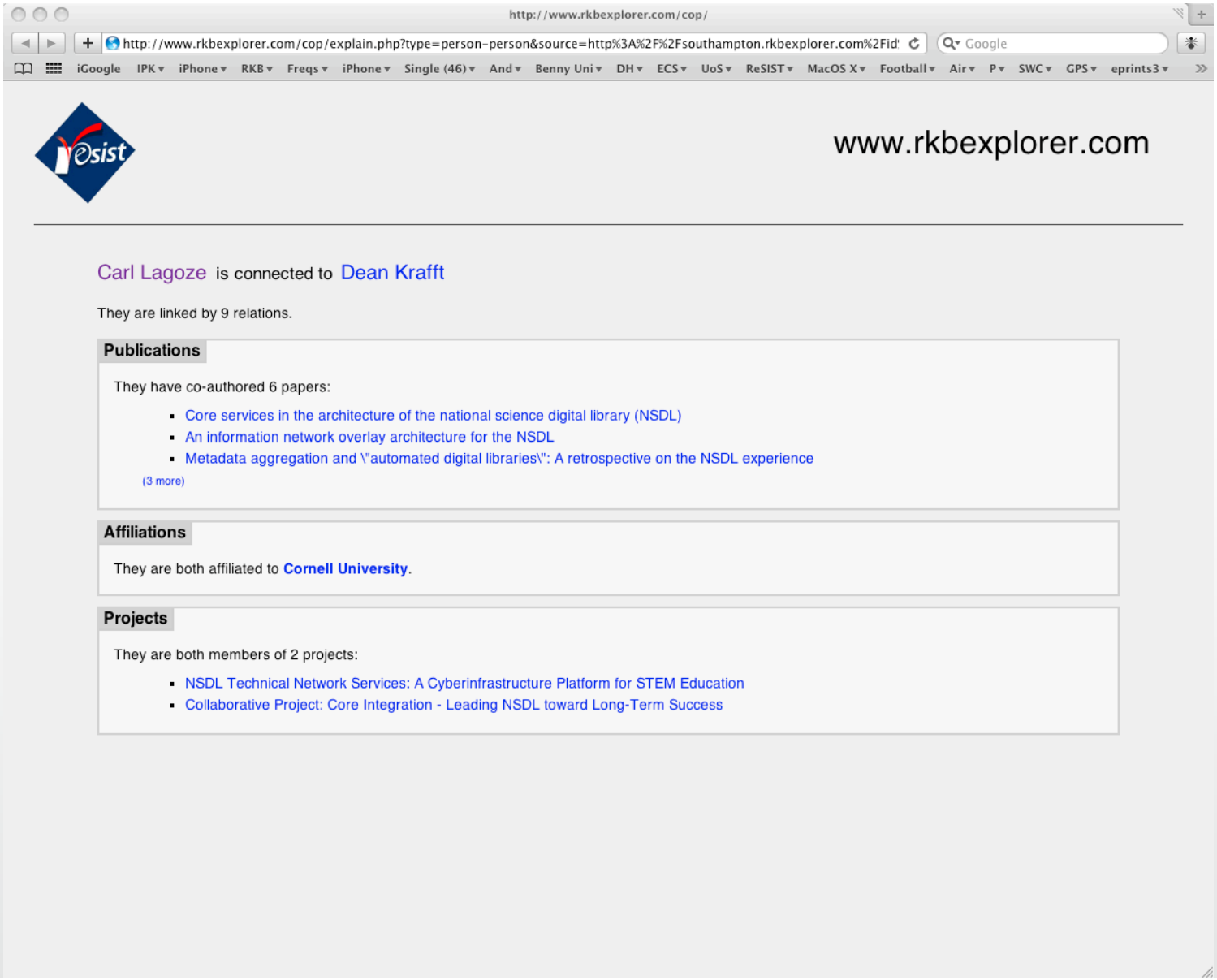
UNIVERSITY OF Southampton
School of Electronics and Computer Science



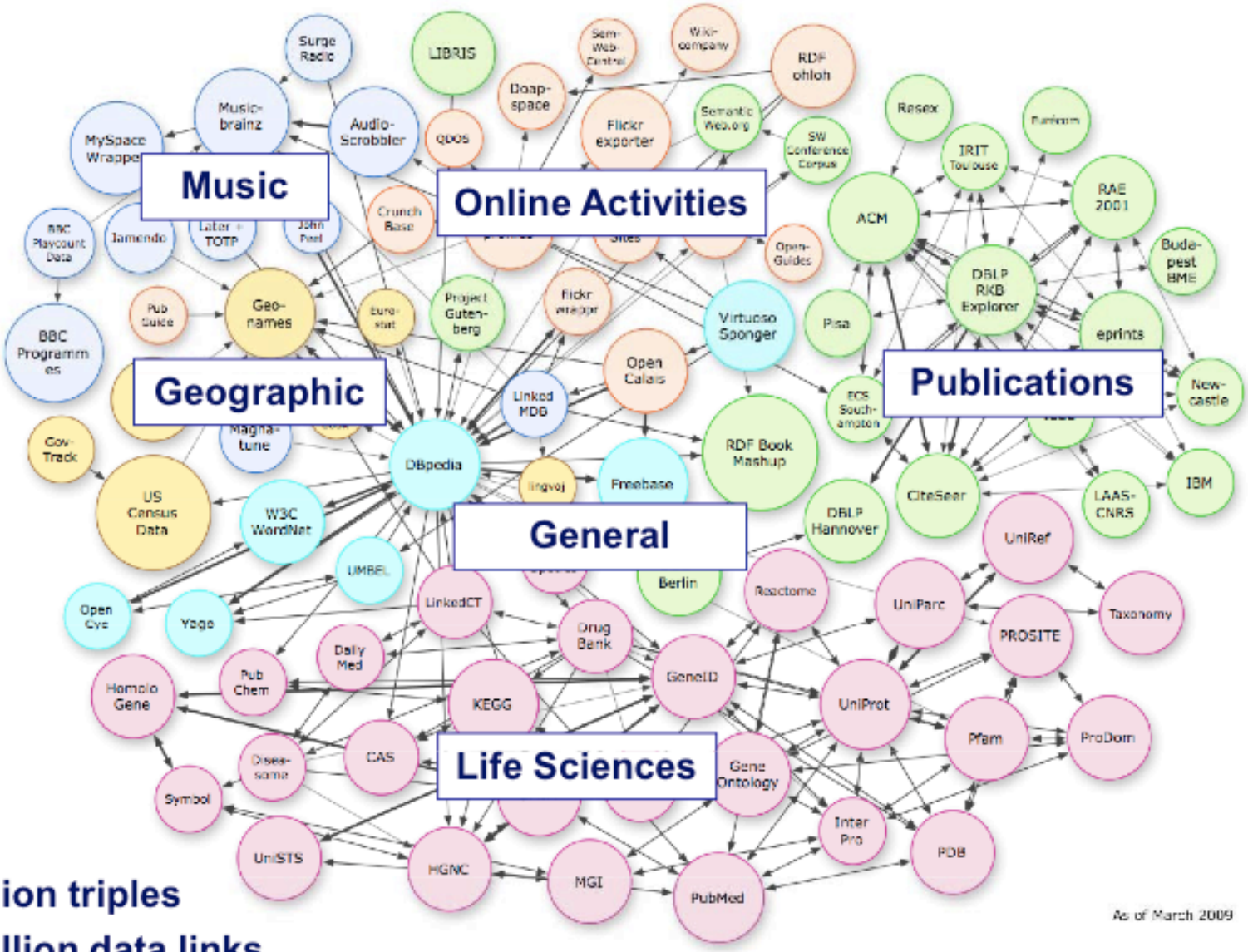
And how they are linked

And Why they are Linked?

