UNIVERSITY OF Southampton
School of Electronics and Computer Science

# RKB, sameAs and dotAC

at
"2009: Beyond the Repository Fringe"

Edinburgh
30-31 July 2009

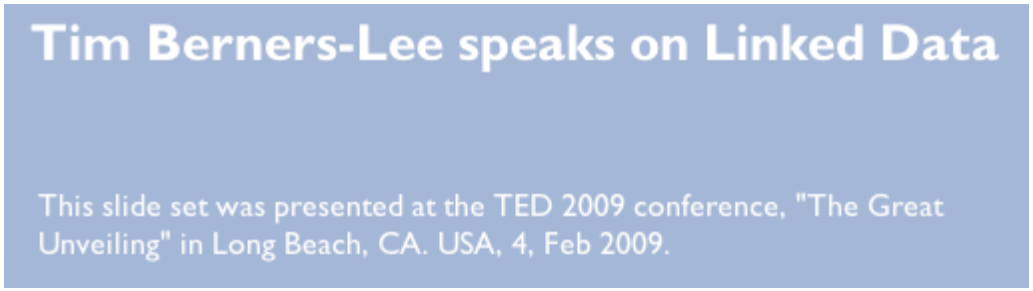
Hugh Glaser & Ian Millard



UNIVERSITY OF Southampton
School of Electronics and Computer Science

# Linked Data

- Tim Berners-Lee




  - http://www.w3.org/2009/Talks/0204-ted-tbl/
  - "the Semantic Web done right, and the Web done right"
- http://esw.w3.org/topic/SweoIG/TaskForces/CommunityProjects/LinkingOpenData

LINKINGOPENDATA
W3C SWEO Community Project

UNIVERSITY OF Southampton
School of Electronics and Computer Science

# Design Issues - Linked Data

*Tim Berners-Lee*
*Date: 2006-07-27, last change: $Date: 2007/05/02 14:30:56 $*
*Status: personal view only. Editing status: imperfect but published.*

- http://www.w3.org/DesignIssues/LinkedData.html

- Use URIs as names for things
- Use HTTP URIs so that people can look up those names.
- When someone looks up a URI, provide useful information.
- Include links to other URIs. so that they can discover more things.

UNIVERSITY OF Southampton
School of Electronics and Computer Science

# RDF

- <subject-uri> <predicate-uri> <object-uri>

  or

- <subject-uri> <predicate-uri> “String”
- Eg

@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .

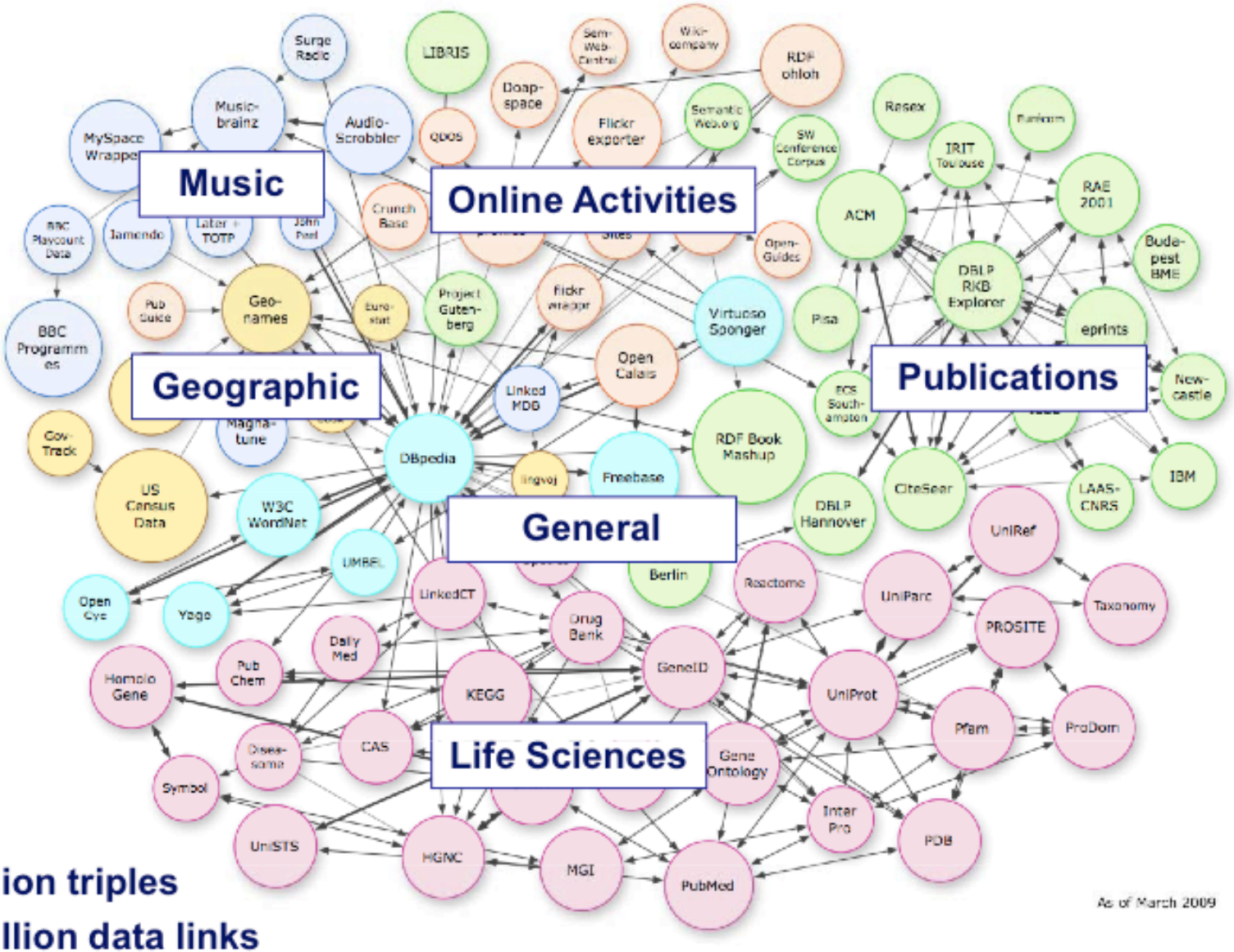
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .

<http://resex.rkbexplorer.com/id/resilience-concept-de7f20d5> rdf:type <http://resex.rkbexplorer.com/ontologies/resex#Resilience-Concept> .

<http://resex.rkbexplorer.com/id/resilience-concept-de7f20d5> rdfs:label "UML model" .

<http://resex.rkbexplorer.com/id/resilience-concept-de7f20d5> <http://resex.rkbexplorer.com/ontologies/resex#has-description> "The UML modeling language describes the system architecture and behaviour in a standard way." .

