UNIVERSITY OF Southampton
School of Electronics and Computer Science

Linked Data:
Publishing and Consuming on the Semantic Web

Tsinghua-Southampton Web Science Laboratory at Shenzhen

Hugh Glaser, Deputy Director
Ian Millard

19th August 2009

http://eprints.ecs.soton.ac.uk/17788/

清華大學深圳研究生院
Graduate School at Shenzhen, Tsinghua University

网络科学深圳实验室

University of Southampton

UNIVERSITY OF Southampton
School of Electronics and Computer Science

- Since 1862 (Hartley Institution)
- Multi-discipline University
  - Arts, Science, Law, Medicine, Engineering, …
- 25,000 Students
- 5,000 Staff
- “Russell Group”

UNIVERSITY OF Southampton
School of Electronics and Computer Science

ECS

- 100+ Faculty
- 1000+ Students (300 Graduate)
- Highly-rated
  - Teaching and Research

http://www.ecs.soton.ac.uk/

UNIVERSITY OF Southampton
School of Electronics and Computer Science

Research Groups

- Communication
- Dependable Systems & Software Engineering
- Electrical Power Engineering
- Electronic Systems & Devices
- Information: Signals, Images, Systems
- Intelligence, Agents, Multimedia
- Learning Societies Lab
- Nano
- Optoelectronics Research Centre
- Pervasive Systems Centre
- Science & Engineering of Natural Systems

Semantic Web

UNIVERSITY OF Southampton
School of Electronics and Computer Science

- w3.org FAQ:
  - The **Semantic Web** provides a common framework that allows **data** to be shared and reused across application, enterprise, and community boundaries.
- Exposing data for agents to use
  - Including databases
- Interoperable
  - Naming schemes
  - Vocabularies
  - Protocols

UNIVERSITY OF Southampton
School of Electronics and Computer Science

RDF (Resource Description Framework)

"Hugh Glaser works for the University of Southampton"

```
person-00021
   name
       "Hugh Glaser"

person-00021
   works-for
       Southampton_University

Southampton_University
   label
       "The University of Southampton"
```

RDF (Resource Description Framework)

"Hugh Glaser works for the University of Southampton"

```
<http://southampton.rkbexplorer.com/id/person-00021>
   <http://xmlns.com/foaf/0.1/name>
       "Hugh Glaser" .

<http://southampton.rkbexplorer.com/id/person-00021>
   <http://aktors.org/ontology/portal#works-for>
       <http://dbpedia.org/resource/Southampton_University> .

<http://dbpedia.org/resource/Southampton_University>
   <http://www.w3.org/2000/01/rdf-schema#label>
       "The University of Southampton" .
```

UNIVERSITY OF Southampton
School of Electronics and Computer Science

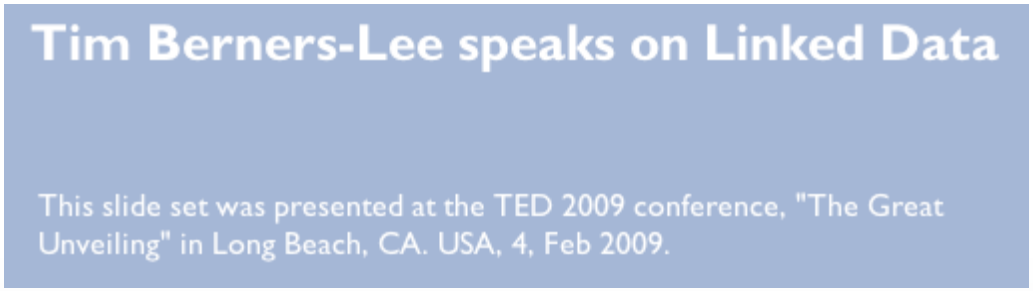
RDF (Resource Description Framework)

- RDF Triples constitute a graph
  - <subject-uri> <predicate-uri> <object-uri>
  - <subject-uri> <predicate-uri> “String”
- *ontologies* define vocabularies, types and relationships
- Agreed URIs facilitate linkage between datasets

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"

UNIVERSITY OF Southampton
School of Electronics and Computer Science

Linked (Open) Data

- The “practical side” of the Semantic Web
- Move from a Web of Text to a Web of Data, where the meaning is specified using metadata (**RDF**)
- Machine to machine communication
  - “Silent revolution” (Tom Ilube @ G20 IdeasLab)
- Facilitates fusion of data sources
  - Maintaining distribution of the data at the providers and others
- Can provide powerful knowledge processing

CALAIS
Powered by Thomson Reuters

The Calais initiative is about enabling semantic applications by providing a metadata generation web service, sample applications using that service to jumpstart development efforts, and support for developers.

UNIVERSITY OF Southampton
School of Electronics and Computer Science

Guardian opens APIs and 'Data Store'

The Guardian launched its Open Platform with APIs for all of its content and a sets of data collected by its editors.

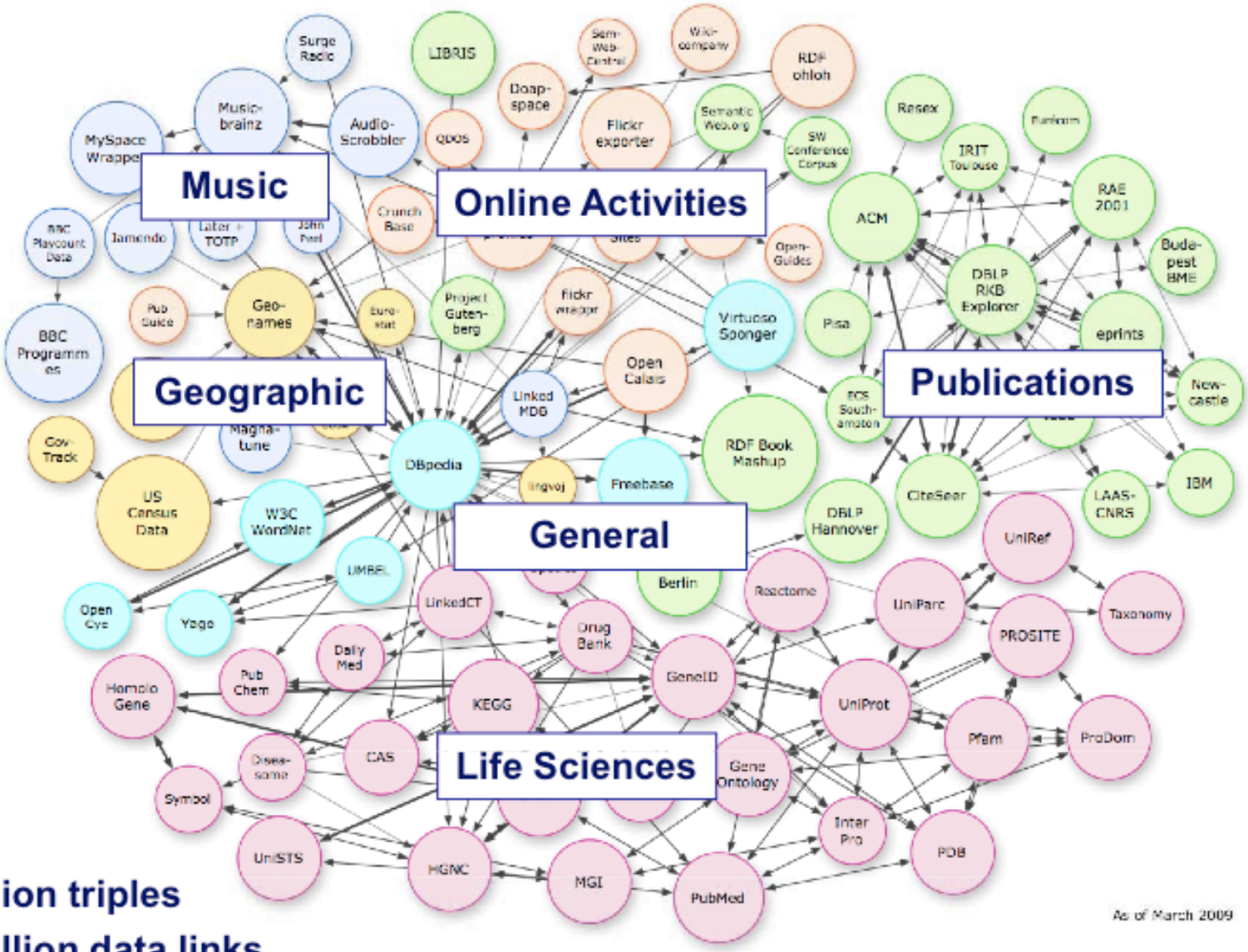
BBC: http://www.bbc.co.uk/music/artists/d4d17620-fd97-4574-92a8-a2cb7e72ce42.rdf

web scale

- REST (ROA, WOA)
  - Web page = Web Service
    - XHTML+RDFa markup for people and machines
  - *resource* as *public record*
    - global visibility and persistence of URI's
- Atom as RESTful API
  - introspection service for feed resource *discovery*
  - HTTP *uniform interface*
    - CRUD analog - PUT/GET/POST/DELETE
  - <entry> as *recordset* </entry> resource state event
- atom+xml and xhtml+xml *representations*
  - Caching, Crawling and Indexing
- Web = DB
  - Linked Open Data SPARQL endpoints, XHTML+RDFa

