ReSIST: Resilience for Survivability in IST

A European Network of Excellence

Contract Number: 026764

Deliverable D23: Knowledge base version 2

Report Preparation Date: December 2007
Classification: Public

Contract Start Date: 1st January 2006
Contract Duration: 36 months
Project Co-ordinator: LAAS-CNRS
Partners: Budapest University of Technology and Economics
City University, London
Technische Universität Darmstadt
Deep Blue Srl
Institut Eurécom
France Telecom Recherche et Développement
IBM Research GmbH
Université de Rennes 1 – IRISA
Université de Toulouse III – IRIT
Vytautas Magnus University, Kaunas
Fundação da Faculdade de Ciencas da Universidade de Lisboa
University of Newcastle upon Tyne
Università di Pisa
QinetiQ Limited
Università degli studi di Roma "La Sapienza"
Universität Ulm
University of Southampton

Contents

1 - Introduction

"Deliverable D8: Draft planning for the next 18 months" states:

> **Deliverable D23: Knowledge base version 2**
>
> RKB version 2. A mature development of the RKB, reflecting the consolidation and maintenance performed by IT-T1, and the extensions to resilience-explicit computing support incorporated in the RKB and ontology, and from the development of challenge problems. Prepared by task IT-T1, with support from IT-T2 and IT-T3.

It further states:

> **Task IT-T1: Resilience Knowledge Base (RKB)**
>
> The plan for the Resilience Knowledge Base was to provide an integrative infrastructure to the ReSIST Project and the wider community. The aim was to deliver convenient access to extensive information from across Europe, in particular about people, projects, institutions, events, publications, etc., of relevance to system resilience researchers in both the dependability and the security domains, and to their sponsoring organisations.
>
> Deliverable D10 (month 12) provided a prototype of the RKB, which clearly showed the potential of the expected final system and in practice went further than was originally planned. The range and size of data sources that were included was an order of magnitude greater than had been expected at this stage, and facilities for adding specialist information, such as courseware information, had been made available to the project. The task also provided a project wiki with a semantic infrastructure.
>
> Specifically, there is now an RKB that has data from sources covering the whole of Computer Science and related topics, both within Europe and beyond. The "Explorer" interface was well received by project members who conducted a review, although a number of possible improvements were identified. Details of the prototype RKB can be found in deliverable D10.
>
> Task IT-T1 is now concerned with gaining value from D10, as well as supporting the resilience-explicit and resilience ontology activities.
>
> The aims of the next phase of the RKB work are twofold:
>
> 1. Maintenance and Consolidation
>
> Content – The resources needed to maintain the RKB information content rise with its magnitude; the success of the first phase has delivered much more content than was envisaged initially, and there are significant additional sources that are targeted for acquisition (IEEE material is already identified).
>
> Software – software systems need to be kept up to date; this relates to operating systems, web server software and triplestore system software.
>
> Hardware – it can be expected that the demand load on the RKB will rise, which may necessitate work to enhance the hardware base (with some more hardware purchased); the much less attractive alternative would be to restrict access and usage.
>
> Specific – Special requirements in the first phase (the wiki, the coursework page, etc.) imposed additional tasks and it is recognised that more work of this type will be needed as the Network continues to make progress; for example, a need has already been identified for a page to automatically gather the Resilience Mechanism Questionnaire data.
>
> 2. ResEx and ResOn support
>
> Support for the next phases of resilience-explicit work was planned initially and will continue, but as the new resilience ontology strand develops, the partners will need further support on both *acquisition* and *publishing*. Acquisition involves adding metadata about components and processes and their attributes to the RKB, as well as related documents. Appropriate ways to publish this data, to users and systems that are engaged in Res-Ex processes, will need to be devised

> At the prototype stage (i.e. D10), the extent to which resources in the RKB are classified according to resilience-related function is less specific than is desirable, and the RKB can only classify resources where other systems have identified concepts of interest to them that are related to resilience. To improve on this, the deployment of a system that classifies documents according to their relation to the detailed resilience terms embedded in the RKB would be required. Deploying such classification technologies is a significant task, but would yield considerable additional value. The new resilience ontology task has been created to identify and refine resilience concepts, and is also exploring the challenge of deploying appropriate classifiers, albeit from outwith the ReSIST Project. Supporting the emerging classifications in order to gain greatest benefit from this work, and any classifying systems that use them, will be an exciting challenge.

