

# RKBExplorer.com:

## Anatomy of a Semantic Web Application

Hugh Glaser & Ian Millard  
4th December 2008, Seoul



# Context

- CSAKTiveSpace
  - AKT Project
  - 2003 Semantic Web Challenge winner
- ReSIST - EU Network of Excellence in Resilient Systems
  - Knowledge-enabled infrastructure
  - Jan 2006 – Dec 2008
- KISTI work on Co-reference
  - 2008



# User Interaction

- Semantic MediaWiki
- Custom form interfaces
- Google Maps
- Raw Knowledge Browser
- **RKBExplorer**
- Why do you think that? information



### Resilience for Survivability in IST

Maximize

21 / 20

#### 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 Moorssel	?
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	?

# Focusing on a Person

**Alexander Romanovsky**
Maximize

50 / 60

**People**

- R de Lemos
- Brian Randell
- Jie Xu
- A Garcia
- R J STROUD
- A F Zorzo
- Cristina Gacek

**Research Areas**

- D.1.5. Object-oriented Programming
- D.2.4. Software/Program Verification
- D.3.2. Language Classifications
- D.4.5. Reliability
- D.2.5. Testing and Debugging
- D.2.11. Software Architectures

**Publications**

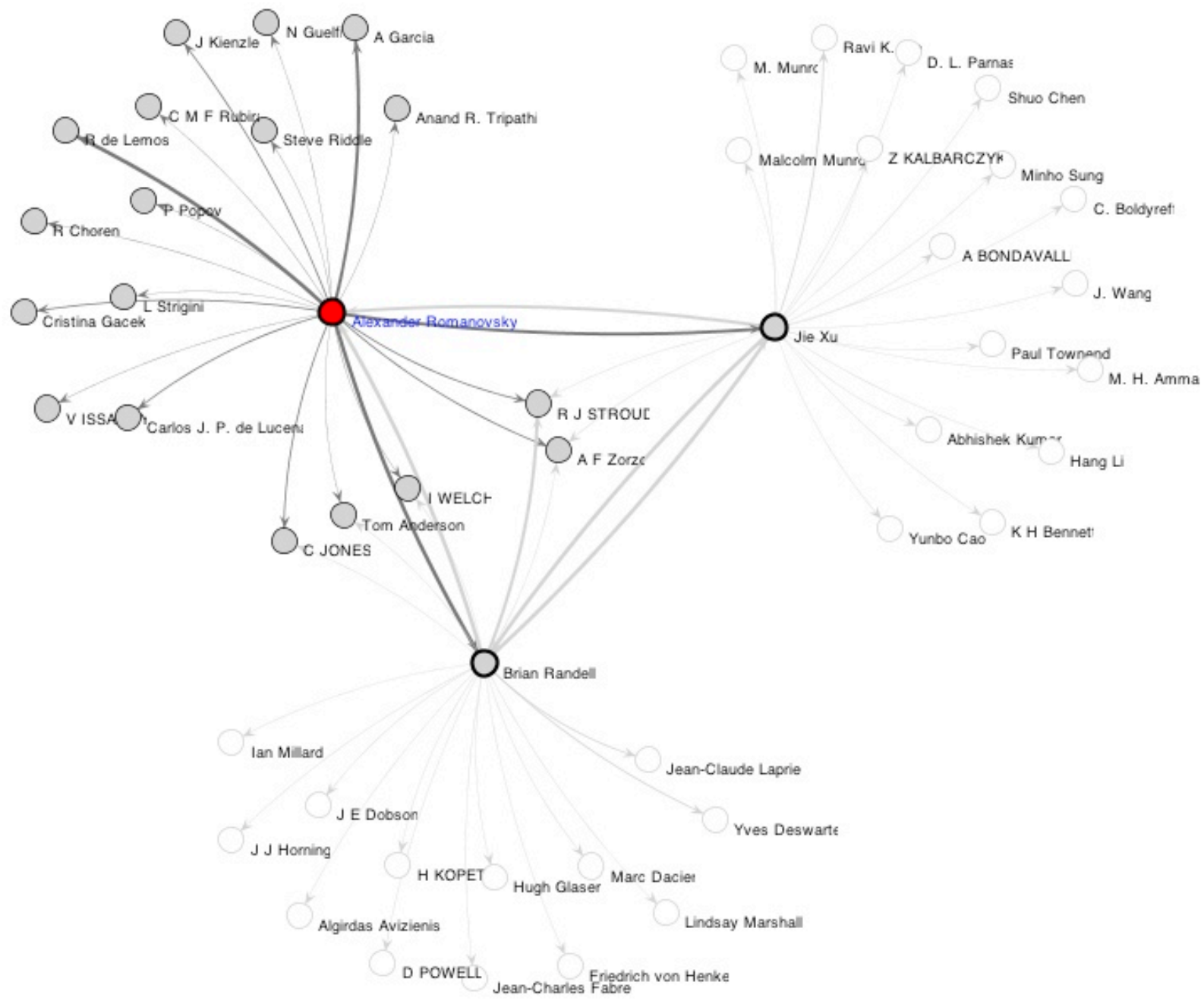
- DEPLOY: Industrial Deployment of Advanced System Engineering Methods for High Productivity and Dependability
- Refinement Patterns for Fault Tolerant Systems
- Special track on Dependable and Adaptive Distributed Systems: editorial message