George Thomas
U.S. government's recovery.gov

LOD Datasets on the Web: March 2009



4.5 billion triples

180 million data links

As of March 2009

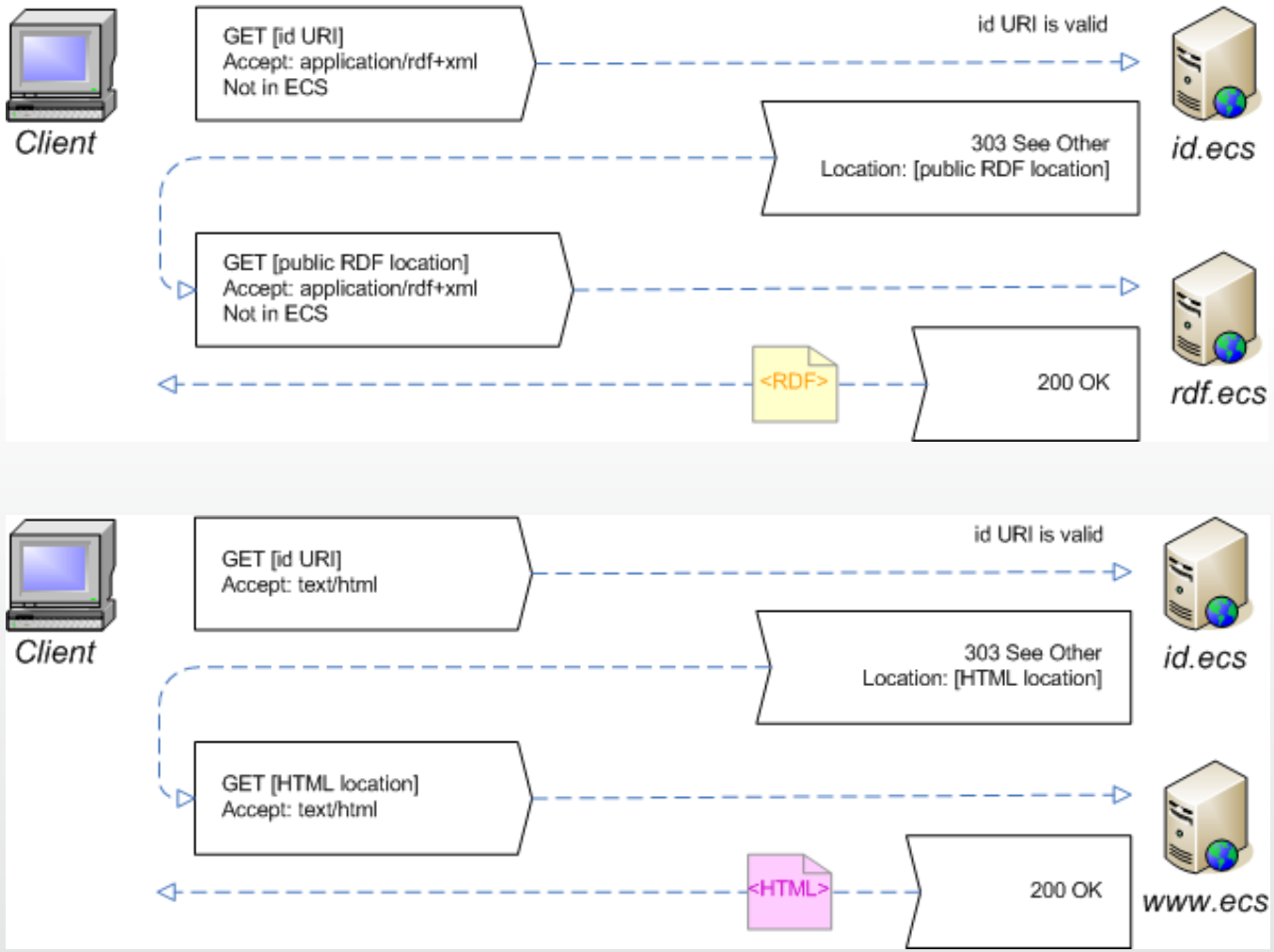
UNIVERSITY OF Southampton
School of Electronics and Computer Science

Linked Data Principles (Berners-Lee)

1. Use URIs as names for things
2. Use HTTP URIs so that people can look up those names
3. When someone looks up a URI, provide useful information, using the standards (RDF, SPARQL)
4. Include links to other URIs, so that they can discover more things

UNIVERSITY OF Southampton
School of Electronics and Computer Science

Linked Data URI Resolution



UNIVERSITY OF Southampton
School of Electronics and Computer Science

ReSIST Project 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

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	semanticweb.org
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”

For example

- Statistics for repository dblp.rkbexplorer.com
  - Last data assertion 2009-06-01 20:05:10
  - Number of triples 27186150
  - Number of symbols 7227252
  - Size of RDF dataset 2.1G

Co-Reference

UNIVERSITY OF Southampton
School of Electronics and Computer Science

- Datasets have people, publications, etc. which also exist in other datasets
- 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 datasets

UNIVERSITY OF Southampton
School of Electronics and Computer Science

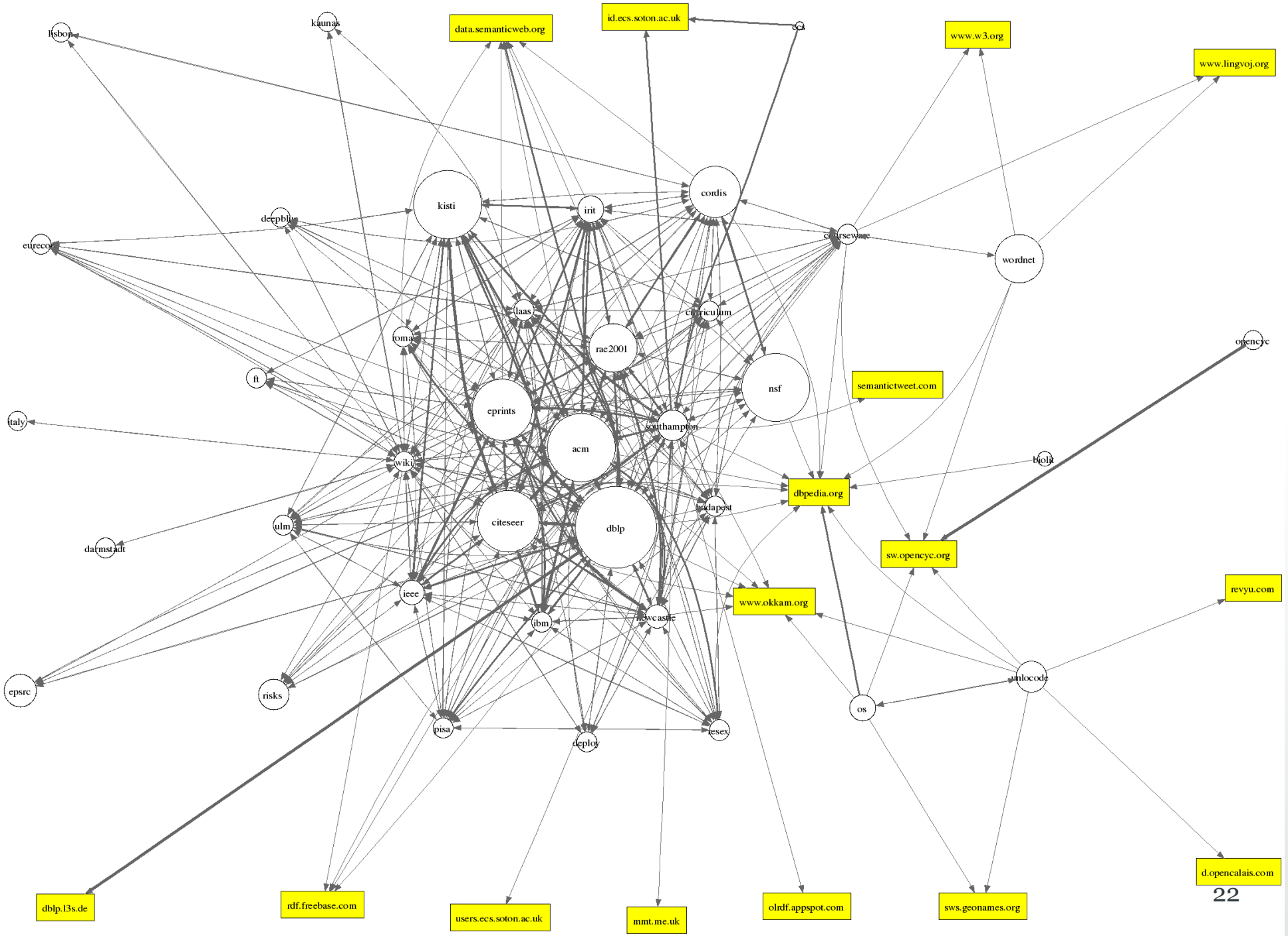
Co-Reference Service (CRS)