# LOD Datasets on the Web: March 2009




**4.5 billion triples**

**180 million data links**

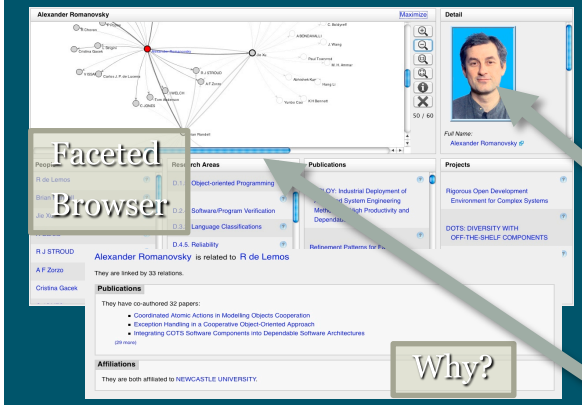
As of March 2009

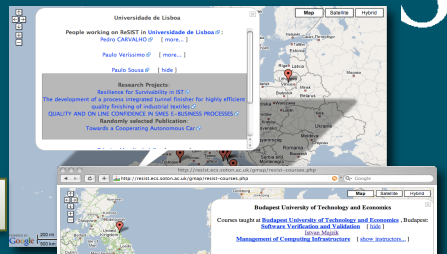
# Knowledge-Enabled Research Support: RKBExplorer.com

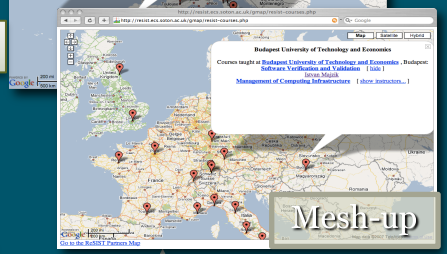
Ian Millard & Hugh Glaser, {icm,hg}@ecs.soton.ac.uk

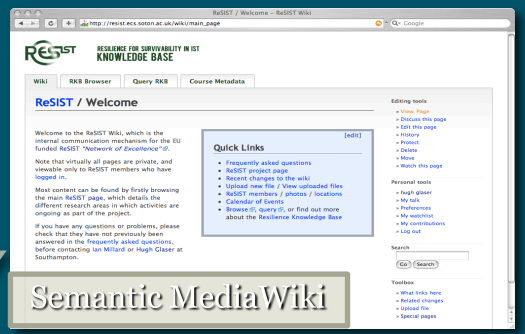
UNIVERSITY OF Southampton

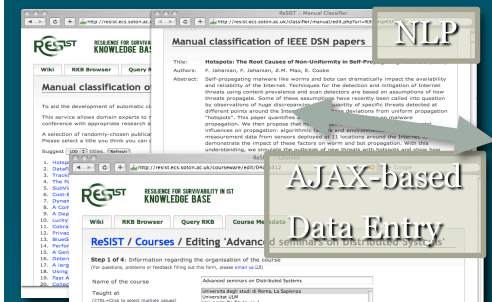
School of Electronics and Computer Science

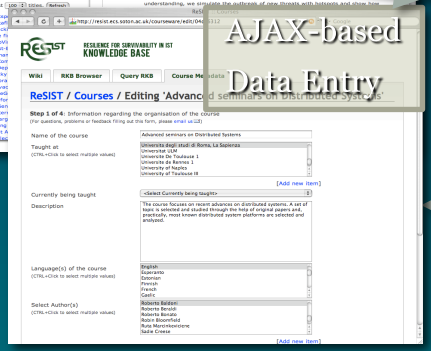


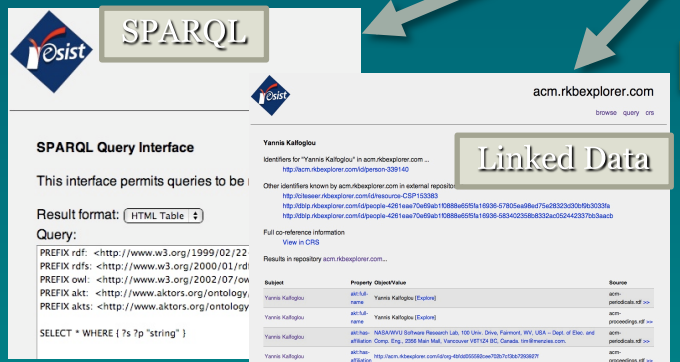




3store

CRS

Coreference Analysis

Offline Conversion & Versioning







Ontology Mapping Service


Resolving URIs KISTI , dbpedia, etc.

Citeseer, CORDIS, DBLP, Partners, UN LoCode, ...

We are pleased to acknowledge the help and work provided by our ReSIST Partners and many others.
This work was supported by the ReSIST Network of Excellence, which is sponsored by the Information Society Technology (IST) priority under contract number IST 4 026764 NOE.

UNIVERSITY OF Southampton
School of Electronics and Computer Science

# Some Underlying Sources

| | |
|---|---|
| **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”

UNIVERSITY OF Southampton
School of Electronics and Computer Science

# Knowledge Sources

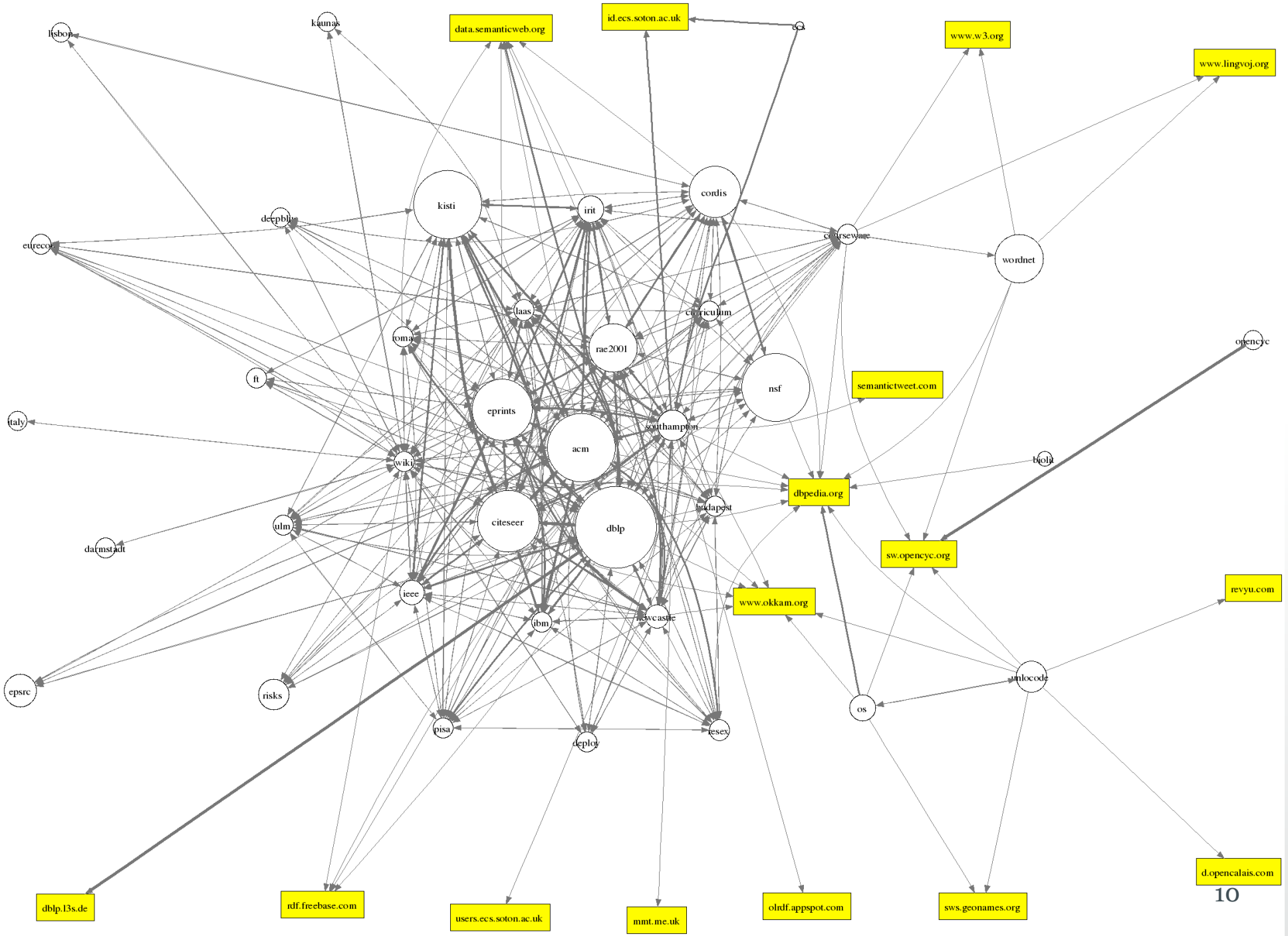
- Partners
- Publications
- Funding Agencies
- Project Wiki
- Courseware
- Resilient-Explicit Computing