**Projects**

- Rigorous Open Development Environment for Complex Systems
- DOTS: DIVERSITY WITH OFF-THE-SHELF COMPONENTS
- ReSIST Resilience for Survivability in IST

**Detail**

Full Name:  
[Alexander Romanovsky](#)



# Knowledge Sources

- Partners
- Publications
- Funding Agencies
- Project Wiki
- Courseware
- Resilient-Explicit Computing
- Wide range, don't just look where you expect to find

# Some Underlying Sources

**acm.rkbexplorer.com**

budapest.rkbexplorer.com

citeseer.rkbexplorer.com

**cordis.rkbexplorer.com**

**courseware.rkbexplorer.com**

darmstadt.rkbexplorer.com

**dblp.rkbexplorer.com**

deepblue.rkbexplorer.com

deploy.rkbexplorer.com

epsrc.rkbexplorer.com

eurecom.rkbexplorer.com

ft.rkbexplorer.com

ibm.rkbexplorer.com

**ieee.rkbexplorer.com**

irit.rkbexplorer.com

italy.rkbexplorer.com

kaunas.rkbexplorer.com

**kisti.rkbexplorer.com**

laas.rkbexplorer.com

lisbon.rkbexplorer.com

newcastle.rkbexplorer.com

**nsf.rkbexplorer.com**

pisa.rkbexplorer.com

rae2001.rkbexplorer.com

**resex.rkbexplorer.com**

roma.rkbexplorer.com

southampton.rkbexplorer.com

ulm.rkbexplorer.com

unlocode.rkbexplorer.com

wiki.rkbexplorer.com

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



## For example

- Statistics for repository [kisti.rkbexplorer.com](http://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

# Semantic Web/Linked Data issues

- The system supports state of the art facilities:
  - Browsing
  - Resolvable URIs
  - SPARQL endpoint
  - CRS (Co-reference knowledge)
  - RDF publishing
  - Semantic Sitemap
  - On a separate domain

# UNIVERSITY OF Southampton

## School of Electronics and Computer Science

**Alexander Romanovsky**

People working on ReSIST in Universidade de Lisboa @ Paulo Verissimo @ [more...] Paulo Sousa @ [hide]

**Research Areas**

- D.1.5. Object-Oriented Programming
- D.2.4. Software/Program Verification

**Publications**

They have co-authored 32 papers:

- Coordinated Atomic Actions in Modelling Objects Cooperation
- Exception Handling in a Collaborative Object-Oriented Approach
- Integrating COTS Software Components into Dependable Software Architectures

**Universidade de Lisboa**

People working on ReSIST in Universidade de Lisboa @ Paulo Verissimo @ [more...] Paulo Sousa @ [hide]

**Research Projects**

Research for Sustainability in IT - The development of a process integrated control system for highly efficient, secure, flexible and reconfigurable factory. QUALITY AND IN LINE COMPLIANCE IN FINE E-BUSINESS PROCESSES - Reconfiguring industrial Production - Towards a Collaborative Autonomous Car.

**Bulgarian University of Technology and Economics**

Current topics in Software Engineering, IT Technological and Environmental Software Software Verification and Validation (S3V) ITSM Management of Computing Infrastructures | more... | more... | more...

**ReSIST / Welcome**

Welcome to the ReSIST Wiki, which is the internal communication mechanism for the EU funded ReSIST - 'Network of Excellence'.

Note that virtually all pages are private, and viewable only to ReSIST members who have logged in.

Most content can be found by freely browsing the main ReSIST pages, which details the different research areas in which activities are ongoing as part of the project.