- 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

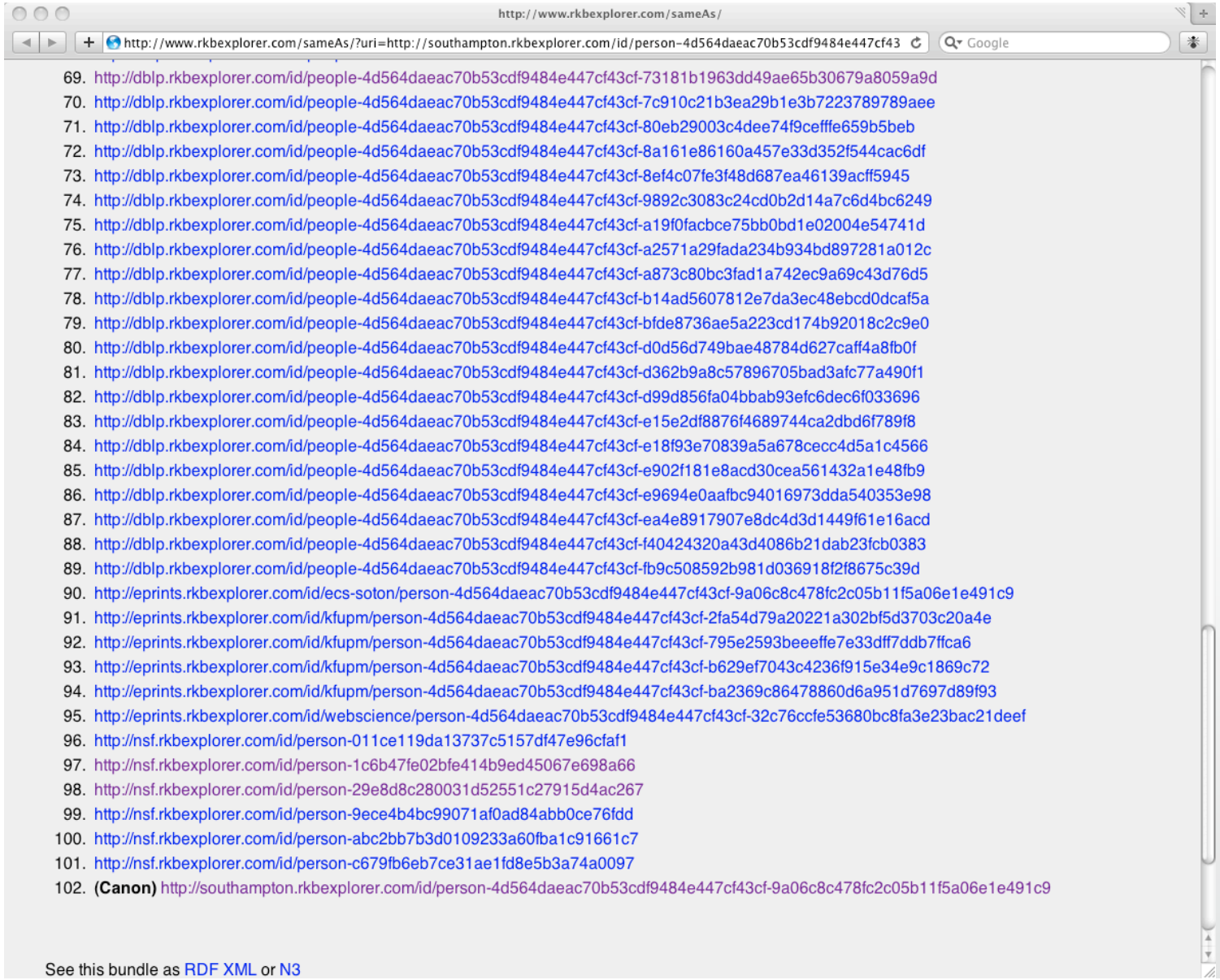
CRS continued

- 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


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

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

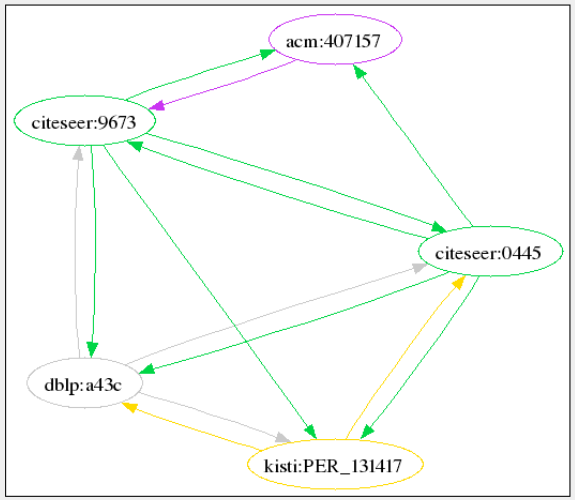
This service computes the equivalence class within the known URIs for a specified URI, by consulting all relevent CRS knowledge bases.

<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 ...

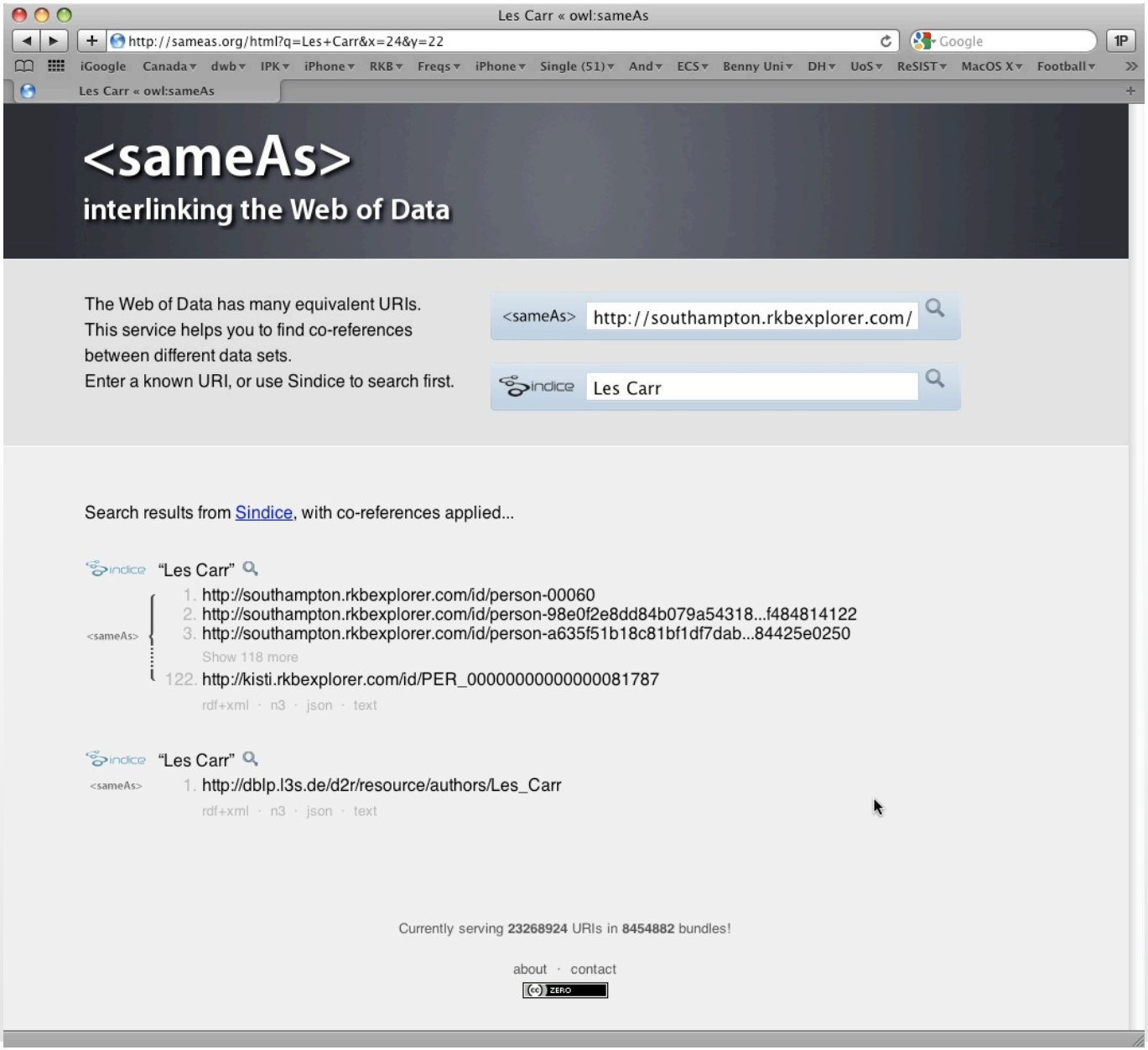
Showing information queried from all repositories ...