UNIVERSITY OF Southampton
School of Electronics and Computer Science



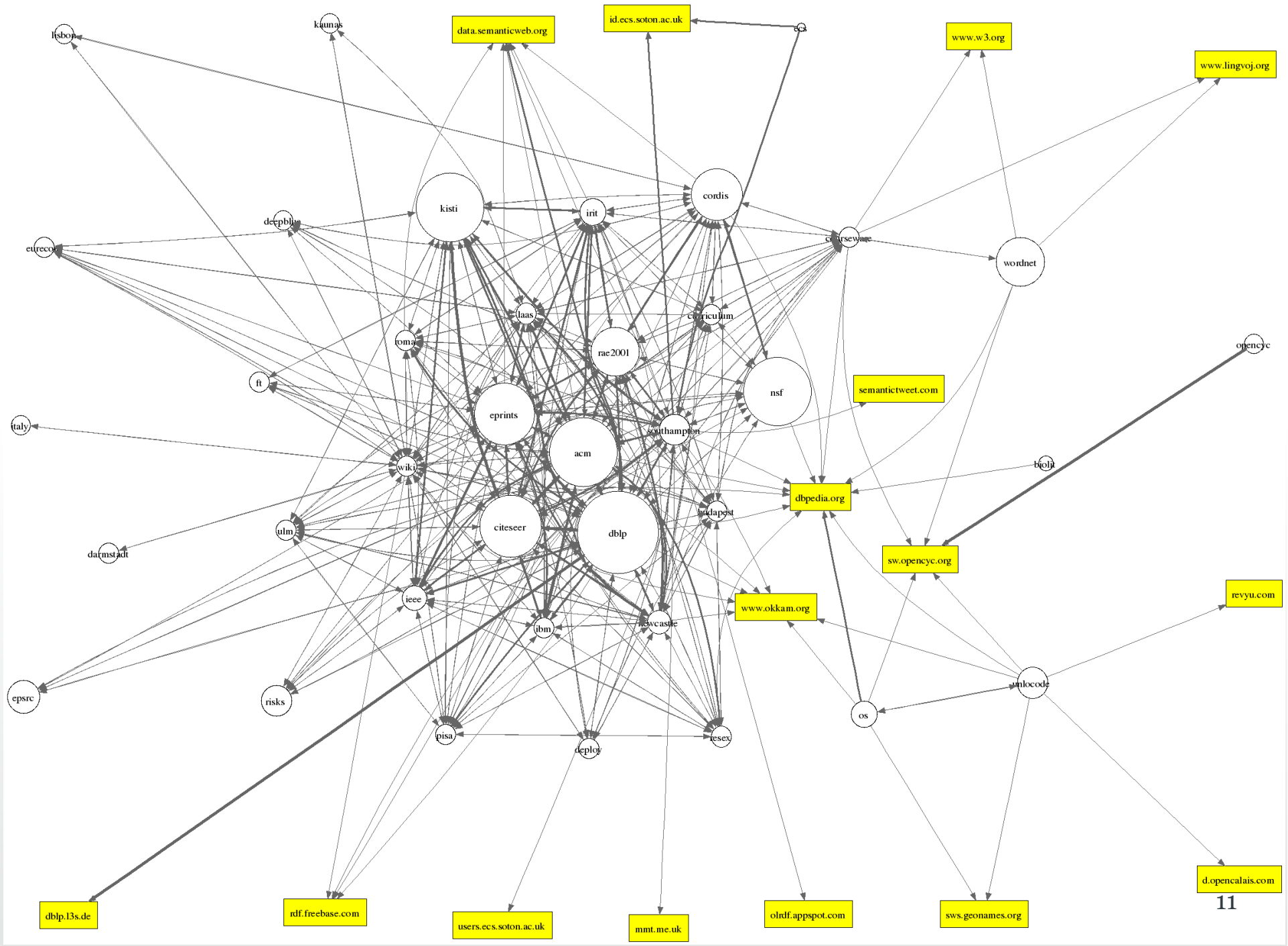
LOD Datasets on the Web: March 2009



4.5 billion triples

180 million data links

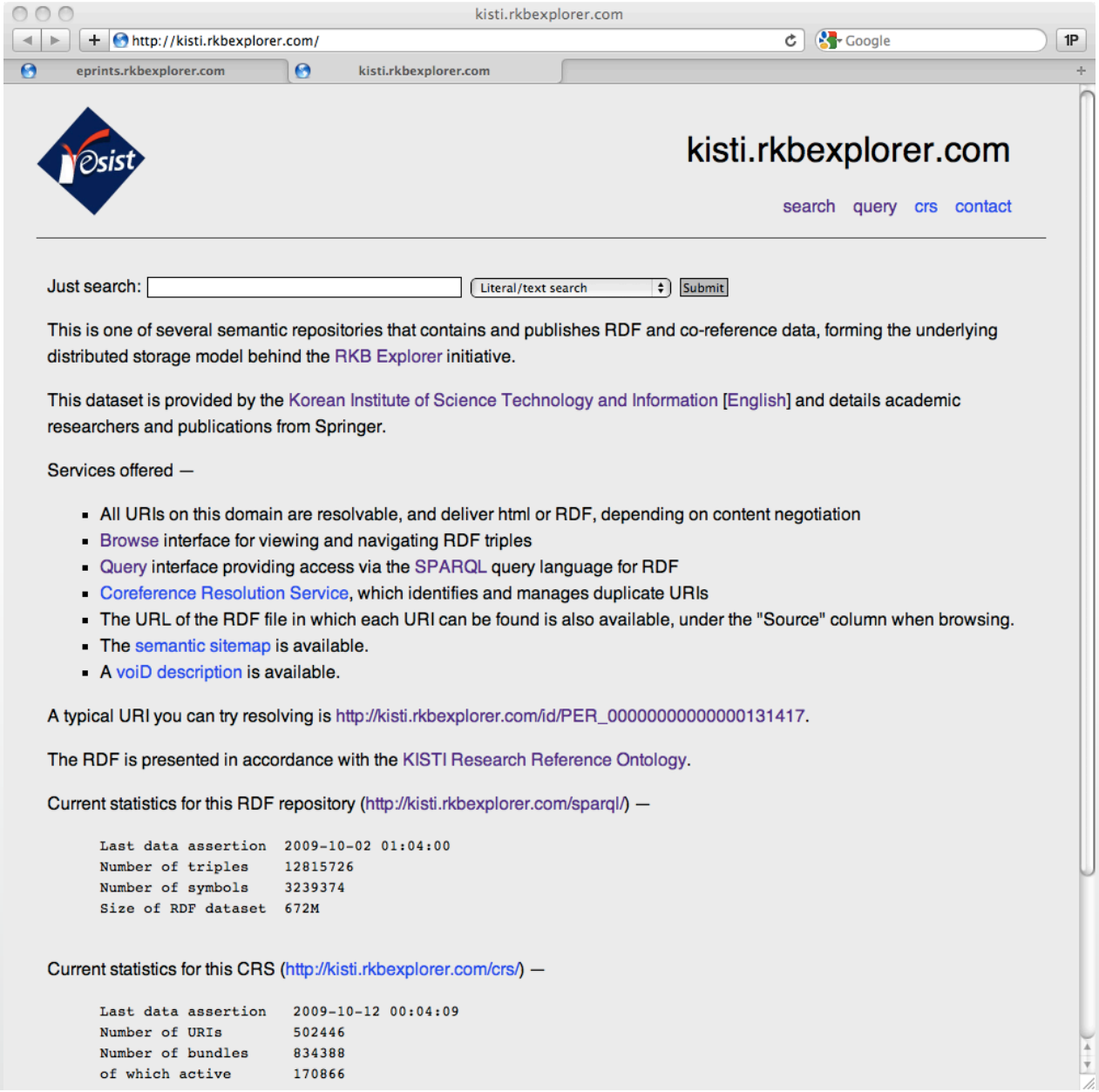
As of March 2009



lisbon
kaunas
data.semanticweb.org
id.ecs.soton.ac.uk
ecs
www.w3.org
www.lingvoj.org
kisti
irit
cordis
deepblue
eurecom
courseware
wordnet
laas
curriculum
roma
rae2001
opencyc
ft
nsf
semantictweet.com
eprints
italy
southampton
acm
wiki
biolit
dbpedia.org
budapest