As it was with D10, the primary components of deliverable D23 are the RKB and associated service pages, and this document discusses their content.

2 – RKB Content

The primary function of the RKB is to provide appropriate content for the project and researchers in resilient systems. D23 provides additional content both acquired from outside sources and generated within the project. It also represents a considerable improvement in the quality of the data over the prototype D10.

2.1 IEEE

The project identified the IEEE FTCS/DSN & Oakland series of conferences as an important specialised resource. The RKB therefore now has the metadata for the series.

This has been used to facilitate the generation of further content by inviting members of the project to establish the relationship between these papers and subject classifications, in particular the resilience and ACM ontologies. The specialised page that was provided for this is shown in Figure 1 and can be found from http://resist.ecs.soton.ac.uk/classifier/manual/; an example of the data gathered is shown in Figure 2.

This activity has led to the training set based on the data collected, which is now being used by IT-T3 to build the taxonomy and classification services.

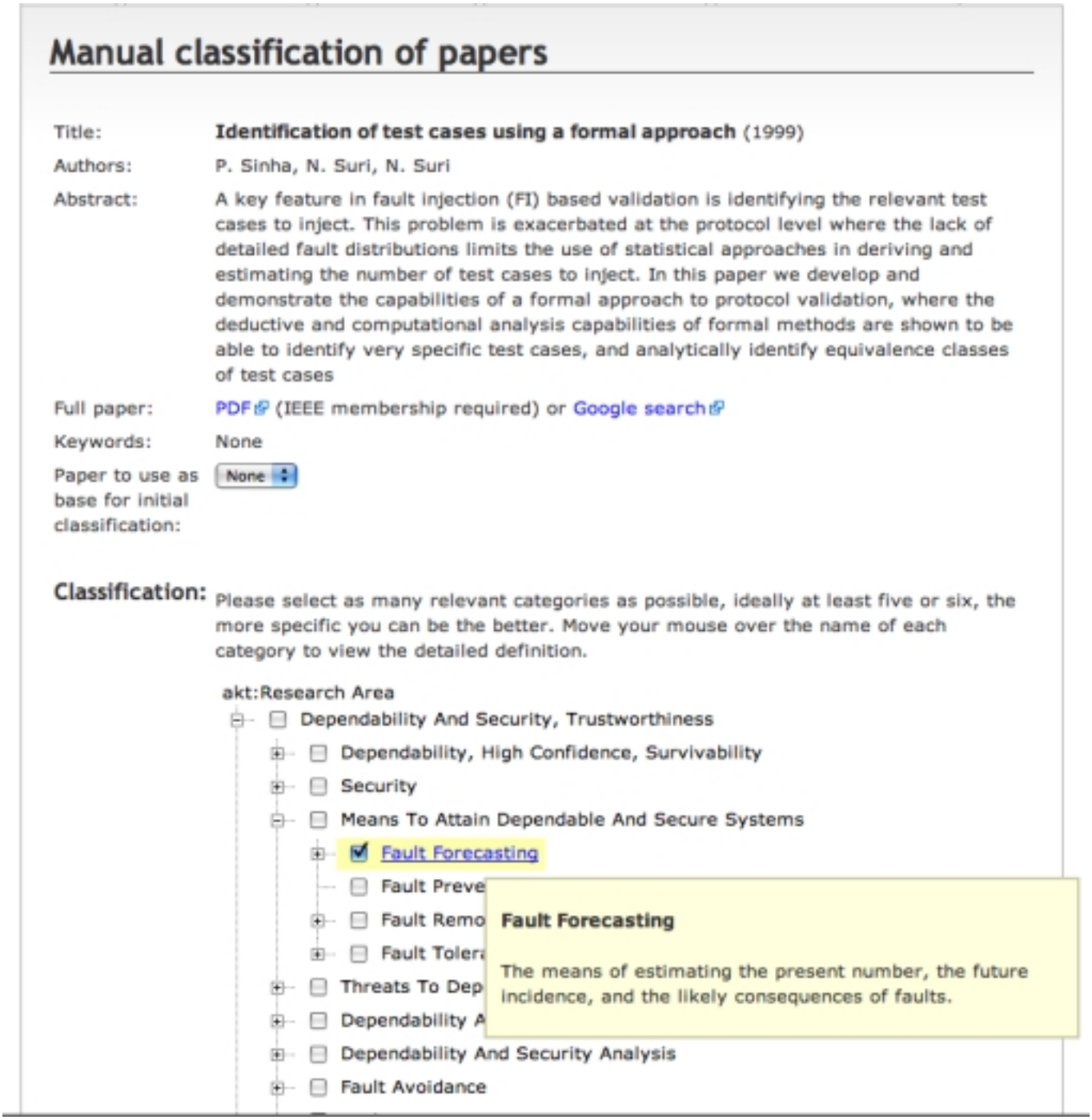


Figure 1 - Paper Classification Form

RKB Browser :: Fast Byzantine consensus

Alternative representations
» RDF export

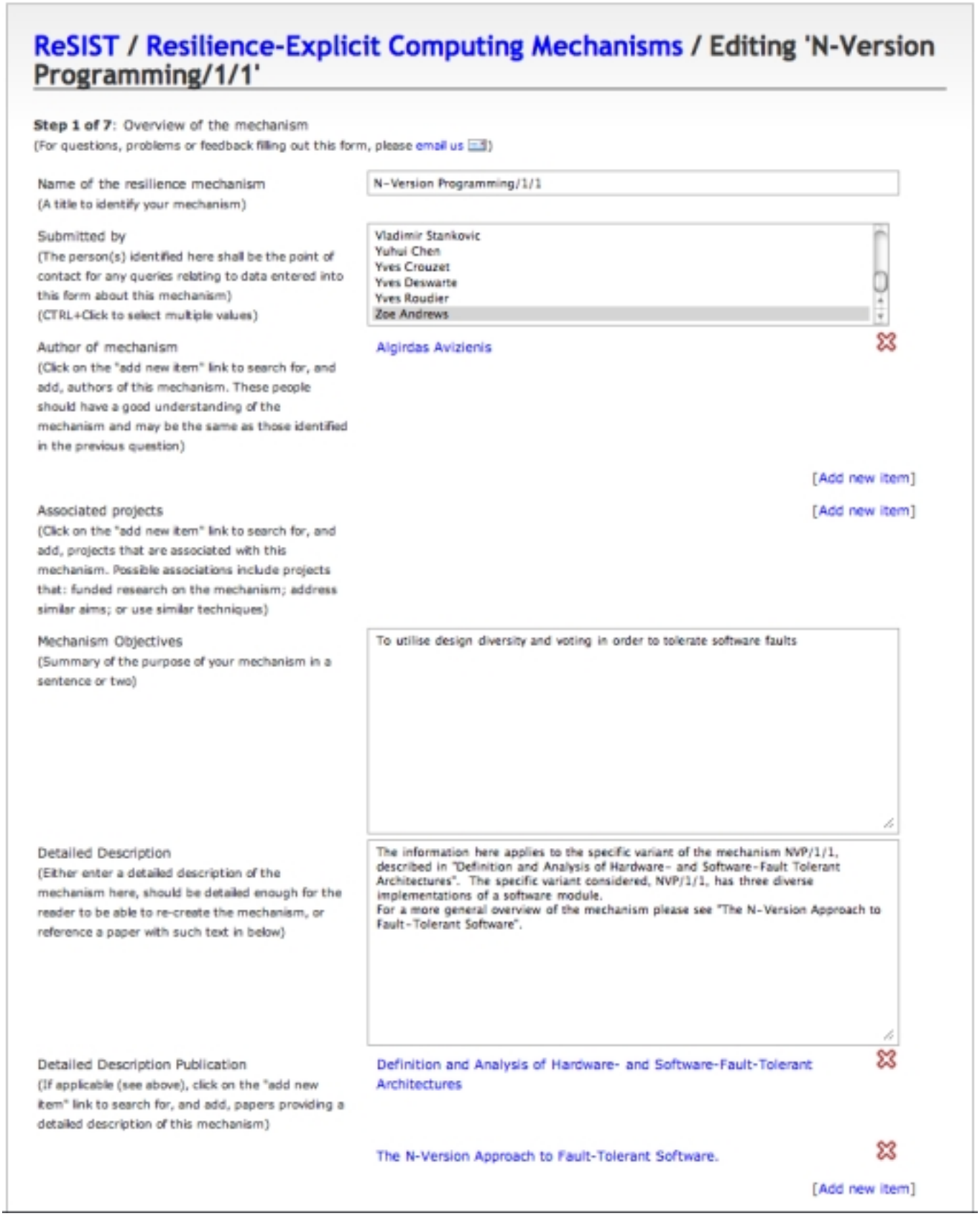
Identifiers...
http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=01467815