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-

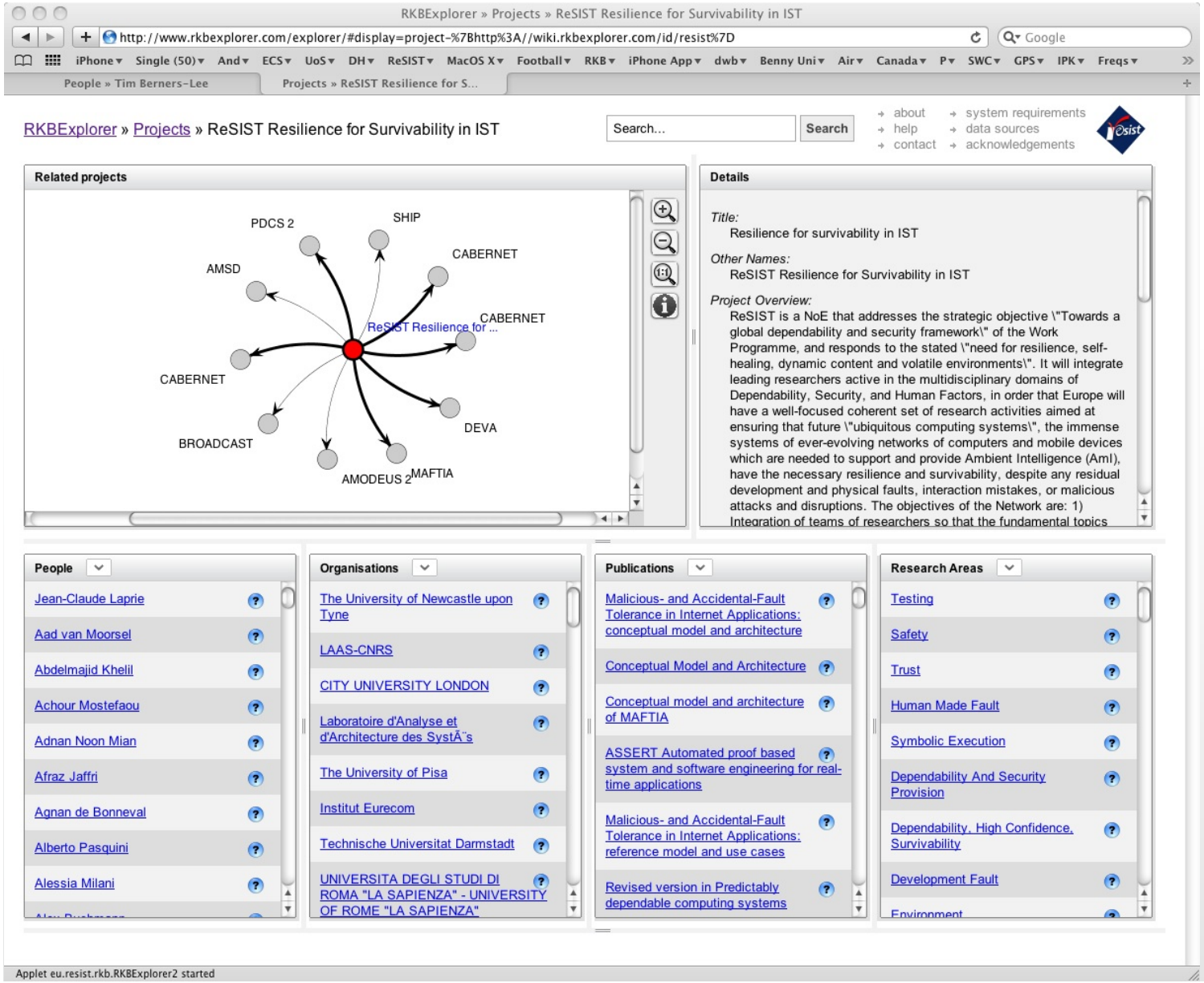
Finding Co-reference

UNIVERSITY OF Southampton
School of Electronics and Computer Science



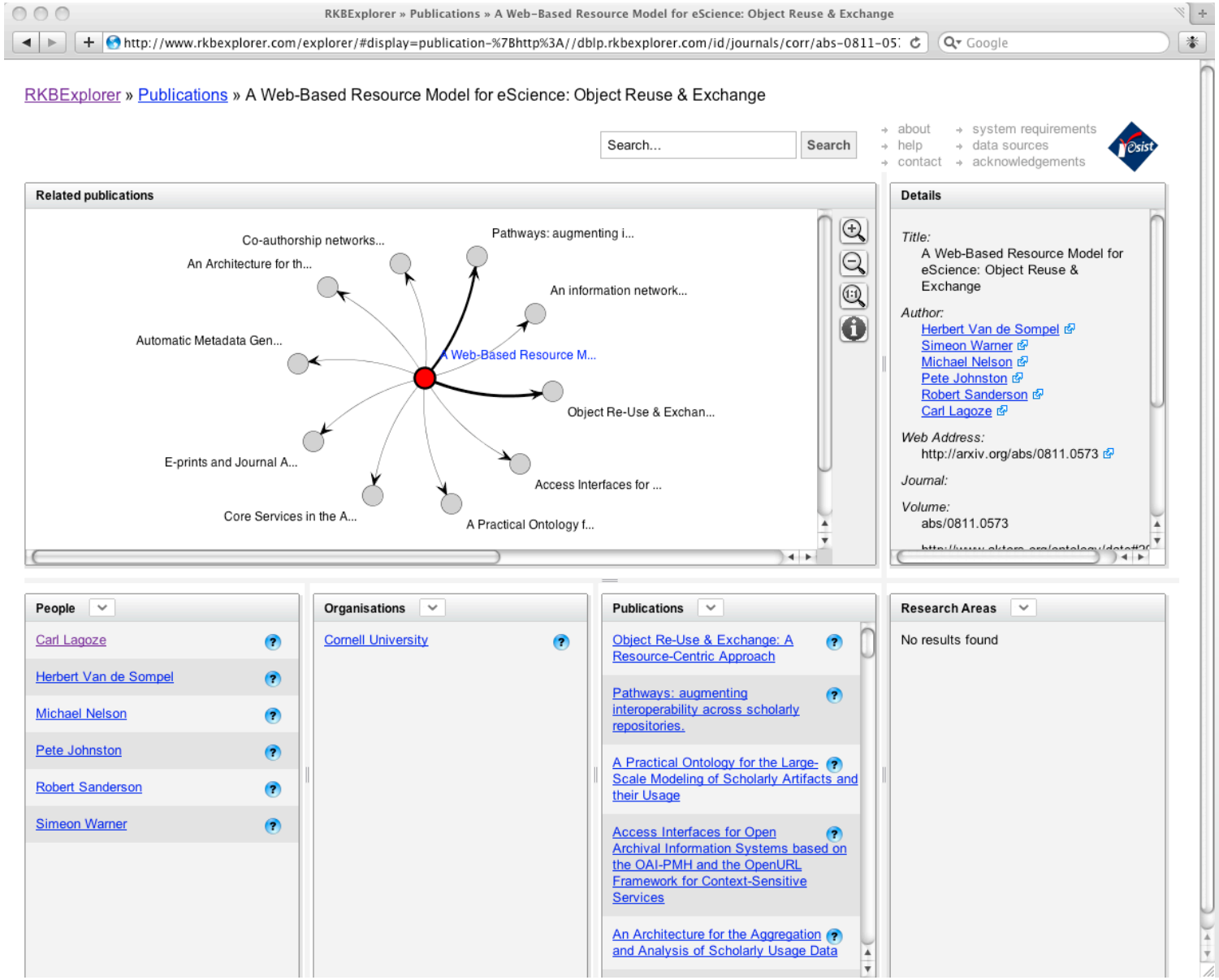
RKBExplorer – the ReSIST Project

UNIVERSITY OF Southampton
School of Electronics and Computer Science



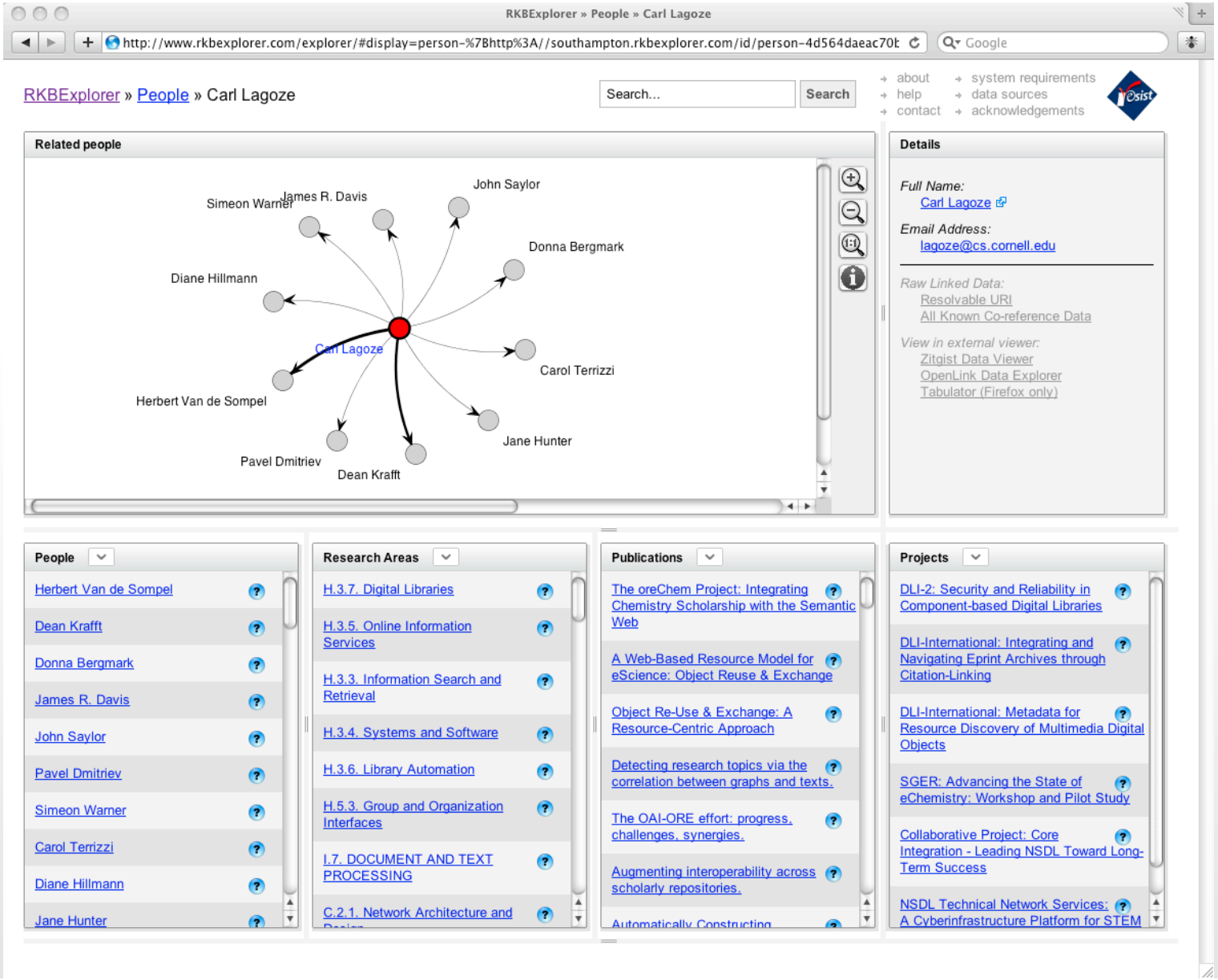
A Paper