ulm
citeseer
dblp
darmstadt
sw.opencyc.org
ieee
www.okkam.org
revyu.com
ibm
newcastle
unlocode
epsrc
risks
os
pisa
deploy
resex
d.opencalais.com
dblp.l3s.de
rdf.freebase.com
users.ecs.soton.ac.uk
mmt.me.uk
olrdf.appspot.com
sws.geonames.org

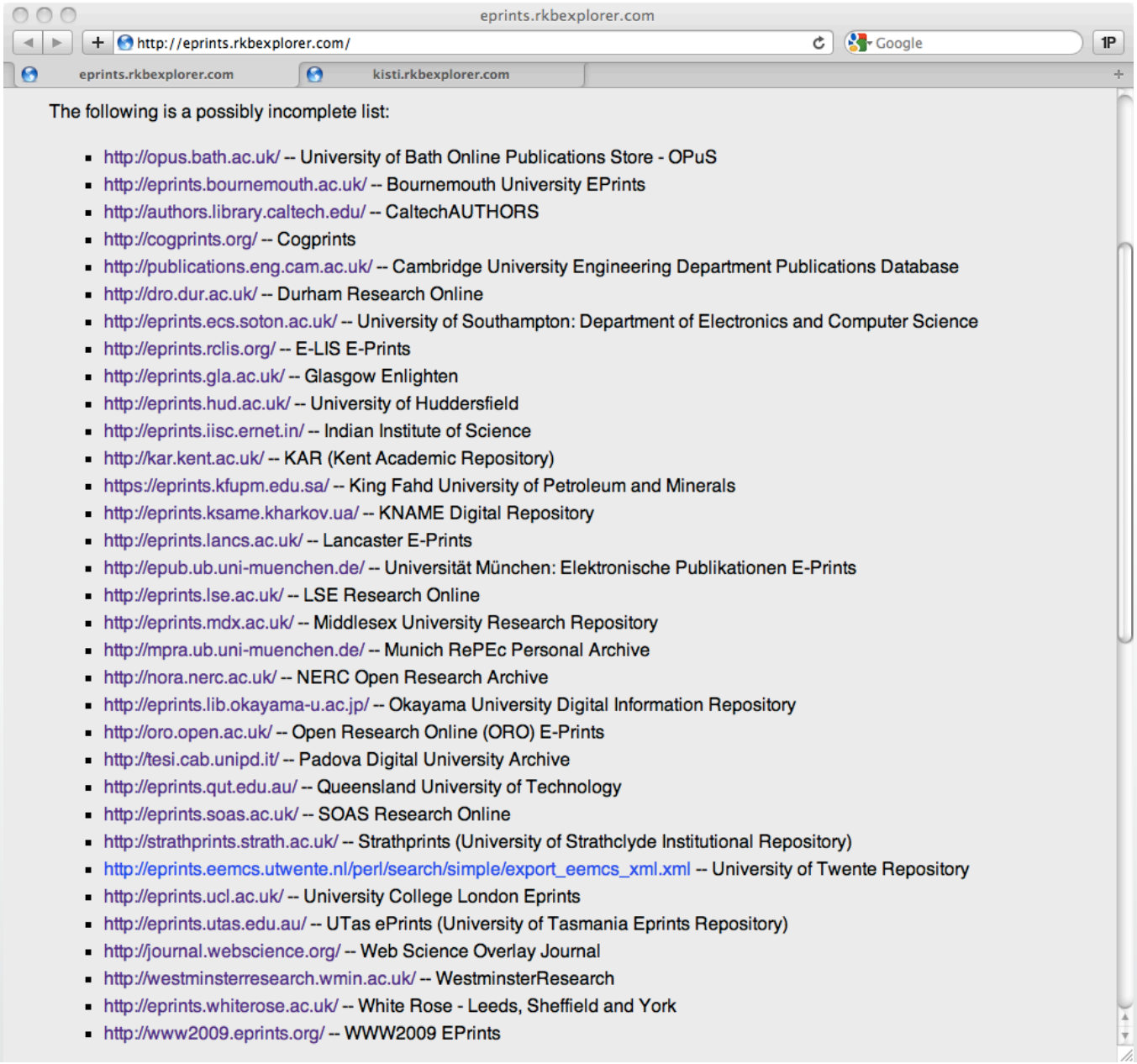
Linked Data Publishing

UNIVERSITY OF Southampton
School of Electronics and Computer Science



ePrints repositories

UNIVERSITY OF Southampton
School of Electronics and Computer Science



The following is a possibly incomplete list:

- http://opus.bath.ac.uk/ -- University of Bath Online Publications Store - OPuS
- http://eprints.bournemouth.ac.uk/ -- Bournemouth University EPrints
- http://authors.library.caltech.edu/ -- CaltechAUTHORS
- http://cogprints.org/ -- Cogprints
- http://publications.eng.cam.ac.uk/ -- Cambridge University Engineering Department Publications Database
- http://dro.dur.ac.uk/ -- Durham Research Online
- http://eprints.ecs.soton.ac.uk/ -- University of Southampton: Department of Electronics and Computer Science
- http://eprints.rclis.org/ -- E-LIS E-Prints
- http://eprints.gla.ac.uk/ -- Glasgow Enlighten
- http://eprints.hud.ac.uk/ -- University of Huddersfield
- http://eprints.iisc.ernet.in/ -- Indian Institute of Science
- http://kar.kent.ac.uk/ -- KAR (Kent Academic Repository)
- https://eprints.kfupm.edu.sa/ -- King Fahd University of Petroleum and Minerals
- http://eprints.ksame.kharkov.ua/ -- KNAME Digital Repository
- http://eprints.lancs.ac.uk/ -- Lancaster E-Prints
- http://epub.ub.uni-muenchen.de/ -- Universität München: Elektronische Publikationen E-Prints
- http://eprints.lse.ac.uk/ -- LSE Research Online
- http://eprints.mdx.ac.uk/ -- Middlesex University Research Repository
- http://mpra.ub.uni-muenchen.de/ -- Munich RePEc Personal Archive
- http://nora.nerc.ac.uk/ -- NERC Open Research Archive
- http://eprints.lib.okayama-u.ac.jp/ -- Okayama University Digital Information Repository
- http://oro.open.ac.uk/ -- Open Research Online (ORO) E-Prints
- http://tesi.cab.unipd.it/ -- Padova Digital University Archive
- http://eprints.qut.edu.au/ -- Queensland University of Technology
- http://eprints.soas.ac.uk/ -- SOAS Research Online
- http://strathprints.strath.ac.uk/ -- Strathprints (University of Strathclyde Institutional Repository)
- http://eprints.eemcs.utwente.nl/perl/search/simple/export_eemcs_xml.xml -- University of Twente Repository
- http://eprints.ucl.ac.uk/ -- University College London Eprints
- http://eprints.utas.edu.au/ -- UTas ePrints (University of Tasmania Eprints Repository)
- http://journal.webscience.org/ -- Web Science Overlay Journal
- http://westminsterresearch.wmin.ac.uk/ -- WestminsterResearch
- http://eprints.whiterose.ac.uk/ -- White Rose - Leeds, Sheffield and York
- http://www2009.eprints.org/ -- WWW2009 EPrints

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”

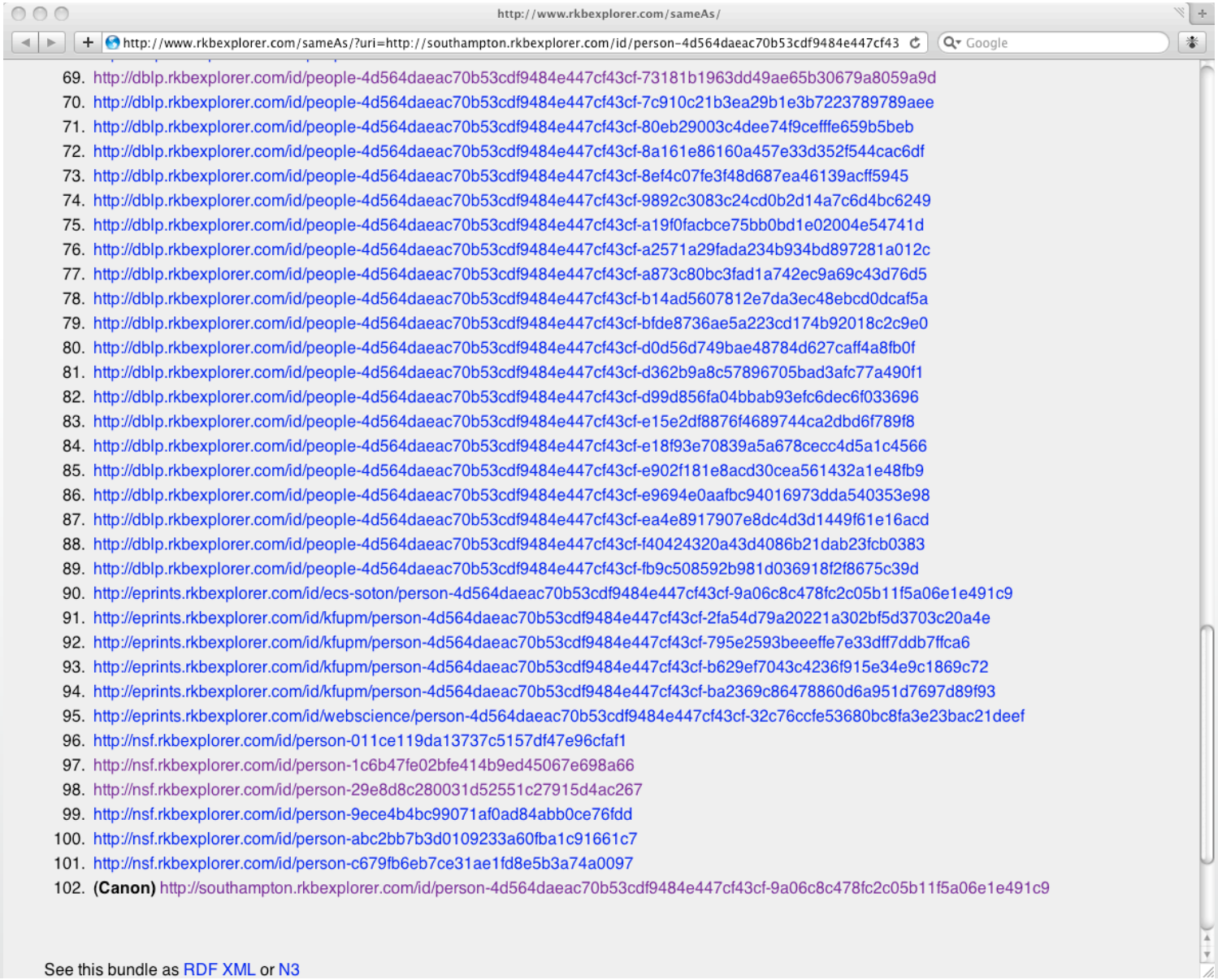
Co-Reference

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

Who is Carl Lagoze?