- Wide range, don't just look where you expect to find

UNIVERSITY OF Southampton
School of Electronics and Computer Science

# For example

- Statistics for repository kisti.rkbexplorer.com
  - Last data assertion 2008-09-18 17:16:41
  - Number of triples 12815162
  - Number of symbols 3239105
  - Size of RDF dataset 671M


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

# 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

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

UNIVERSITY OF Southampton
School of Electronics and Computer Science

# CRS – Consistent Reference Service

- A service to manage and publish co-referent information
- Identify co-referent pairs using a set of tools
- Assert into the CRS
- Query the CRS
  - $URI_i$ -> { $URI_1$, ..., $URI_i$, ..., $URI_n$ }
- Recommend a Canon

UNIVERSITY OF Southampton
School of Electronics and Computer Science

# CRS continued

- CRS Policies are defined by context
  - Often one per Triplestore
  - Can be many per Triplestore for different purposes
  - May not be associated with a particular Triplestore
- Maintenance
  - Provenance
  - Rollback
- Can be used to infer owl:sameAs

# Open System

UNIVERSITY OF Southampton
School of Electronics and Computer Science

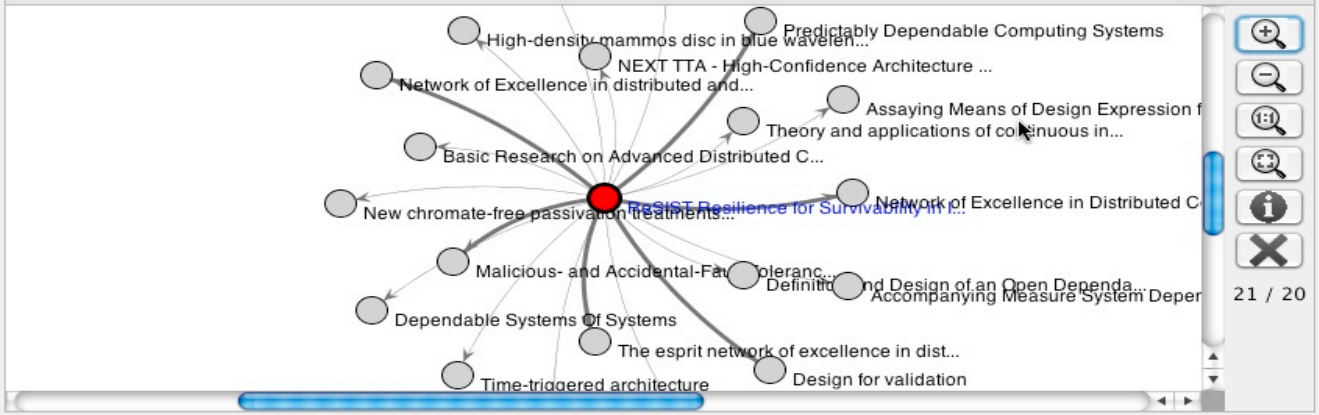
- RKBExplorer is only one interface
  - And not a required part
- Services:
  - Details for a paper (the right hand pane in RKBExplorer):
    - http://www.rkbexplorer.com/detail/?uri=http://southampton.rkbexplorer.com/id/eprints-12614
  - Network of people for a publication (lower pane):
    - http://www.rkbexplorer.com/network/?uri=http://southampton.rkbexplore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

# www.rkbexplorer.com/explorer/



## Resilience for Survivability in IST

Maximize




## Detail

*Title:*
Resilience for Survivability in IST

*Other Names:*
ReSIST Resilience for Survivability in IST

*Funding:*
4500000

*Project Leader:*
Jean-Claude Laprie

*Leading Organisation:*
Centre National de la recherche Scientifique

## People

- Jean-Claude Laprie
- Aad van Moorsel
- Abdelmajid Khelil
- Achour Mostefaou
- Adnan Noon Mian
- Afraz Jaffri
- Agnan de Bonneval
- Alberto Pasquini

## Research Areas

- Information processing, information systems
- Development Failure
- Dependability And Security Analysis
- Dependability And Security Provision

## Publications

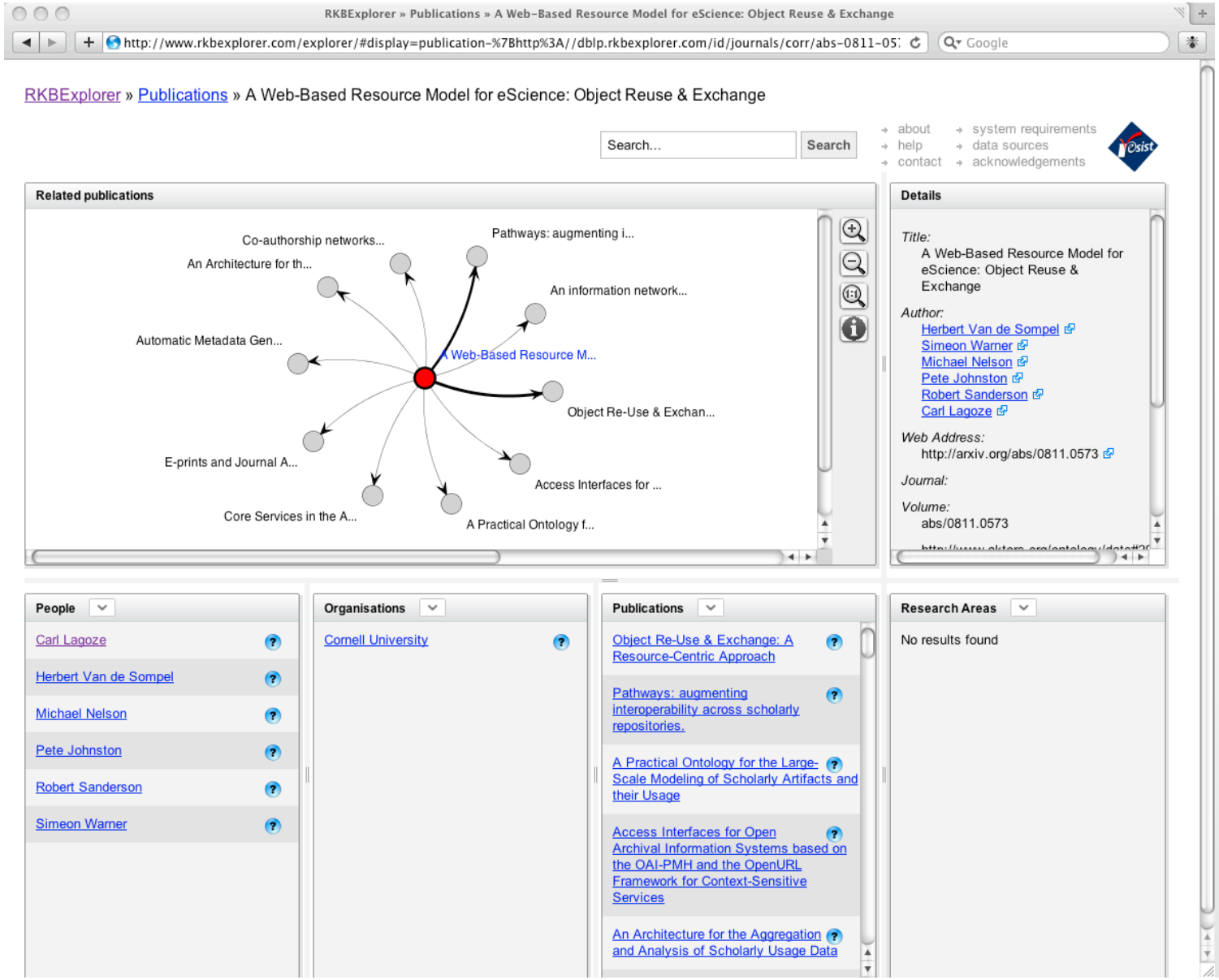
- Revised version in Predictably dependable computing systems
- ASSERT Automated proof based system and software engineering for real-time applications
- Guide de la sûreté de fonctionnement