UNIVERSITY OF Southampton
School of Electronics and Computer Science



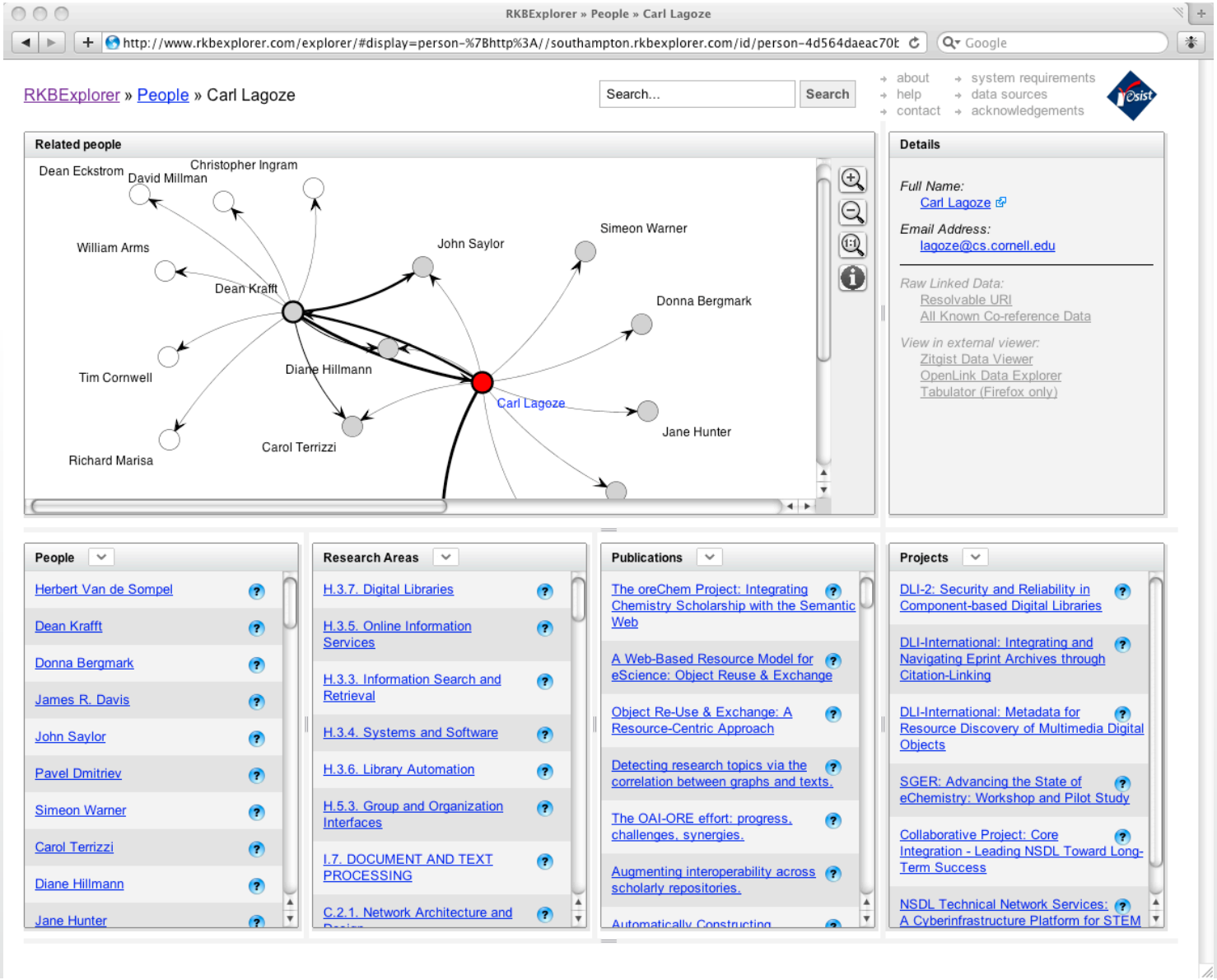
Now Look at an Author

UNIVERSITY OF Southampton
School of Electronics and Computer Science



Or a Couple of People

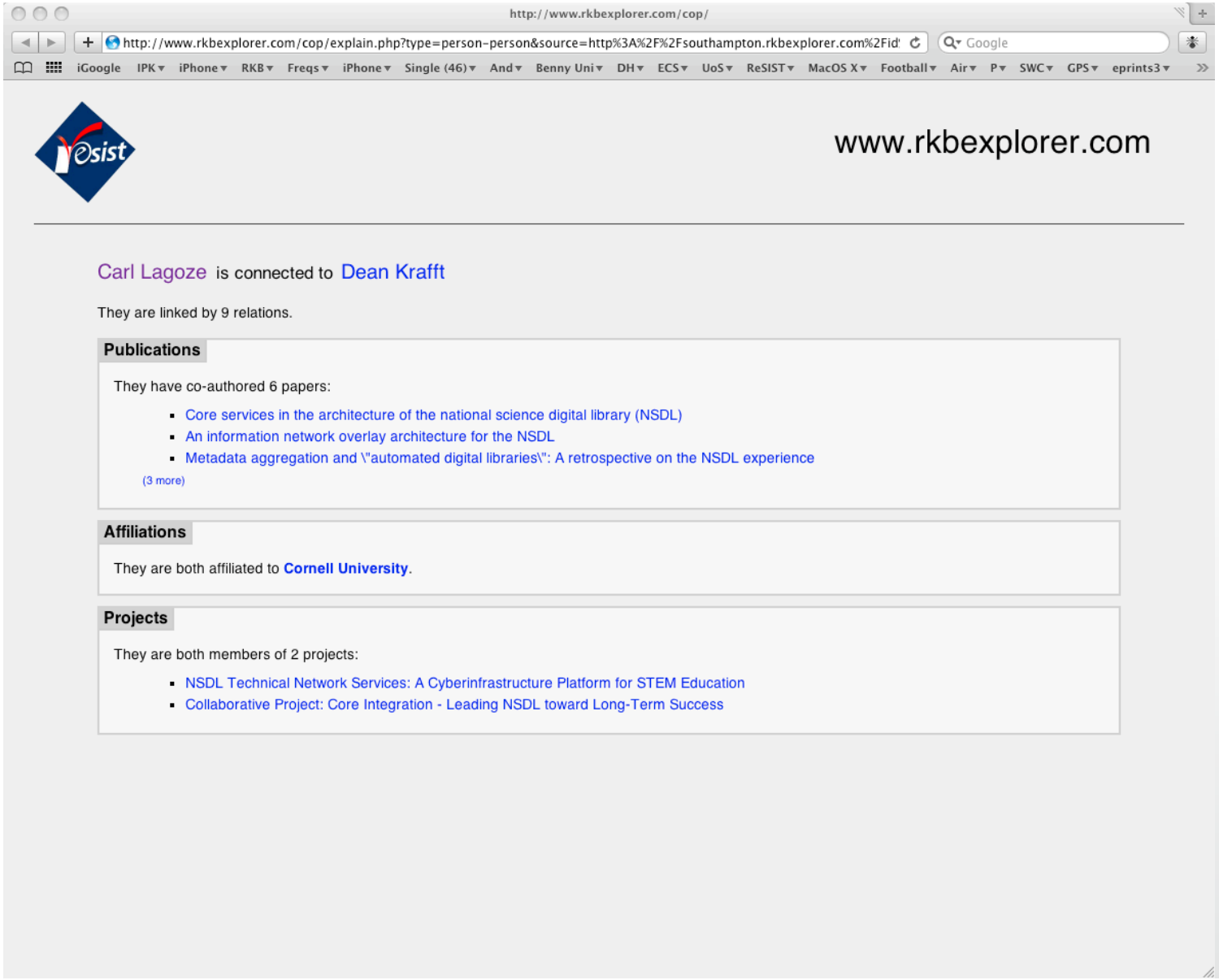
UNIVERSITY OF Southampton
School of Electronics and Computer Science



And how they are linked

And Why they are Linked?