UNIVERSITY OF Southampton
School of Electronics and Computer Science



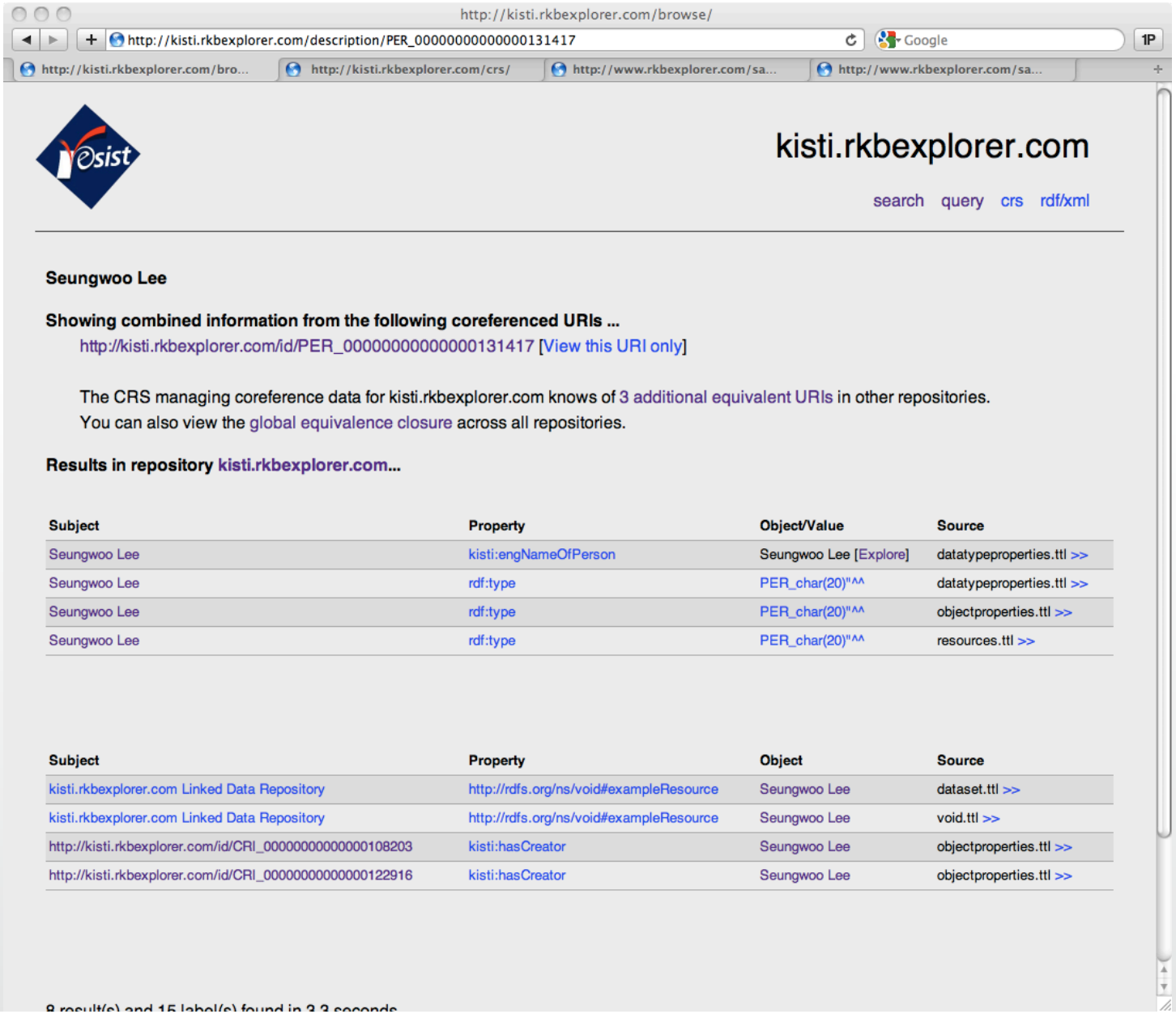
UNIVERSITY OF Southampton
School of Electronics and Computer Science

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

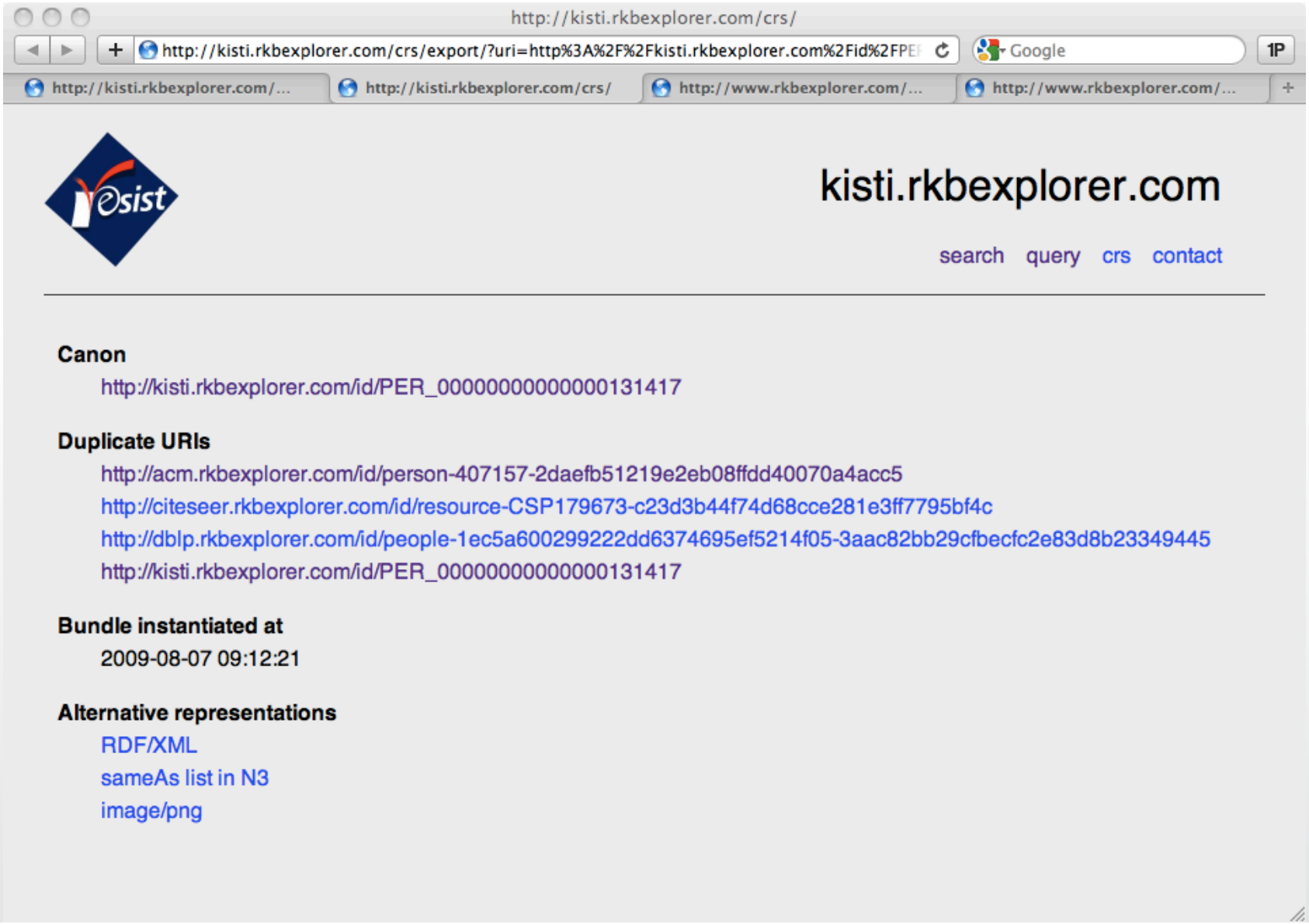
KISTI Browse

UNIVERSITY OF Southampton
School of Electronics and Computer Science



KISTI CRS

UNIVERSITY OF Southampton
School of Electronics and Computer Science



Co-Reference Closure

UNIVERSITY OF Southampton
School of Electronics and Computer Science

Complete Co-Reference Information

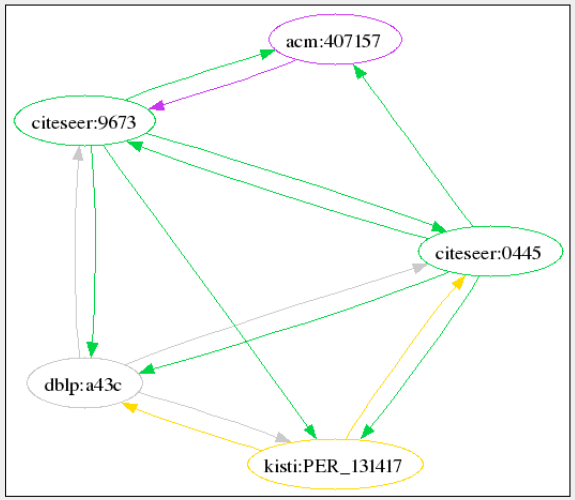
This service computes the equivalence class within the known URIs for a specified URI, by consulting all relevent CRS knowledge bases.