## Projects

- Malicious- and Accidental-Fault Tolerance for Internet Applications
- The esprit network of excellence in distributed computing systems architectures
- Design for validation
- Network of Excellence in

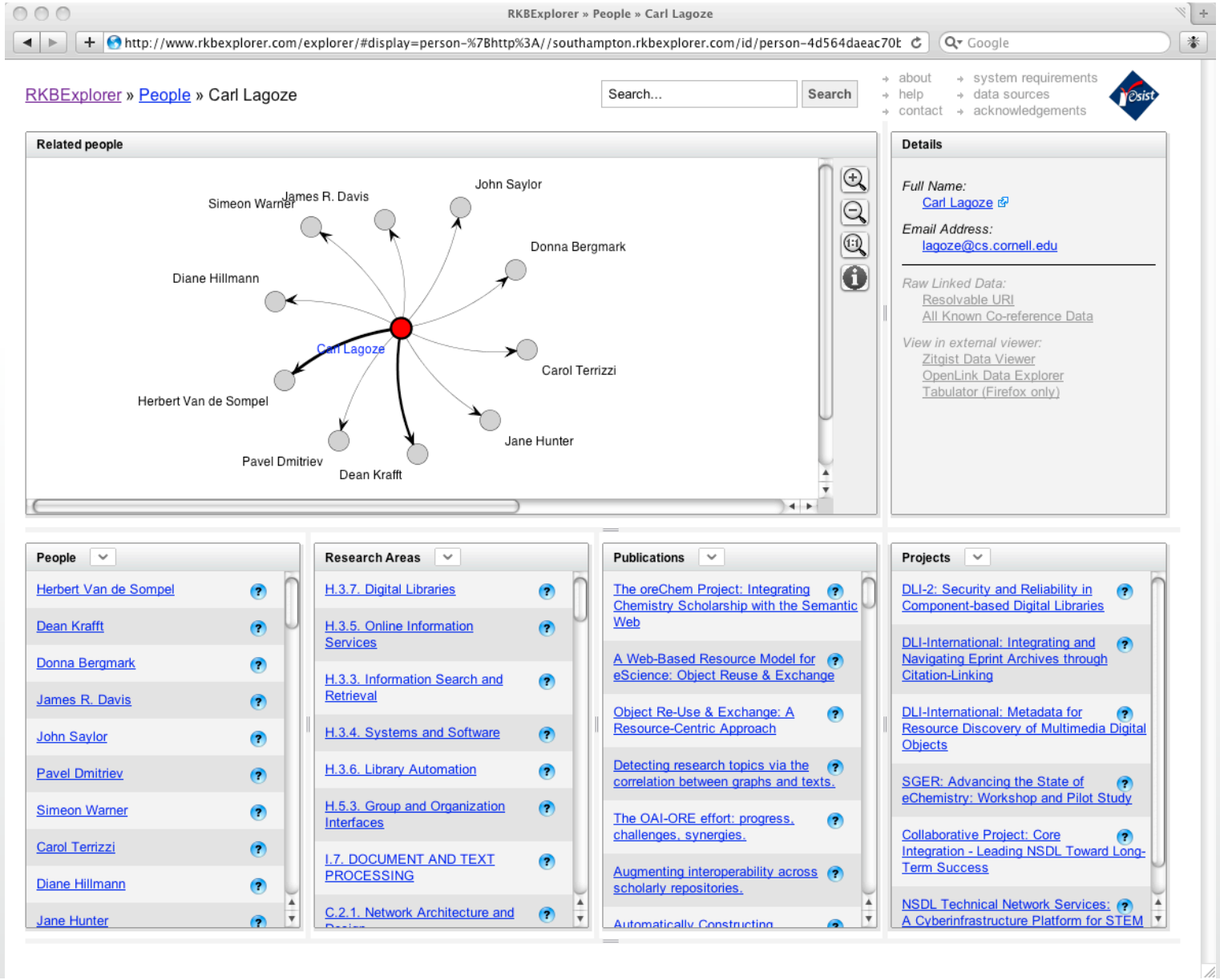
# Or a Paper

UNIVERSITY OF Southampton