UNIVERSITY OF Southampton
School of Electronics and Computer Science



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

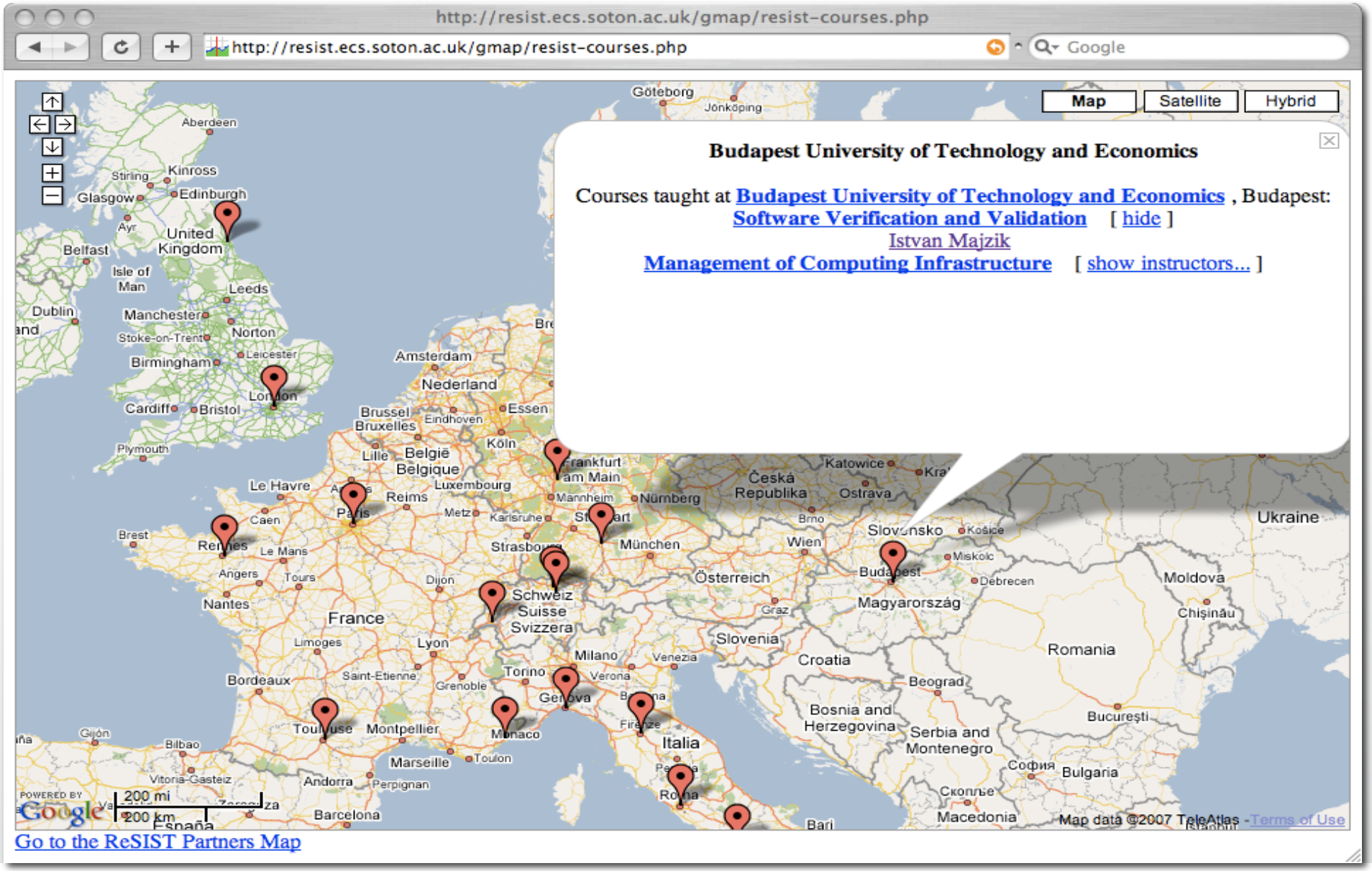
UNIVERSITY OF Southampton
School of Electronics and Computer Science

Cross-Dataset Services

- Search
- Aggregated summary view
  - http://www.rkbexplorer.com/detail/?uri=http://southampton.rkbexplorer.com/id/eprints-12614
- Networks of related URIs
  - http://www.rkbexplorer.com/network/?uri=http://southampton.rkbexplorer.com/id/eprints-12614&type=publication-person
- Co-reference
- …

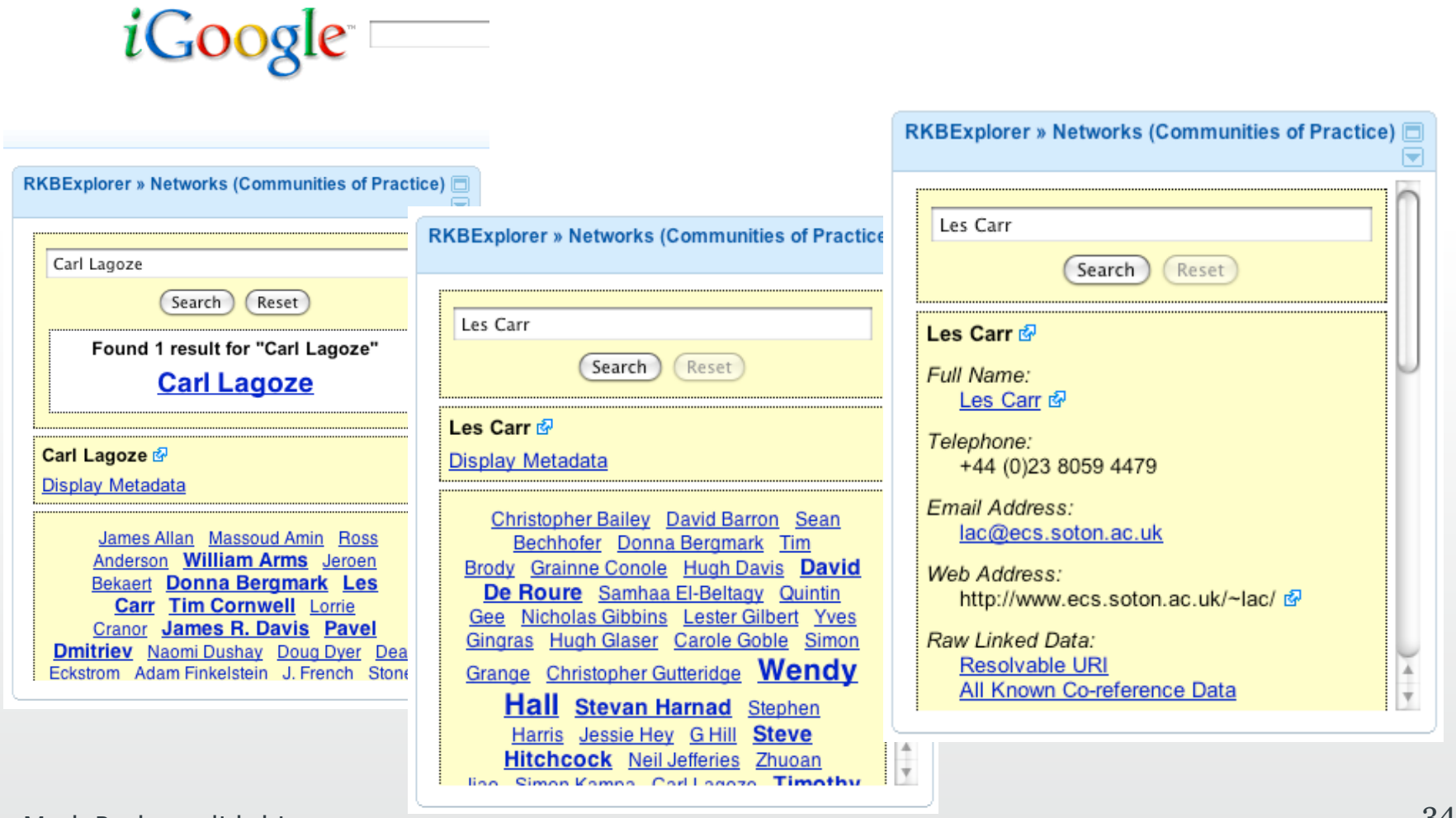
Where are Courses Taught?