<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 ...

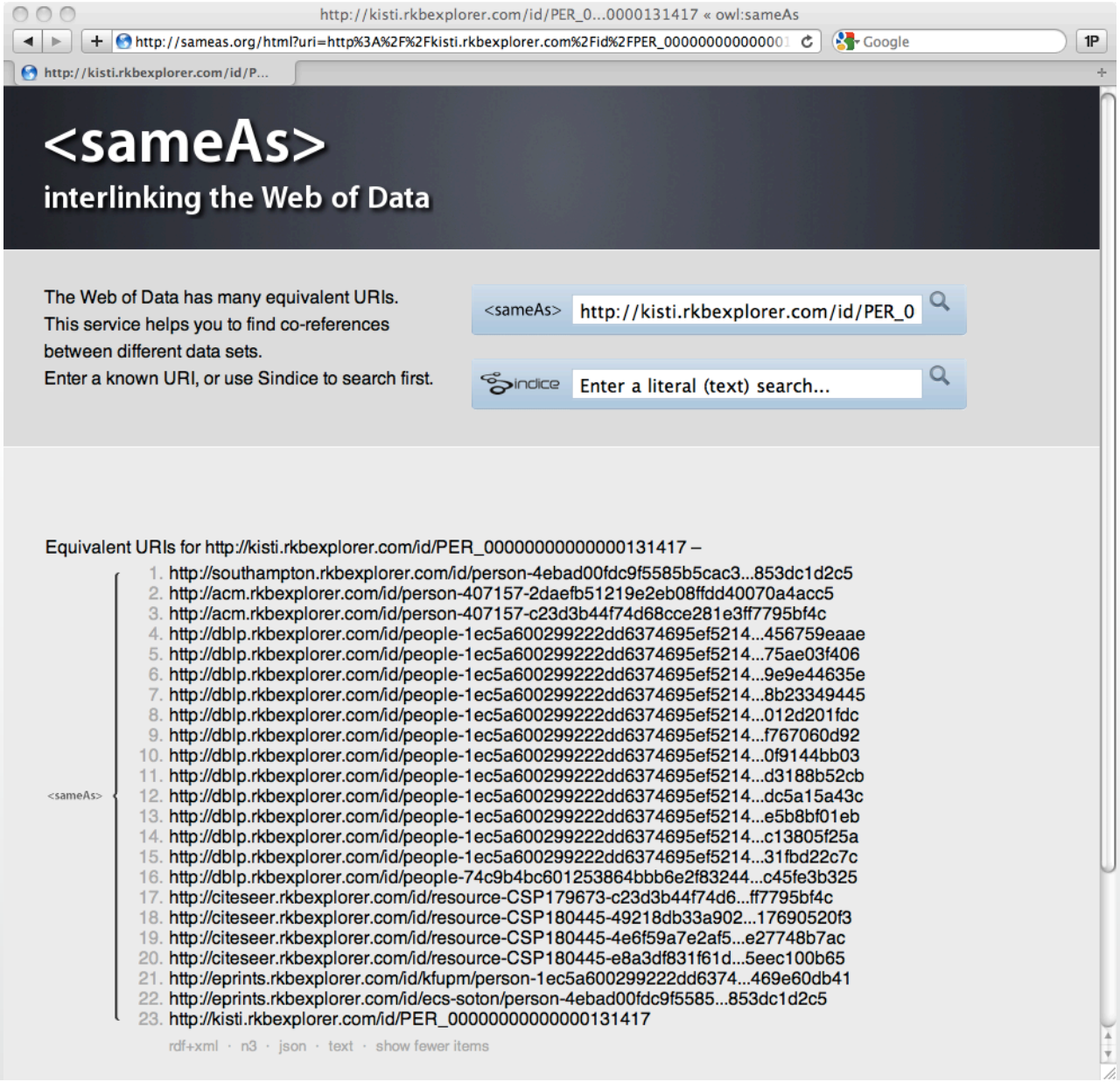
Showing information queried from all repositories ...

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-

Global Co-reference

UNIVERSITY OF Southampton
School of Electronics and Computer Science



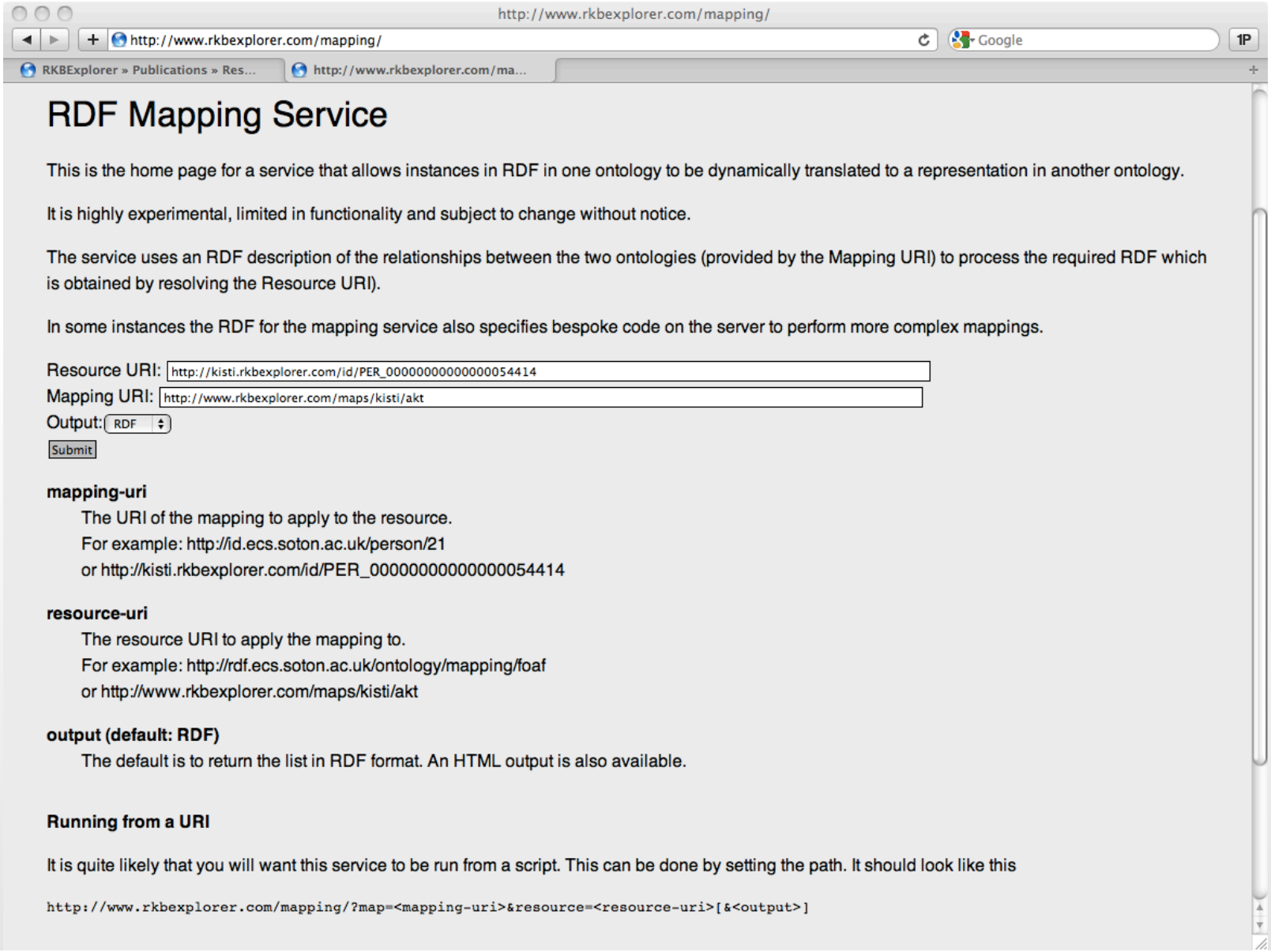
UNIVERSITY OF Southampton
School of Electronics and Computer Science