Subject	Property	Object/Value	Source
Fast Byzantine consensus	http://www.aktors.org/ontology/extension#has-abstract	We present the first consensus protocol that reaches asynchronous Byzantine consensus in two communication steps in the common case. We prove that our protocol is optimal in terms of both number of communication step, and number of processes for 2-step consensus. The protocol can be used to build a replicated state machine that requires only three communication steps per request in the common case.	ieee-dsn-publications.rdf >>
Fast Byzantine consensus	akt:addresses-research-area	Dependability, High Confidence, Survivability	7e4dcd9f.ntriples >>
Fast Byzantine consensus	akt:addresses-research-area	Fault Masking	7e4dcd9f.ntriples >>
Fast Byzantine consensus	akt:addresses-research-area	Fault Tolerance, Resilience, Self Repair, Self Healing	7e4dcd9f.ntriples >>
Fast Byzantine consensus	akt:addresses-research-area	Means To Attain Dependable And Secure Systems	7e4dcd9f.ntriples >>
Fast Byzantine consensus	akt:addresses-research-area	Security	7e4dcd9f.ntriples >>
Fast Byzantine consensus	akt:has-author	L. Alvisi	ieee-dsn-publications.rdf >>
Fast Byzantine consensus	akt:has-author	J.-P. Martin	ieee-dsn-publications.rdf >>
Fast Byzantine consensus	akt:has-date	2005	ieee-dsn-publications.rdf >>
Fast Byzantine consensus	akt:has-title	Fast Byzantine consensus	ieee-dsn-publications.rdf >>
Fast Byzantine consensus	akt:has-web-address	http://ieeexplore.ieee.org/iel5/9904/31476/01467815.pdf [Visit]	ieee-dsn-publications.rdf >>
Fast Byzantine consensus	akt:paper-in-proceedings	International Conference on Dependable Systems and Networks, 2005	ieee-dsn-publications.rdf >>

Figure 2 – IEEE Paper Metadata

2.2 ResEx – Resilience-Explicit Computing

Further content was provided by way of the ResEx activity, and facilities to support this are provided. Apart from the simple provision of data, new metadata types can be added to the RKB, and complex structures can be acquired and represented. Figure 3 shows the form being used, and Figure 4 shows a part of the metadata collected. These can be accessed by following the links from http://resist.ecs.soton.ac.uk/resex/.



ReSIST / Resilience-Explicit Computing Mechanisms / Editing 'N-Version Programming/1/1'

Step 1 of 7: Overview of the mechanism
(For questions, problems or feedback filling out this form, please email us)

Name of the resilience mechanism
(A title to identify your mechanism)
N-Version Programming/1/1

Submitted by
(The person(s) identified here shall be the point of contact for any queries relating to data entered into this form about this mechanism)
(CTRL+Click to select multiple values)
Vladimir Stankovic
Yuhui Chen
Yves Crouzet
Yves Deswarte
Yves Roudier
Zoe Andrews

Author of mechanism
(Click on the "add new item" link to search for, and add, authors of this mechanism. These people should have a good understanding of the mechanism and may be the same as those identified in the previous question)
Algirdas Avizienis
[Add new item]

Associated projects
(Click on the "add new item" link to search for, and add, projects that are associated with this mechanism. Possible associations include projects that: funded research on the mechanism; address similar aims; or use similar techniques)
[Add new item]

Mechanism Objectives
(Summary of the purpose of your mechanism in a sentence or two)
To utilise design diversity and voting in order to tolerate software faults

Detailed Description
(Either enter a detailed description of the mechanism here, should be detailed enough for the reader to be able to re-create the mechanism, or reference a paper with such text in below)
The information here applies to the specific variant of the mechanism NVP/1/1, described in "Definition and Analysis of Hardware- and Software-Fault Tolerant Architectures". The specific variant considered, NVP/1/1, has three diverse implementations of a software module.
For a more general overview of the mechanism please see "The N-Version Approach to Fault-Tolerant Software".