UNIVERSITY OF Southampton
School of Electronics and Computer Science

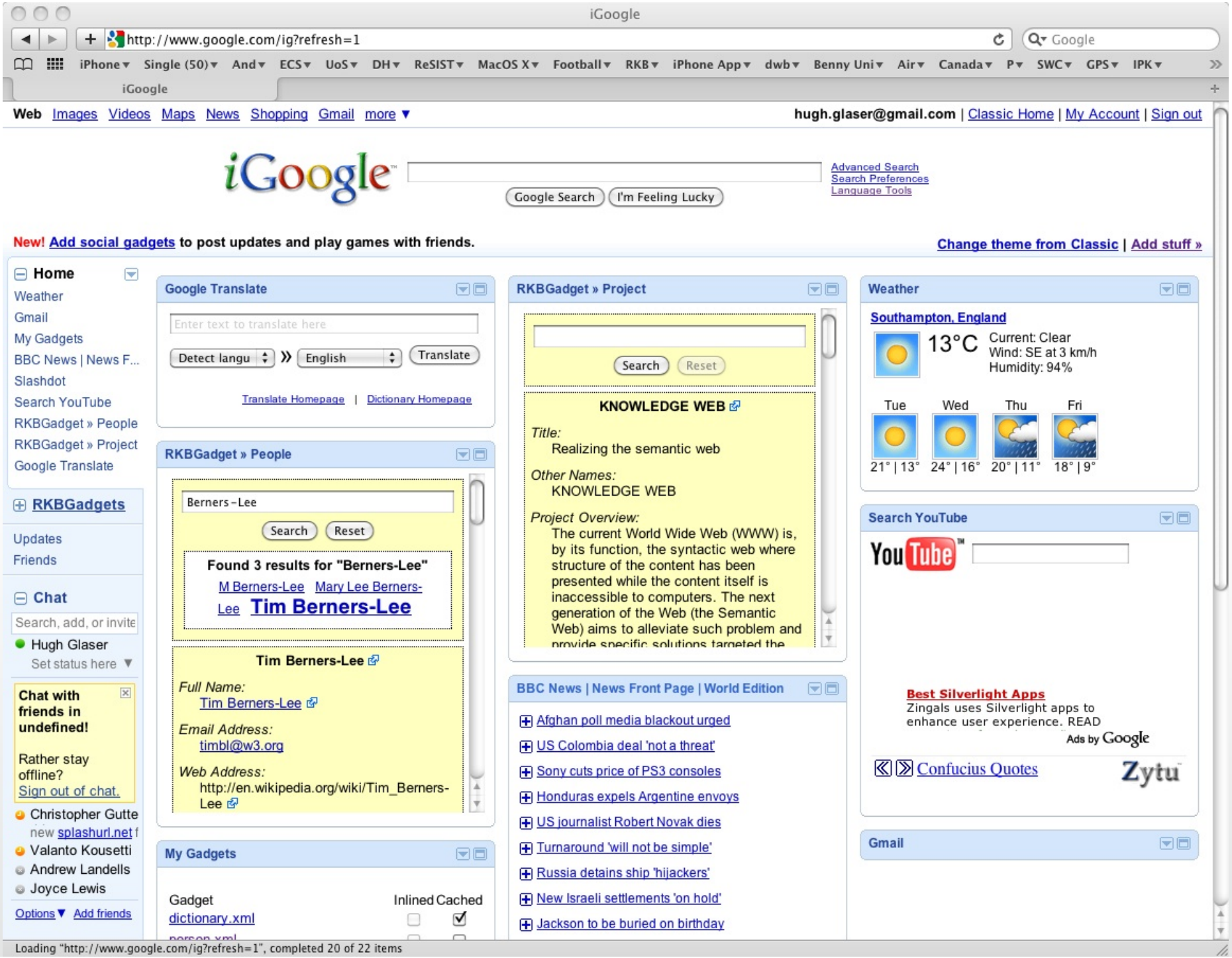


UNIVERSITY OF Southampton
School of Electronics and Computer Science

Gadget – find out about people

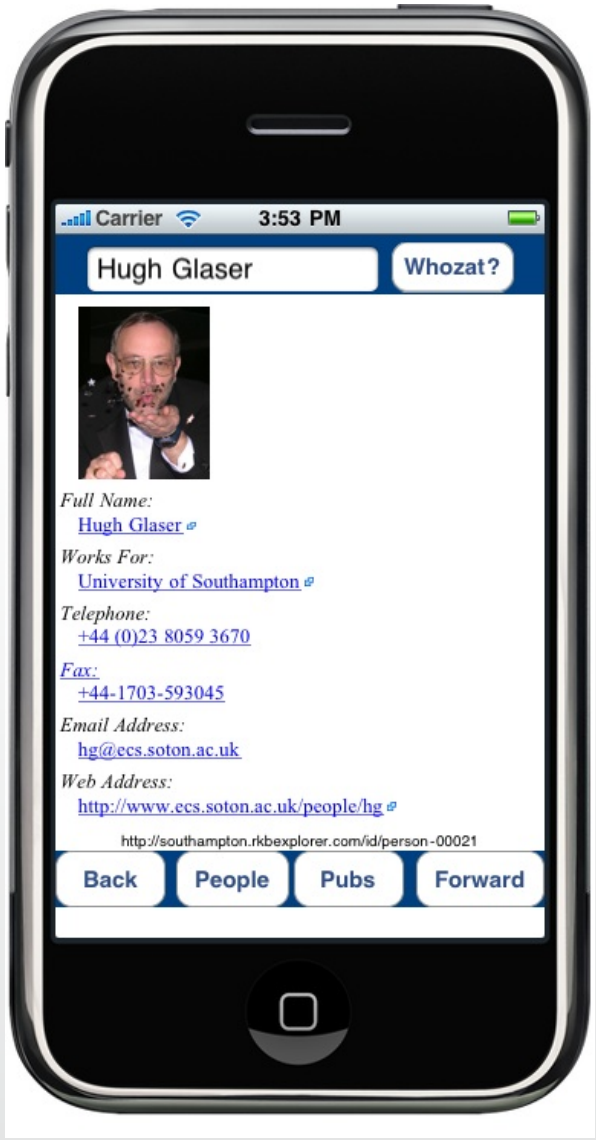


Mark Borkum did this


iGoogle
http://www.google.com/ig?refresh=1
Google
iPhone Single (50) And ECS UoS DH ReSIST MacOS X Football RKB iPhone App dwb Benny Uni Air Canada P SWC GPS IPK
iGoogle
Web Images Videos Maps News Shopping Gmail more
hugh.glaser@gmail.com | Classic Home | My Account | Sign out
iGoogle
Google Search
I'm Feeling Lucky
Advanced Search
Search Preferences
Language Tools
New! Add social gadgets to post updates and play games with friends.
Change theme from Classic | Add stuff »
Home
Weather
Gmail
My Gadgets
BBC News | News F...
Slashdot
Search YouTube
RKBGadget » People
RKBGadget » Project
Google Translate
RKBGadgets
Updates
Friends
Chat
Search, add, or invite
Hugh Glaser
Set status here
Chat with friends in undefined!
Rather stay offline?
Sign out of chat.
Christopher Gutte
new splashurl.net f
Valanto Kousetti
Andrew Landells
Joyce Lewis
Options Add friends
Google Translate
Enter text to translate here
Detect langu
English
Translate
Translate Homepage | Dictionary Homepage
RKBGadget » People
Berners-Lee
Search
Reset
Found 3 results for "Berners-Lee"
M Berners-Lee Mary Lee Berners-Lee Tim Berners-Lee
Tim Berners-Lee
Full Name:
Tim Berners-Lee
Email Address:
timbl@w3.org
Web Address:
http://en.wikipedia.org/wiki/Tim_Berners-Lee
My Gadgets
Gadget
Inlined
Cached
dictionary.xml
RKBGadget » Project
Search
Reset
KNOWLEDGE WEB
Title:
Realizing the semantic web
Other Names:
KNOWLEDGE WEB
Project Overview:
The current World Wide Web (WWW) is, by its function, the syntactic web where structure of the content has been presented while the content itself is inaccessible to computers. The next generation of the Web (the Semantic Web) aims to alleviate such problem and
BBC News | News Front Page | World Edition
Afghan poll media blackout urged
US Colombia deal 'not a threat'
Sony cuts price of PS3 consoles
Honduras expels Argentine envoys
US journalist Robert Novak dies
Turnaround 'will not be simple'
Russia detains ship 'hijackers'
New Israeli settlements 'on hold'
Jackson to be buried on birthday
Weather
Southampton, England
13°C
Current: Clear
Wind: SE at 3 km/h
Humidity: 94%
Tue Wed Thu Fri
21° | 13° 24° | 16° 20° | 11° 18° | 9°
Search YouTube
YouTube
Best Silverlight Apps
Zingals uses Silverlight apps to enhance user experience. READ
Ads by Google
Confucius Quotes
Zytu
Gmail
Loading "http://www.google.com/ig?refresh=1", completed 20 of 22 items

iPhone App

UNIVERSITY OF Southampton
School of Electronics and Computer Science



UNIVERSITY OF Southampton
School of Electronics and Computer Science

IPK

- New collaboration
  - Tsinghua/Southampton
- IP Knowledge Base
  - Facilitate IP take-up by SMEs
- Augment planned portal with Linked Data capabilities