Dealing With Different Ontologies

- The RKBExplorer application uses particular ontologies
  - Mostly AKT Reference Ontology
  - Some KBs will use different ontologies
  - Eg kisti.rkbexplorer.com
  - Uses KISTI Research Ontology
- One solution
  - Represent the ontology mapping in RDF (as far as possible)
  - Resolve the URI through a mapping service to get RDF in the required ontology

On-The-Fly Translation

UNIVERSITY OF Southampton
School of Electronics and Computer Science



UNIVERSITY OF Southampton
School of Electronics and Computer Science

Open System

- RKBExplorer is only one interface
  - And not a required part
- Other Interfaces (using the services)
  - Either as RDF, or JSON, csv, etc.
- Services
  - http://www.rkbexplorer.com/demos/

UNIVERSITY OF Southampton
School of Electronics and Computer Science

Open System

- 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

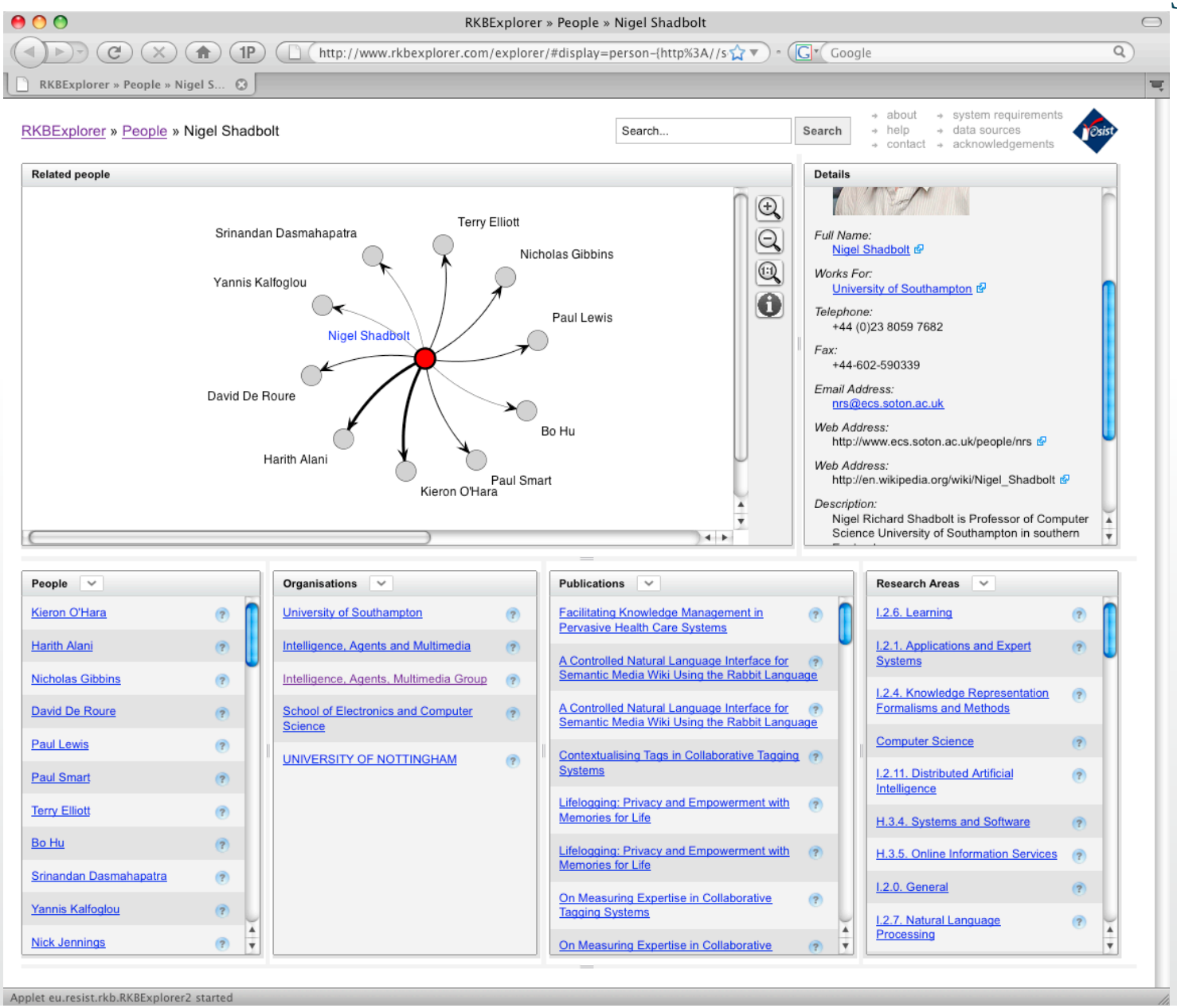
UNIVERSITY OF Southampton
School of Electronics and Computer Science

Concluding Remarks

- Successful data integration
- Much added value
- Linked Data standards made it possible
- Worry about your co-reference