Detailed Description Publication
(If applicable (see above), click on the "add new item" link to search for, and add, papers providing a detailed description of this mechanism)
Definition and Analysis of Hardware- and Software-Fault-Tolerant Architectures
The N-Version Approach to Fault-Tolerant Software.
[Add new item]

Figure 3 – Resilience-Explicit Mechanism Input Form (Page 1)

N-Version Programming/1/1	resex:assessability-comment	For n-version programming to be effective you need diverse implementations of the same software module. Thus you should be able to assess the level of diversity between software modules to attain confidence that the modules are in fact diverse.
N-Version Programming/1/1	resex:description-of-variants	The basic n-version programming architecture is more general than the NVP/1/1 mechanism described here, allowing more versions and variations in the voting threshold and hardware platform. A number of extensions and variations of the basic architecture have also been proposed. Clearly, these variations have a significant influence on the fault-tolerant capabilities and cost overheads of this approach.n"Definition and Analysis of Hardware- and Software-Fault-Tolerant Architectures" describes some of these variants.
N-Version Programming/1/1	resex:detailed-description-pub	Definition and Analysis of Hardware- and Software-Fault-Tolerant Architectures
N-Version Programming/1/1	resex:detailed-description-pub	The N-Version Approach to Fault-Tolerant Software.
N-Version Programming/1/1	resex:detailed-description	The information here applies to the specific variant of the mechanism NVP/1/1, described in "Definition and Analysis of Hardware- and Software-Fault Tolerant Architectures". The specific variant considered, NVP/1/1, has three diverse implementations of a software module.nFor a more general overview of the mechanism please see "The N-Version Approach to Fault-Tolerant Software".
N-Version Programming/1/1	resex:diagrams-pub	Definition and Analysis of Hardware- and Software-Fault-Tolerant Architectures
N-Version Programming/1/1	resex:diversity-comment	N-version programming uses diverse implementations of a software module in order to improve its reliability.
N-Version Programming/1/1	resex:evolvability-comment	N-version programming provides support for evolution by allowing a system designer to replace deployed variants of the software module in the n-version programming architecture with suitable new ones. Such variants could even be discovered at runtime and changed over whilst the system is still running.
N-Version Programming/1/1	resex:formal-description-pub	Software Fault Tolerance
N-Version Programming/1/1	resex:has-additional-requirements	The software variants, whilst having diverse implementations, should all provide suitably similar service. In certain cases it is essential that these variants are functionally identical.
N-Version Programming/1/1	resex:has-application-domain	J.1.0. Business
N-Version	resex:has-	J.1.7. Military

Figure 4 – Resilience-Explicit Mechanism Metadata Fragment

2.3 The RKB

There has been significant reprocessing of the data that was gathered in the prototype phase, reflecting lessons learnt, as well as changes in the technical context. This reprocessing uncovered some further problems that proved to be more challenging than might have been expected.

The reprocessing falls into three categories, each of which supports the needs of the others, as can be seen in the descriptions later:

1) Building of a separate knowledge base (KB) for each source of the data;
2) "Minting" of new URIs for each concept constructed;
3) Identification and management of multiple URIs that refer to the same concept (the co-reference problem).

Separate KBs

The RKB is now divided into a separate KB for each source. This has a number of advantages, including:

individual KBs no longer needs to be co-located with the other KBs;

the KB can be moved to the data provider's site, or elsewhere;

KB management can be delegated and is easier;

the RDF can be accessed by resolving URIs (see below);

RDF assertion can be more efficient;

querying can be localised, reducing the possibility of future scalability problems.