UNIVERSITY OF Southampton
School of Electronics and Computer Science

"The" IPK

- In fact many sub-IPKs
  - Every source published separately
    - IP owners and SMEs (cf the LOD diagram)
  - Supported by the Project
    - As a service
    - Can be pushed out to stakeholders
- Central added value of knowledge processing for recommendation
- "Push", not just search
  - Represent knowledge about customers

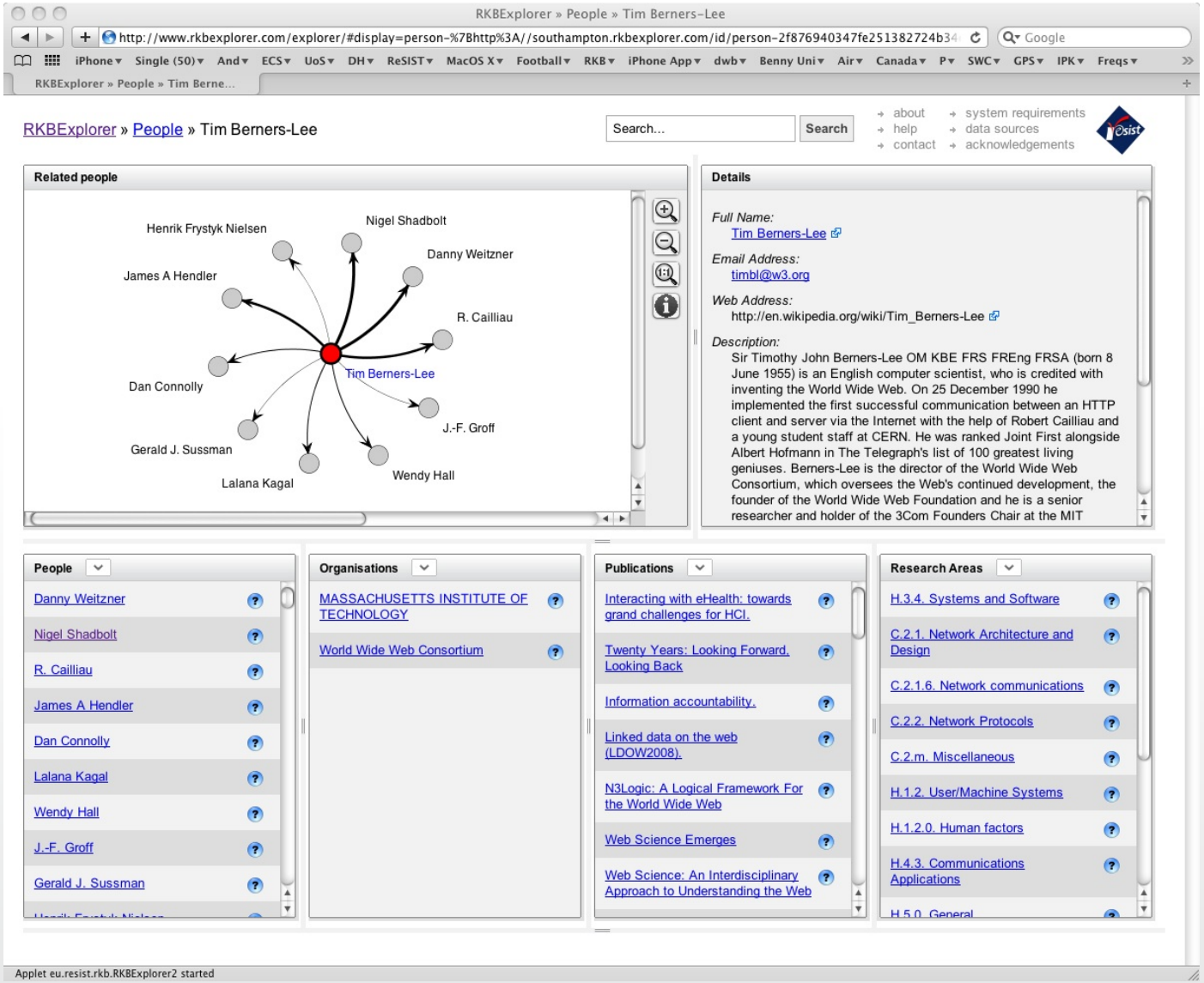
UNIVERSITY OF Southampton
School of Electronics and Computer Science

Concluding Remarks

- Linked Data is not hard
  - Producing - easy
  - Publishing – not quite so easy
  - Using - easy
- Powerful data integration technology
- Major adoption now happening
  - Eg US & UK governments

RKBExplorer.com – Try it!

UNIVERSITY OF Southampton
School of Electronics and Computer Science

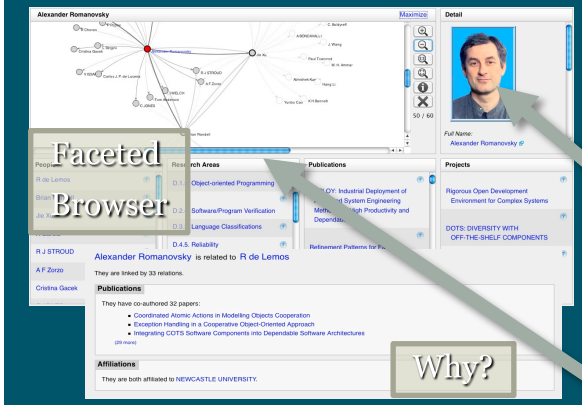


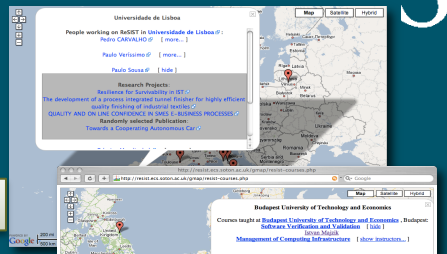
http://eprints.ecs.soton.ac.uk/17788/

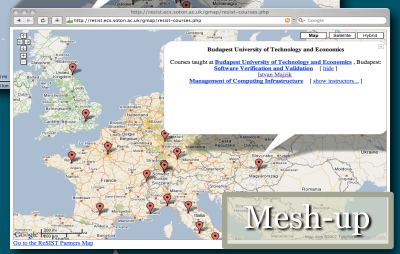
Knowledge-Enabled Research Support: RKBExplorer.com

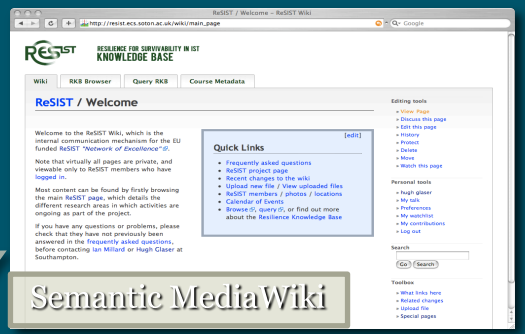
Ian Millard & Hugh Glaser, {icm,hg}@ecs.soton.ac.uk

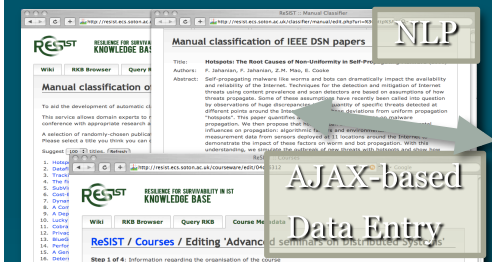
UNIVERSITY OF Southampton
School of Electronics and Computer Science

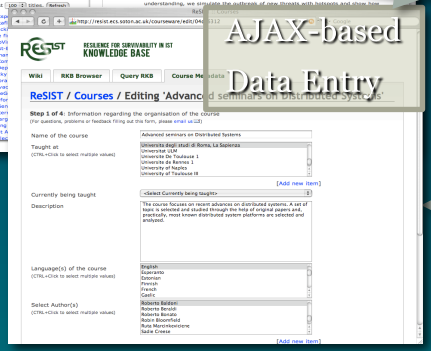










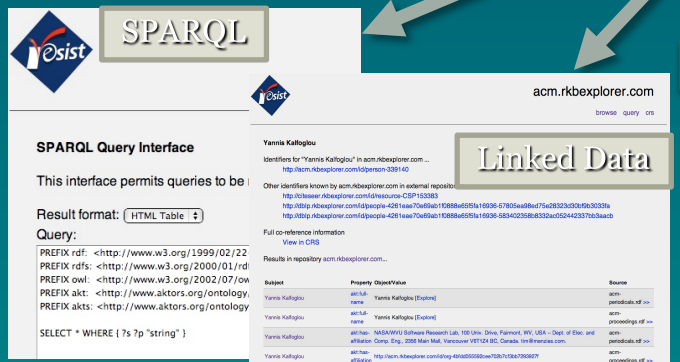


3store

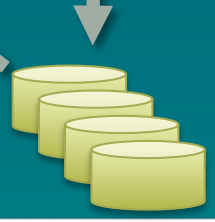
CRS

Coreference Analysis

Offline Conversion & Versioning



Ontology Mapping Service



Ontologies

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.