School of Electronics and Computer Science



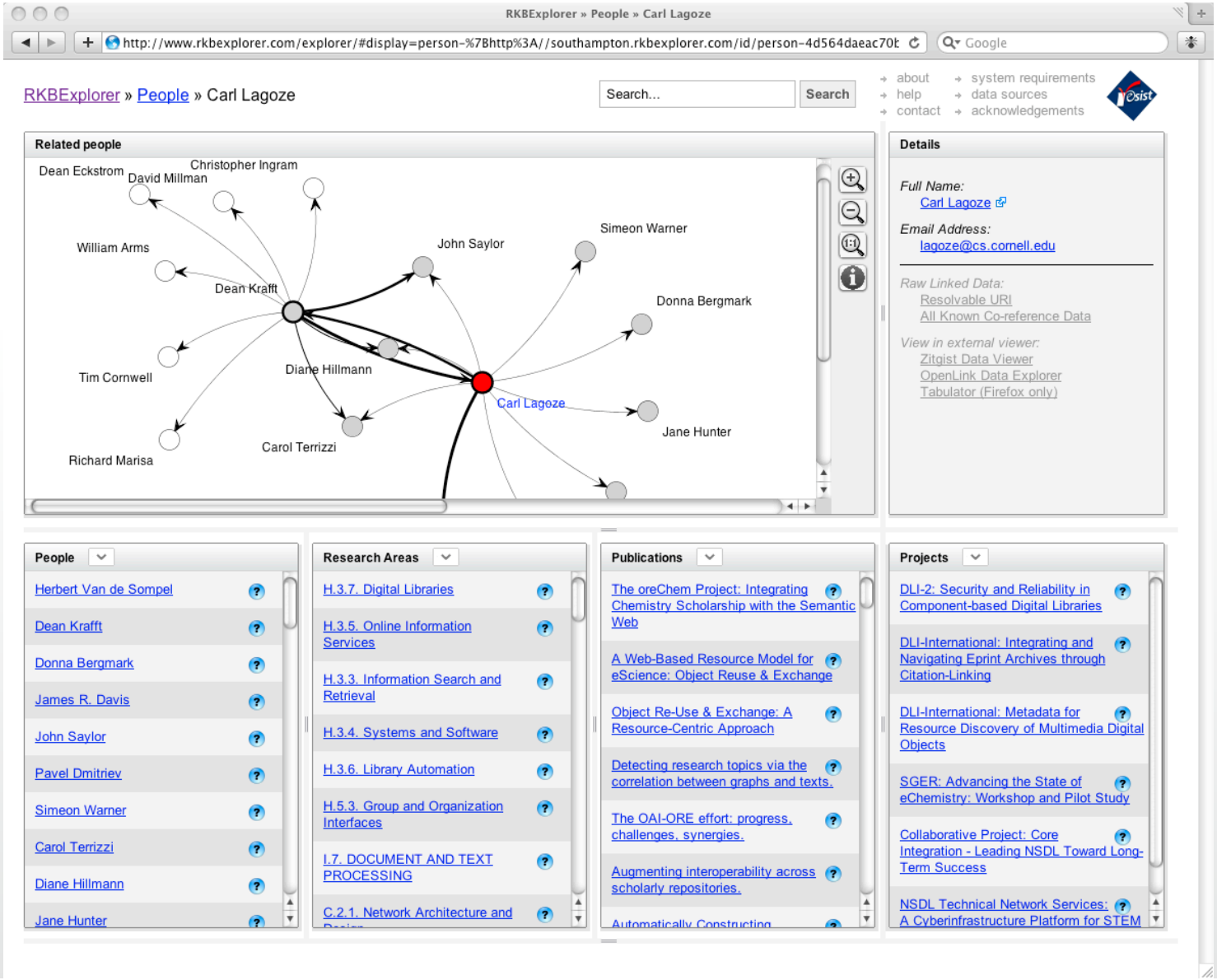
# Now Look at an Author

UNIVERSITY OF Southampton
School of Electronics and Computer Science



# Or a Couple of People

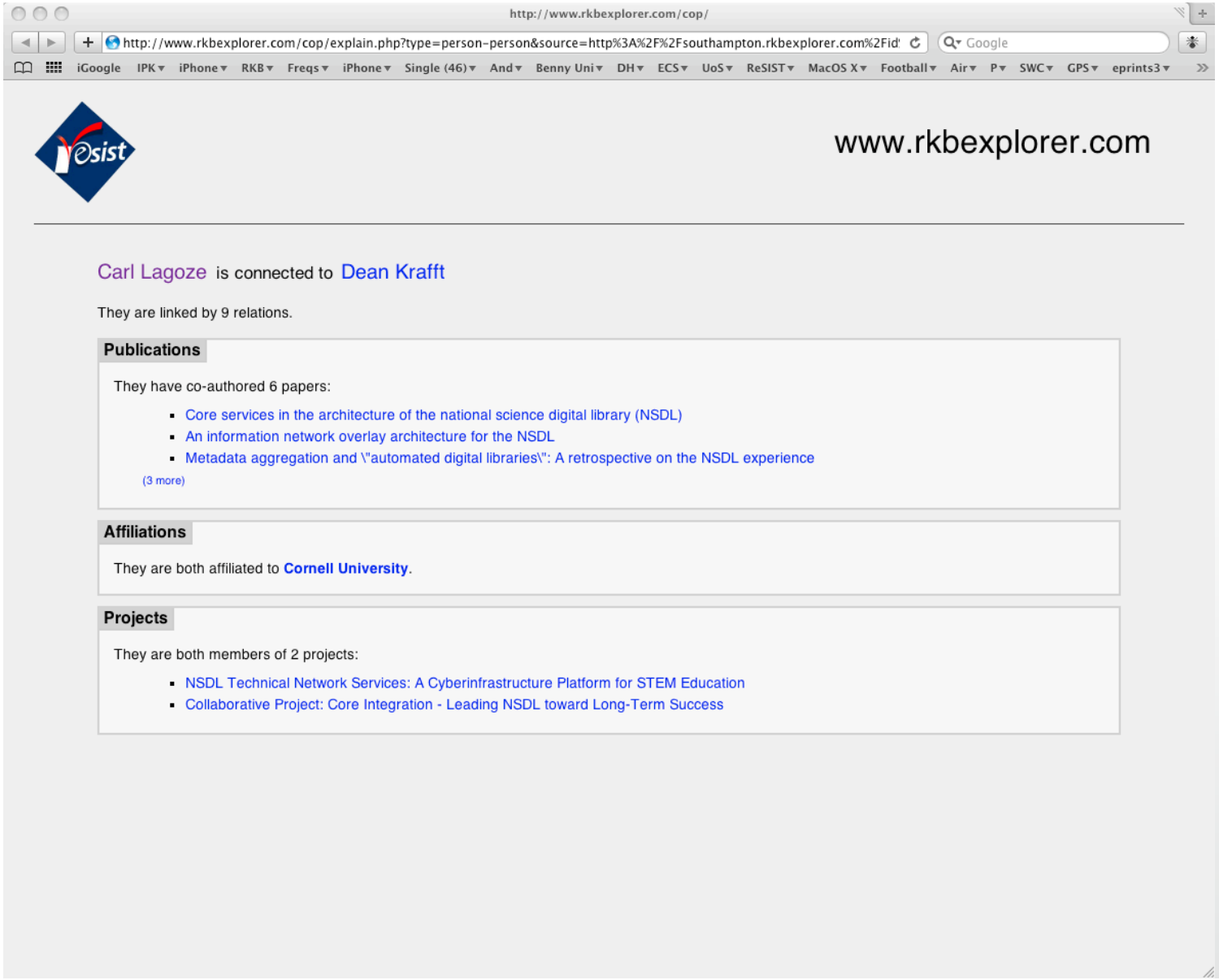
UNIVERSITY OF Southampton
School of Electronics and Computer Science