http://eprints.ecs.soton.ac.uk/17761/

UNIVERSITY OF Southampton
School of Electronics and Computer Science

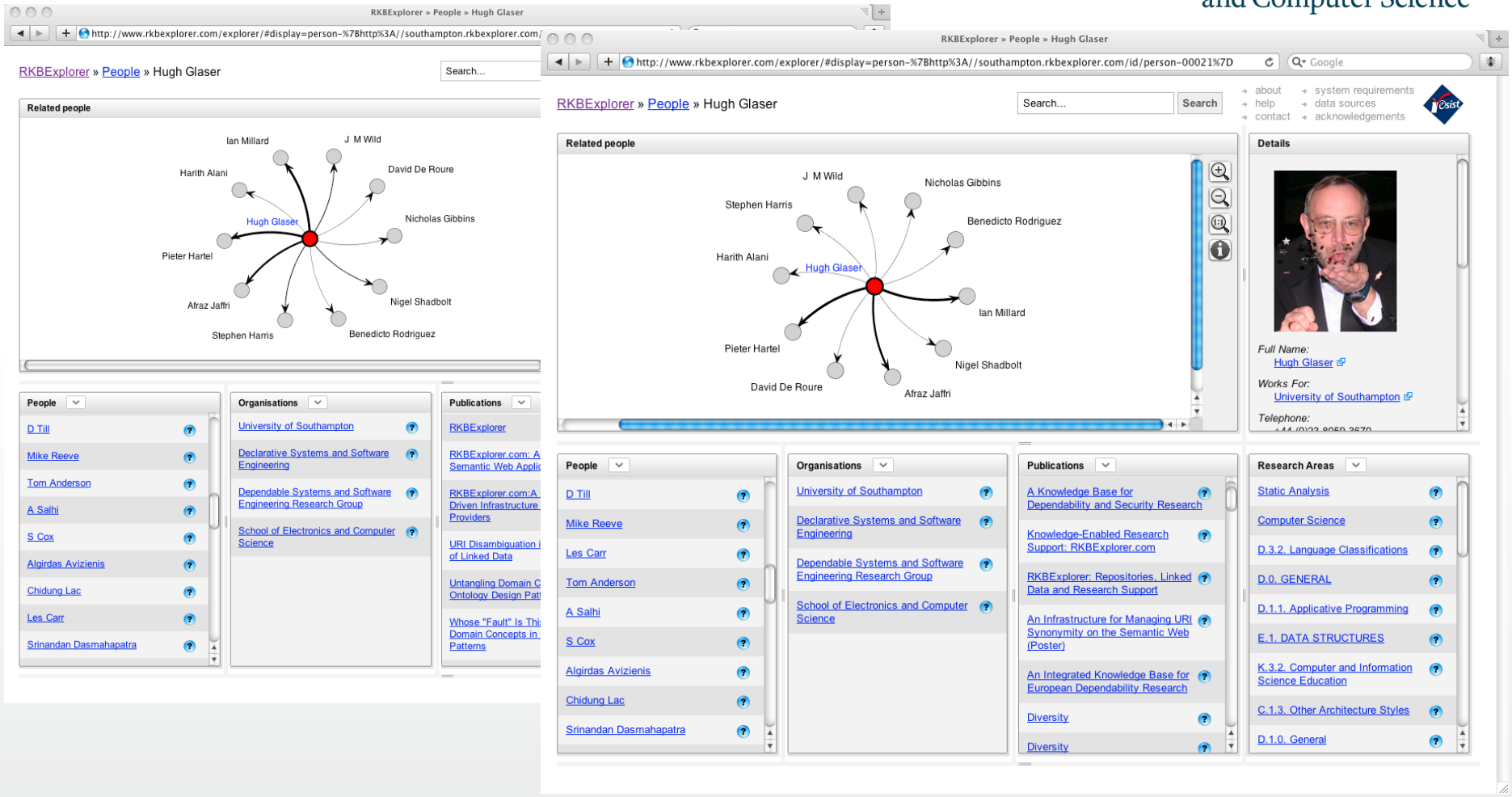


RKBExplorer.com

Try it!

Update in Action

UNIVERSITY OF Southampton
School of Electronics and Computer Science

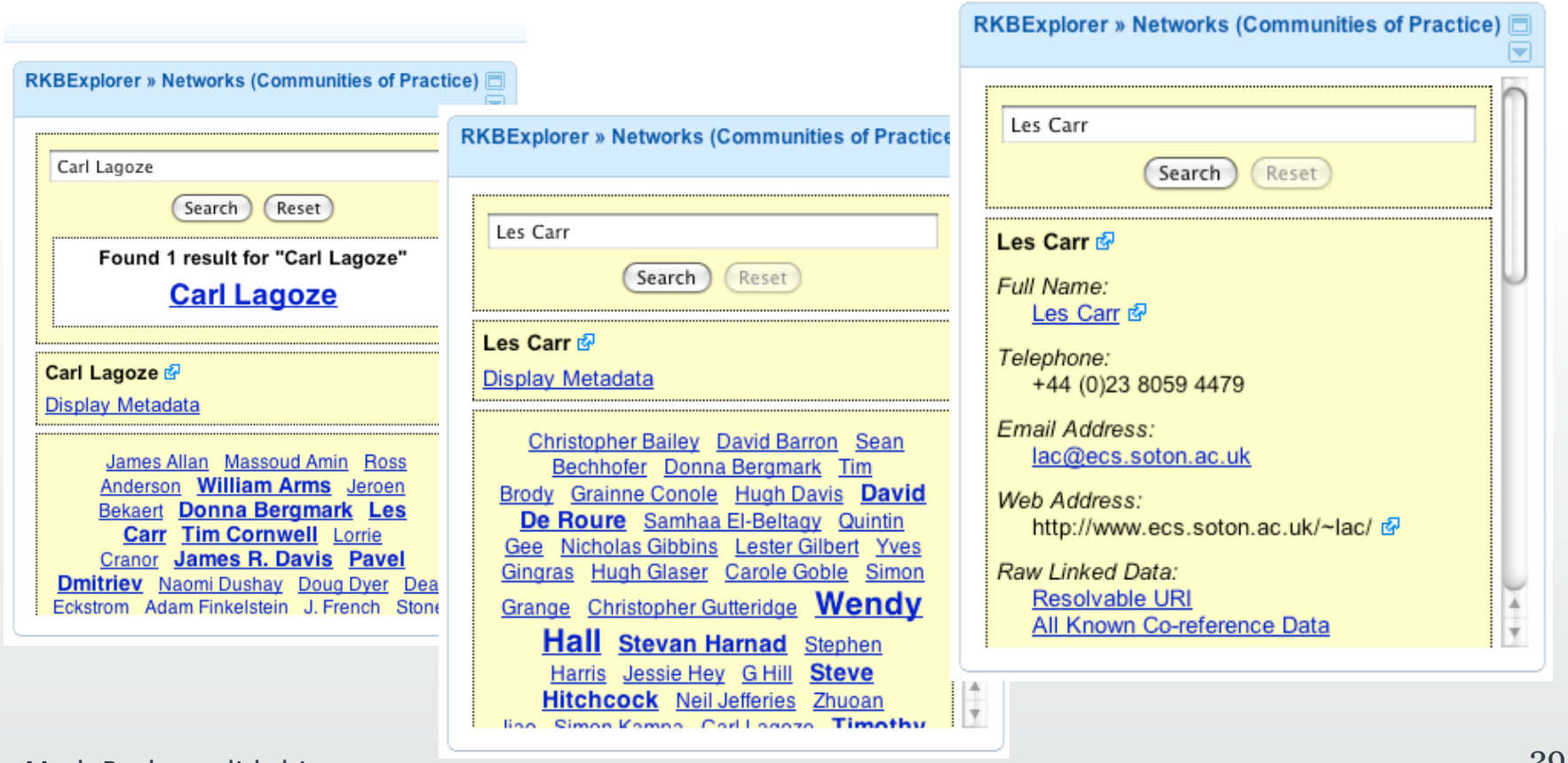


Before and after inserting this paper in the Southampton ePrints repository and RKB has noticed
Note the position of Les Carr in Hugh's related People

UNIVERSITY OF Southampton
School of Electronics and Computer Science

Gadget – find out about people

iGoogle



Mark Borkum did this