If you have any questions or problems, please check that they have not previously been answered in the frequently asked questions, before contacting Ian Millard or Hugh Gasser at Southampton.

**Quick Links**

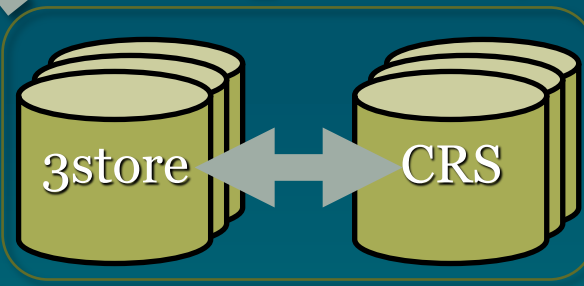
- Frequently asked questions
- ReSIST project page
- Recent changes to the wiki
- Upload new files / View uploaded files
- ReSIST members / photos / locations
- Calendar of events
- Browse by quality ID or find out more about the Resilience Knowledge Base

**Manual classification of IEEE DSN papers**

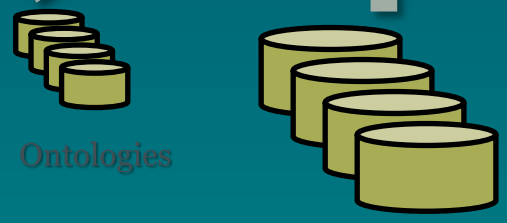
This page highlights the Root Causes of Non-Uniformity in Self-Propagating Hardware (2000)

Authors: A. Jamborek, A. Jamborek, J. P. Kim, P. Jamborek

Self-propagating hardware (the worms and bots) can dramatically impact the availability and reliability of the Internet. Techniques for the detection and mitigation of Internet worms using control plane and data plane based methods are presented. Some of these assumptions have recently been called into question by observations of large outbreaks in the form of specific worms detected at different points around the Internet. We call these deviations from uniform propagation 'anomalies'. This paper identifies and explains these anomalies in detail. The impact of anomalies on propagation is examined by simulation. The impact of anomalies on the performance of self-propagating hardware is also examined. This paper identifies and explains these anomalies in detail. The impact of anomalies on the performance of self-propagating hardware is also examined. This paper identifies and explains these anomalies in detail.



**Offline Conversion & Versioning**



**acm.rkbexplorer.com**

**SPARQL Query Interface**

This interface permits queries to be

Result format: **HTML Table**

Query:

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX akt: <http://www.aktors.org/ontology#>
PREFIX akts: <http://www.aktors.org/ontology#>
```

**Results in repository acm.rkbexplorer.com...**

Subject	Property	Object/Value	Source
Yannis Kalligou	akt:foaf:name	Yannis Kalligou (Enrich)	acm:periodicals.rdf
Yannis Kalligou	akt:foaf:mail	Yannis.Kalligou@act.com	acm:periodicals.rdf
Yannis Kalligou	akt:foaf:workplace	NASA/WFO Software Research Lab, 100 Lincoln Drive, Fairfax, VA, USA - Dept. of Elec. and Electron. Comp. Eng. 2001-04-01	acm:periodicals.rdf
Yannis Kalligou	akt:foaf:workplace	University of Thessaly, 22500 Lamia, Greece	acm:periodicals.rdf
Yannis Kalligou	akt:foaf:workplace	Sask. Innov. Tech. Cent. for Health, Info. and Appl. Soft. of Informatics, The Univ. of Saskatchewan, Sask., S4N 0A8, Canada, Sask.	acm:periodicals.rdf
Yannis Kalligou	akt:foaf:workplace	University of Thessaly, 22500 Lamia, Greece	acm:periodicals.rdf

**Resolving URIs**  
KISTI, dbpedia, etc.

CiteSeer, CORDIS, DBLP, Partners, UN LoCode, ...

# Communication

- Ontologies
  - General Scientific Endeavour
  - Domain-specific
  - Support (geospatial, etc)
- Open Local Knowledge – HTTP
  - Resolvable URIs
  - SPARQL
- Uses Remote Knowledge
  - Resolves URIs with caching

# Components 1

- Semantic Web infrastructure throughout
- Triplestore for each source
  - Putting the Web in Semantic Web
  - Stores RDF – (Subject, Predicate, Object)
  - We use 3store
- Linked Data
  - 303 and content negotiation architecture with caching

## Components 2

- Co-Reference Subsystem
  - CRS – more later
- Community of Practice Analysis
  - Why do you think that?
- Ontology Mapping
  - Dealing with other Ontologies
- NLP for text classification
- Caching everywhere

## Components 3

- Application Middleware
  - URI Equivalence Closure
  - RDF Graph Closure
- Semantic Sitemap
  - Facilitate Search Engines

# Co-Reference

- Co-Reference is a Big Problem
  - Identifying multiple URIs for one resource
  - Rejecting incorrectly conflated resources
  - Publishing
  - Using
- Coldstart
  - A serious problem
  - Nothing is linked to anything



# Co-Reference Closure

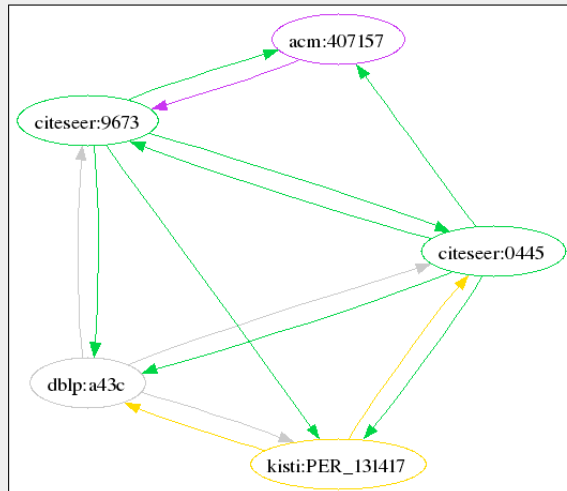
## Complete Co-Reference Information

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

### 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\\_000000000000131417](http://kisti.rkbexplorer.com/id/PER_000000000000131417)

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 Corous-Based Learning Method of Compound Noun			dblp-publications-

# 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
  - $URI_i \rightarrow \{ URI_1, \dots, URI_i, \dots, URI_n \}$
- Recommend a Canon

## 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

# Dealing With Non-SPARQL KBs

- The RKBExplorer application uses SPARQL to query the KBs
  - But needs to access data from KBs that only offer resolvable URIs
- So resolve such a URI
- Cache the RDF with associated resolved RDF locally
- Query the local cache

# Dealing With Different Ontologies

- The RKBExplorer application uses a particular ontology
  - Some KBs will use different ontologies
  - Eg [kisti.rkbexplorer.com](http://kisti.rkbexplorer.com)
- One solution
  - Represent the ontology relationship in RDF (as far as possible)
  - Resolve the URI through the mapping service to get RDF in the required ontology

# Concluding Remarks

- Major Data Fusion using Semantic Web Technologies
- Many things can be cast in a Semantic Web framework
- Linked Data works pretty well
- RDF works pretty well
- A little Ontology goes a long way
- Co-Reference is the biggest problem
  - But is tractable

# RKBExplorer.com/explorer/ – Try it!

The screenshot shows a web browser window displaying the RKBExplorer website. The browser's address bar shows the URL <http://www.rkbexplorer.com/explorer/>. The website header includes the 'osist' logo and navigation links for 'people', 'research areas', 'publications', 'projects', and 'search'. A 'recently viewed' section contains links for 'reset' and 'help'.

The main content area features a network diagram titled 'Resilience for Survivability in IST'. The diagram is a hub-and-spoke model with a central red node labeled 'ReSIST Resilience for Survivability in IST'. Numerous peripheral nodes are connected to this central node, including 'High-density mammos disc in blue wavele...', 'Predictably Dependable Computing Systems', 'NEXT TTA - High-Confidence Architecture ...', 'Network of Excellence in distributed and...', 'Theory and applications of continuous in...', 'Assaying Means of Design Expression f...', 'Basic Research on Advanced Distributed C...', 'New chromate-free passivation treatments', 'Network of Excellence in Distributed C...', 'Malicious- and Accidental-Fault Tolerance and Design of an Open Dependability Accompanying Measure System Depen...', 'Dependable Systems of Systems', 'The esprit network of excellence in dist...', 'Time-triggered architecture', and 'Design for validation'. A 'Maximize' button is visible in the top right corner of the diagram area.

To the right of the diagram is a 'Detail' panel for the selected project. It contains the following information:

- 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

Below the diagram and detail panel are four vertical panels, each with a list of items and a search icon:

- People:** Jean-Claude Laprie, Aad van Moorseel, 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.

At the bottom of the page, there is a footer with links: 'about | news | system requirements | acknowledgements | contact'. A small status bar at the very bottom reads 'Applet eu.resist.rkb.RKBExplorer started'.

<http://eprints.ecs.soton.ac.uk/16946>