And how they are linked

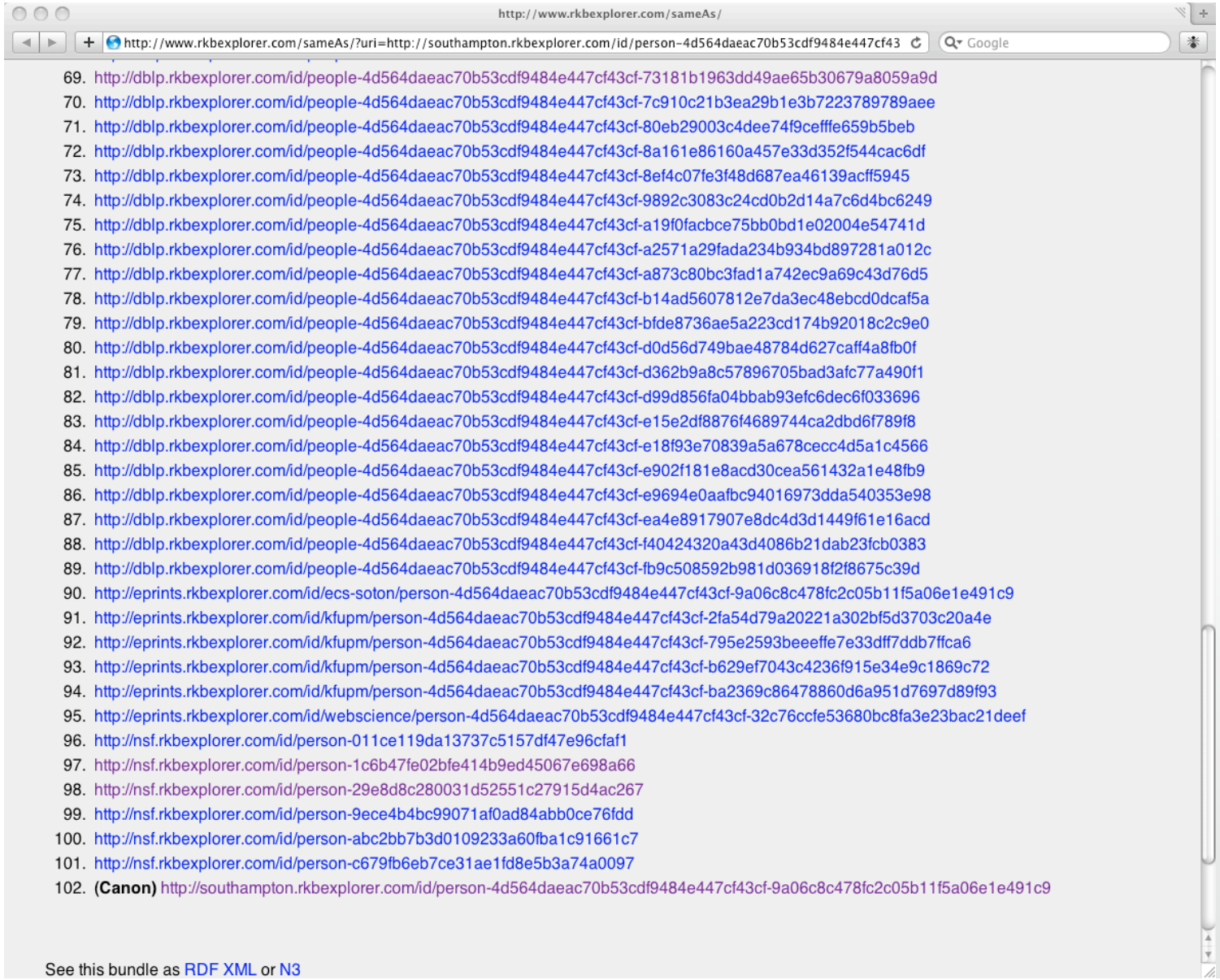
# And Why they are Linked?

UNIVERSITY OF Southampton
School of Electronics and Computer Science



# Who is Carl Lagoze?

UNIVERSITY OF Southampton
School of Electronics and Computer Science



# Co-Reference Closure

UNIVERSITY OF Southampton
School of Electronics and Computer Science

**Complete Co-Reference Information**

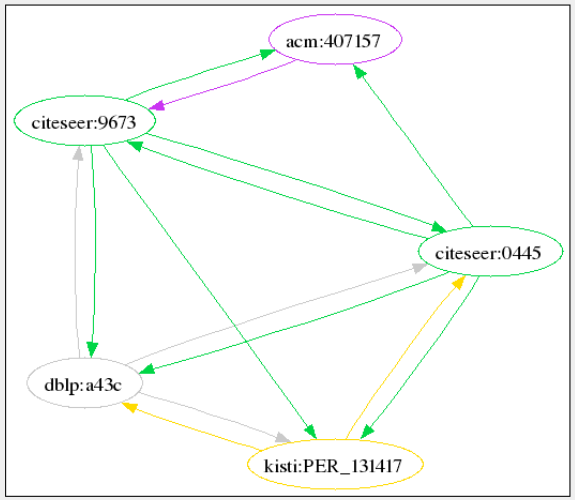
This service computes the equivalence class within the known URIs for a specified URI, by consulting all relevent CRS knowledge bases.

<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 ...**

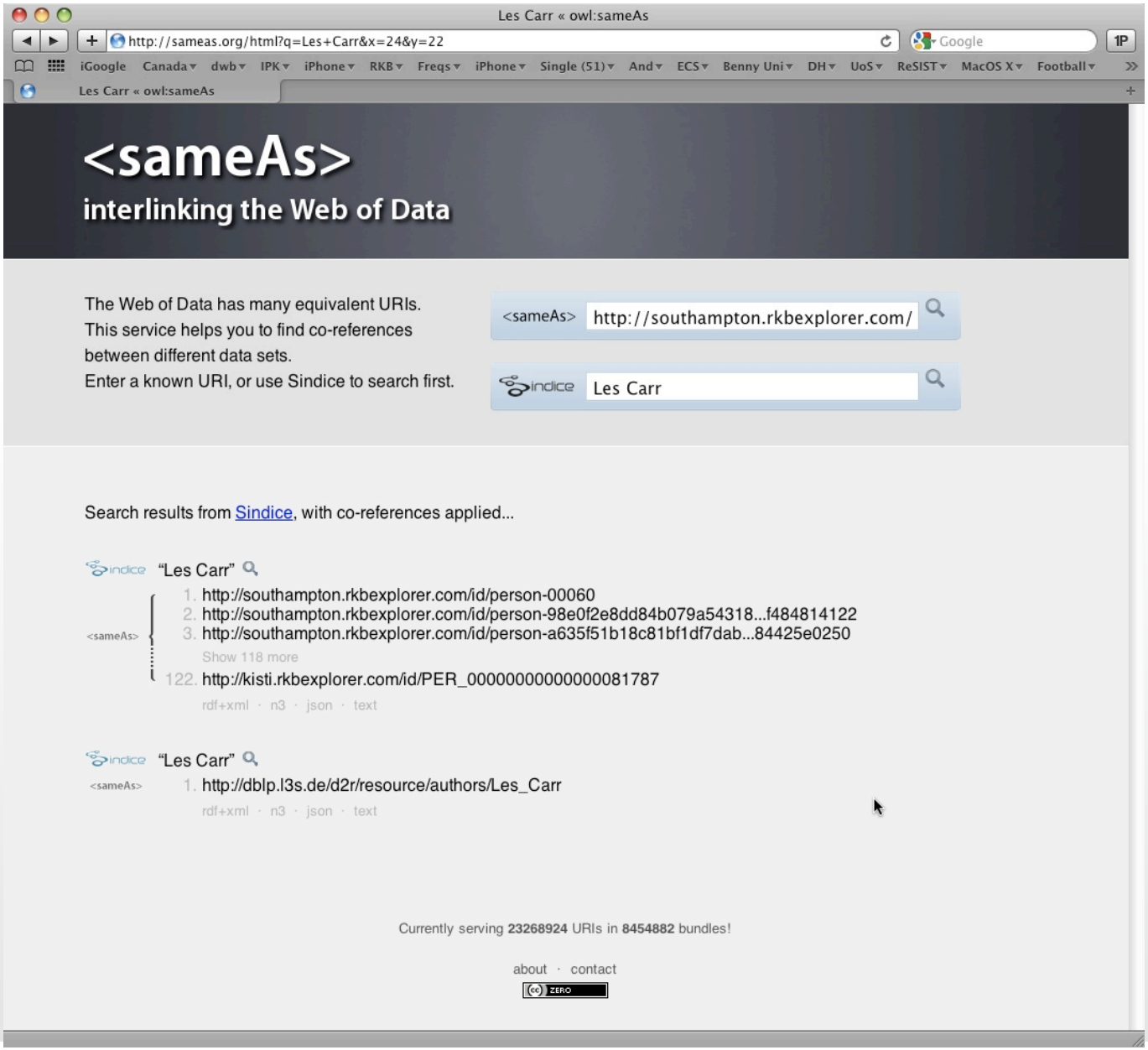
**Showing information queried from all repositories ...**

| 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- |

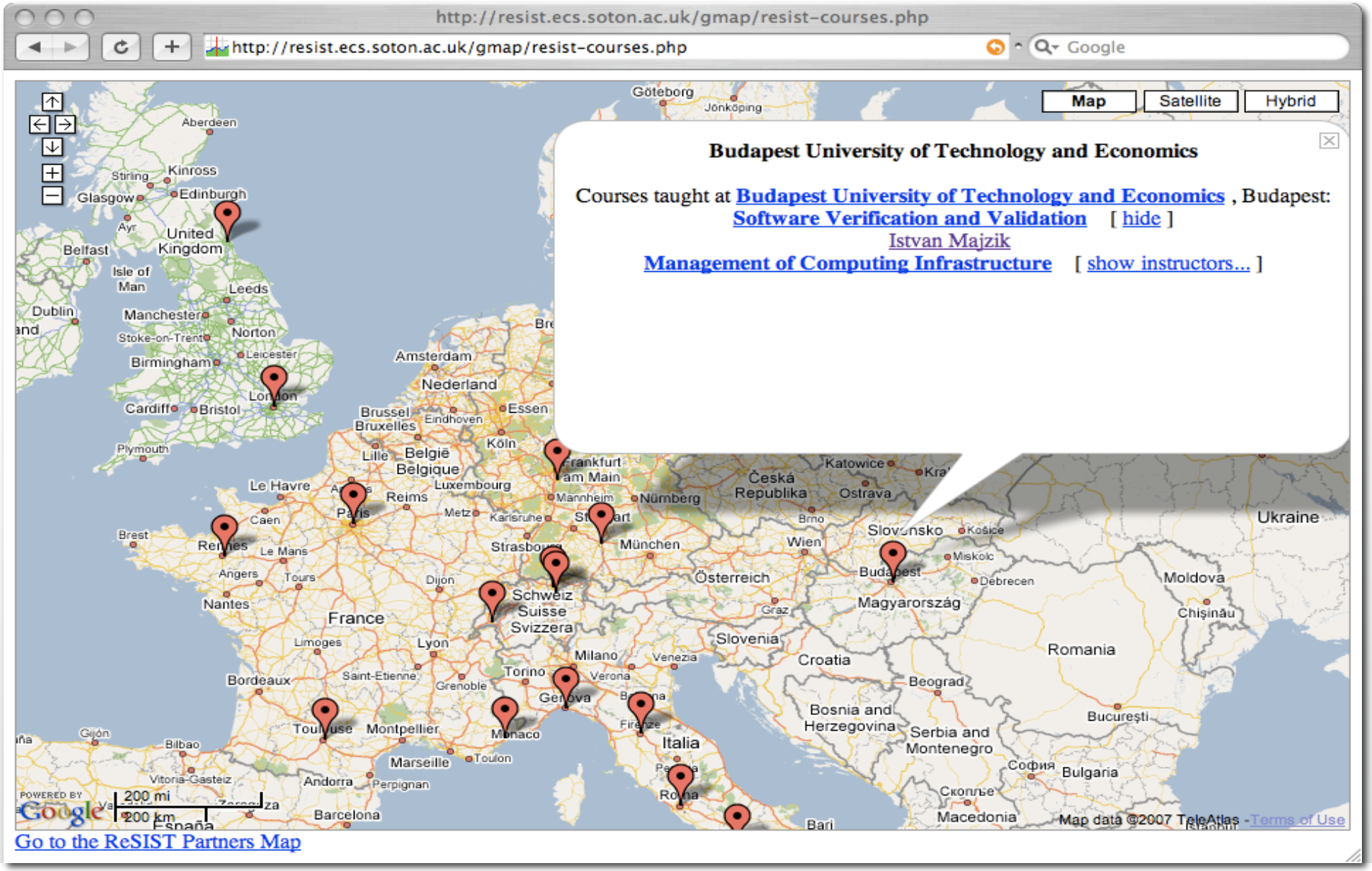
# Finding Co-reference

UNIVERSITY OF Southampton
School of Electronics and Computer Science



# Where is it Taught?

UNIVERSITY OF Southampton
School of Electronics and Computer Science

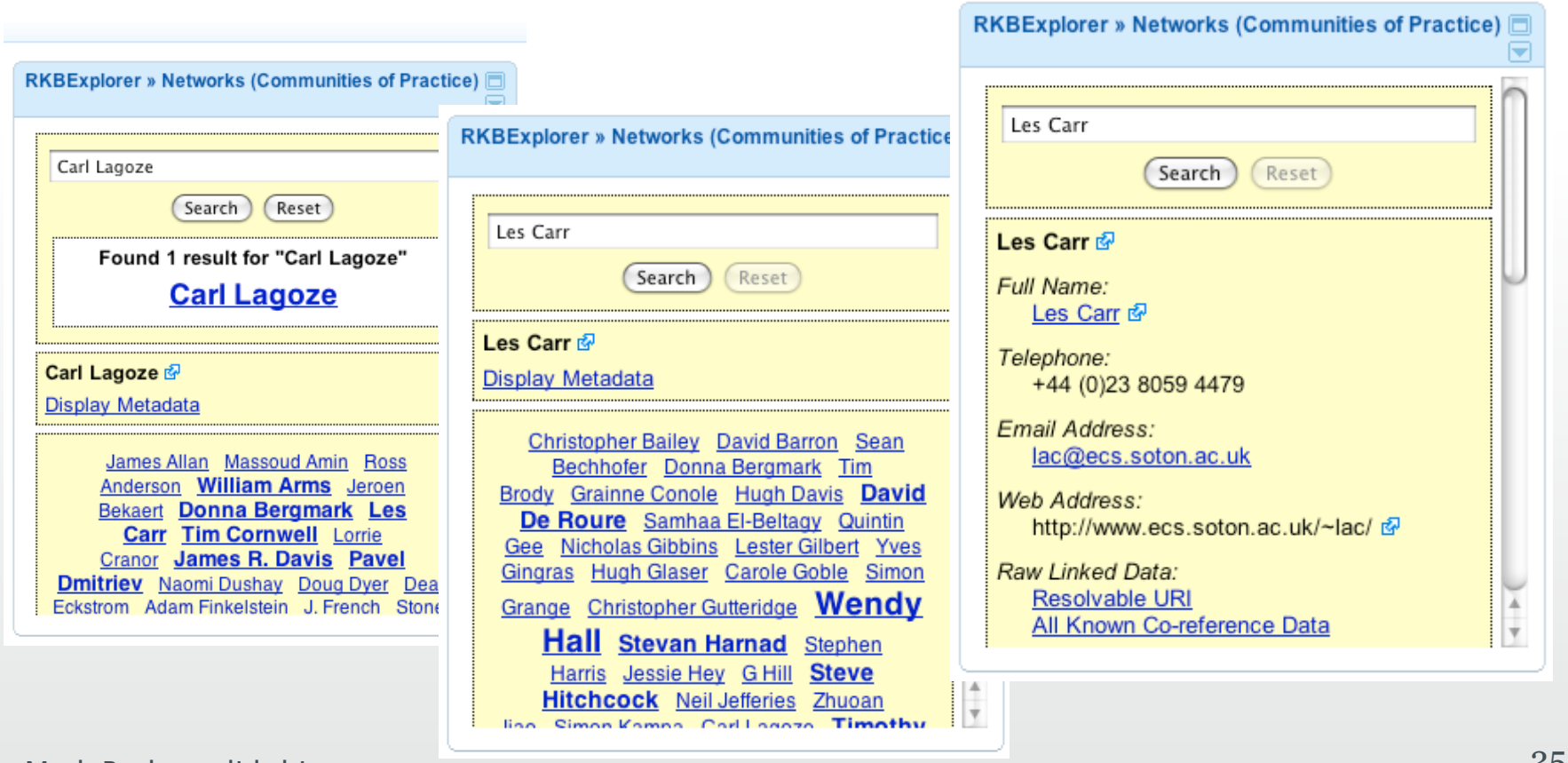



Go to the ReSIST Partners Map

UNIVERSITY OF Southampton
School of Electronics and Computer Science

# Gadget – find out about people

iGoogle




Mark Borkum did this

This is a page that gives a simple demonstration showing papers which have been deemed related through textual analysis by IAI Saarbrucken. Up to the top 20 are listed for each paper, when they meet a simple thresholding:

1 – very strong – 0.9 – strongly – 0.7 – related – 0.6 – ignored – 0

The 1980 paper **Exception Handling and Software-Fault Tolerance** [browse]

is very strongly related to

- [browse] 2003 "Automatic detection and masking of non-atomic exception handling" [PDF]
- [browse] 1989 "Formal Verification of Programs with Exceptions"
- [browse] 1983 "Programming Reliable and Robust Software in ADA"

is strongly related to

- [browse] 1998 "Improving software robustness with dependability cases" [PDF]
- [browse] 1999 "Wrapping windows NT software for robustness" [PDF]
- [browse] 1981 "Exception Handling and Error Recovery Techniques in Modular Systems - An Application to the Isaure System"
- [browse] 2003 "Deadlock resolution via exceptions for dependable Java applications" [PDF]
- [browse] 2002 "Robust software - no more excuses" [PDF]

is related to

- [browse] 1995 "Fault tolerance in concurrent object-oriented software through coordinated error recovery" [PDF]
- [browse] 2004 "Implementing simple replication protocols using CORBA portable interceptors and Java serialization" [PDF]
- [browse] 1984 "Fault Tolerance Using Communicating Sequential Processes"
- [browse] 2001 "Middleware support for voting and data fusion" [PDF]

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?

UNIVERSITY OF Southampton
School of Electronics and Computer Science

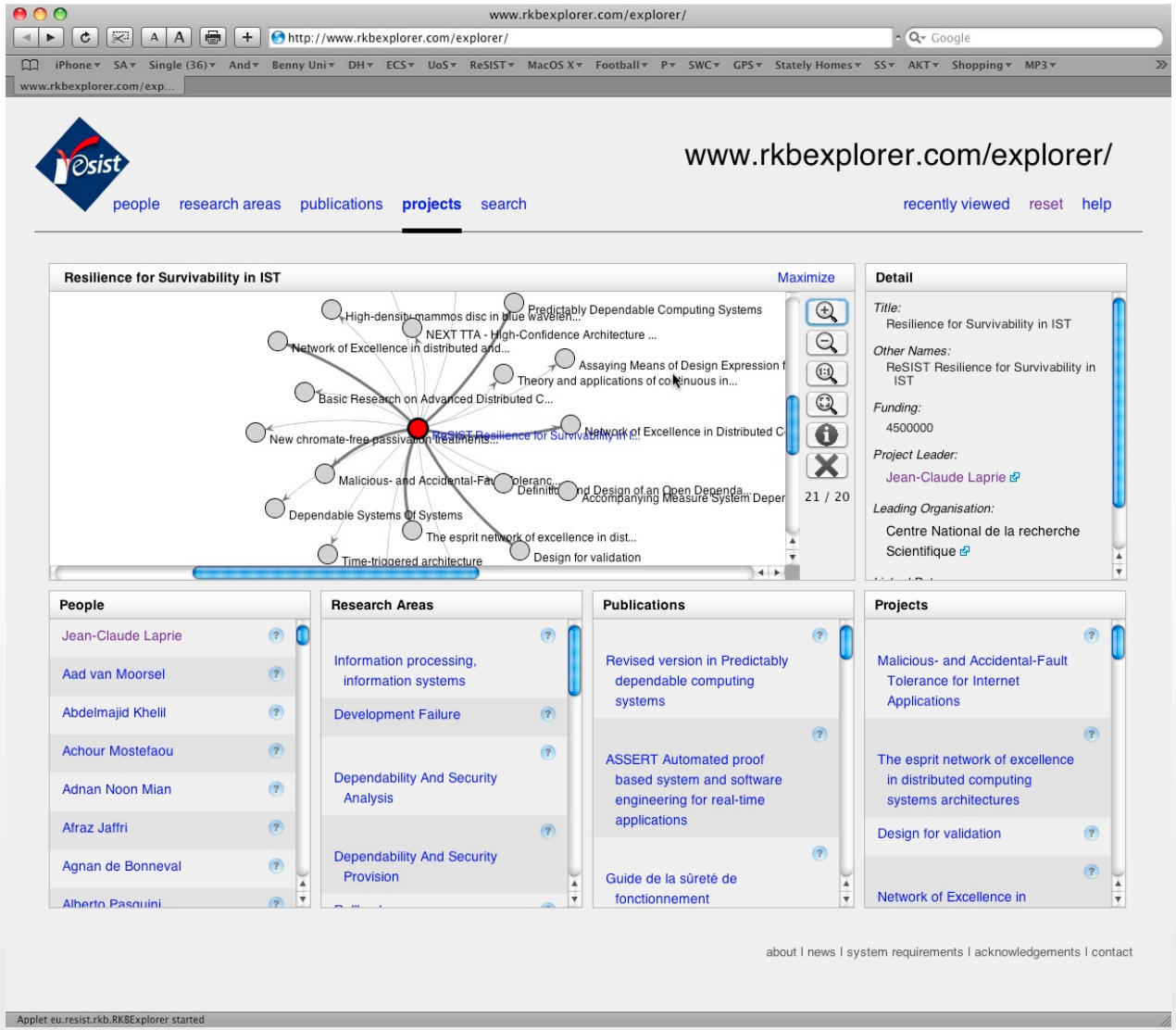
# And so ...


**dotAC: Exploring the UK Research Landscape**

contact@dotac.info

# RKBExplorer.com – Try it!

UNIVERSITY OF Southampton
School of Electronics and Computer Science




http://eprints.ecs.soton.ac.uk/17732/