More detailed access to the RKB data can be found from the main page at http://www.rkbexplorer.com/explorer/. From here access is available to the 25 separate but linked knowledge bases that can be used by the RKB and Explorer.

For each KB, a base page can be found which gives a summary of the KB and lists the facilities available for access. Taking the example of the ieee.rkbexplorer.com KB, they are listed as follows:

1) All the URIs used are resolvable. This, along with the SPARQL query interface (see (3) below), enables the RKB data to be exposed for anyone to use, as well as the Explorer and other ReSIST facilities;

2) It is possible to browse the particular KB, using a raw triple browser;

3) A SPARQL query interface is provided, exposing the RKB in a form that can be used by external applications;

4) The linkage of this KB with other ReSIST KBs and beyond is exposed in the Consistent Reference Service (CRS);

5) The full RDF is also published, along with Sitemap information to allow search engines to index it efficiently.

New URIs

It became apparent during the building of the prototype that few of the data providers had appropriate URIs that could be used for concepts in the knowledge base. Where they did, it was not possible to guarantee that the meaning of the URI would not change, or even if it was correct (see DBLP below). It therefore became necessary to establish our own URI management, so that we could control the meaning of the data we were effectively publishing. We therefore chose to "mint" a new URI for each concept, placed at an appropriate sub-domain referring to a data source KB, as above.

As an example, one of the IEEE papers has a URI of http://ieee.rkbexplorer.com/id/publication-00466948. This is a real URL that can be used to access the metadata we hold about the paper. This metadata includes the fact that IEE have a copy of the paper at http://ieeexplore.ieee.org/iel5/3246/9797/00466948.pdf. In fact, the system supports content negotiation on our URI: it will return *html* or *rdf* descriptions, according to the content type requested by the requesting agent (using a 303 redirect). This conforms to the latest standards in the Linking Open Data initiative.

The use of such resolvable URIs exposes the RKB knowledge in an effective way, avoiding the need to engage in SPARQL queries or process large RDF files, as before. For example the Tabulator RDF browser (http://www.w3.org/2005/ajar/tab) can explore the RKB by starting with a URI that returns RDF directly, such as http://ieee.rkbexplorer.com/data/publication-00466948 .

Identification of multiple URIs

In a KB such as the RKB, when multiple datasets are combined, the URIs from different sources must be processed to determine if they in fact refer to the same real world entities. For D10 we had deployed tools that performed this analysis, and made the results available to the RKB Explorer.

The most important and disturbing lesson learnt from the prototype was that, although these tools were effective, the data they were working on was far worse than expected.

In a publication resource such as DBLP (http://www.informatik.uni-trier.de/~ley/db/), it is widely appreciated that the team has worked hard to ensure that each person is only the subject of one record, and that each person has a record of their own. When a human browses the data, they find this is largely true, and are able to discount the cases where mistakes have been made. Unfortunately, when such data is aggregated with other sources, the network effect which is being created can cause errors to be more significant and pervasive.

For example, when we looking at a particular CORDIS project, the RKBExplorer indicated that there were some NSF projects that were related, without apparent reason. Further investigation showed the reason: "Tom Anderson" was named as investigator on both. Clearly these were different "Tom Anderson"s, but unfortunately DBLP had joined the publications from the two individuals as if they were one (a single URI for both), and so the RKB had accepted the judgement. Further investigation revealed that the problem was so widespread that the only way to use this valuable resource was to ignore such judgements that DBLP was making altogether.

The result was that our system was re-engineered to generate a new URI for each author name on each paper, each paper, and many other concepts, and the RKB took on the task of doing the entire co-reference resolution. Such a "cold start" co-reference activity is in fact different from the incremental system available, and so a new system has been built and is now deployed.

Each KB was augmented with an associated Consistent Reference Service (CRS) to make this co-reference knowledge available to the RKBExplorer and other users.

An example of the newcastle.rkbexplorer.com CRS entry for Brian Randell is given in Figure 5. It can be seen that the URI that has been generated from the Newcastle source has been established as being the same as many URIs representing authors with the string "Brian Randell" (or similar) in the DBLP source. It can also be seen that the IEEE papers have the same author, and both LAAS and Ulm University have provided information including references to him. Finally, the CRS connects all these up with the appropriate person in the project wiki.

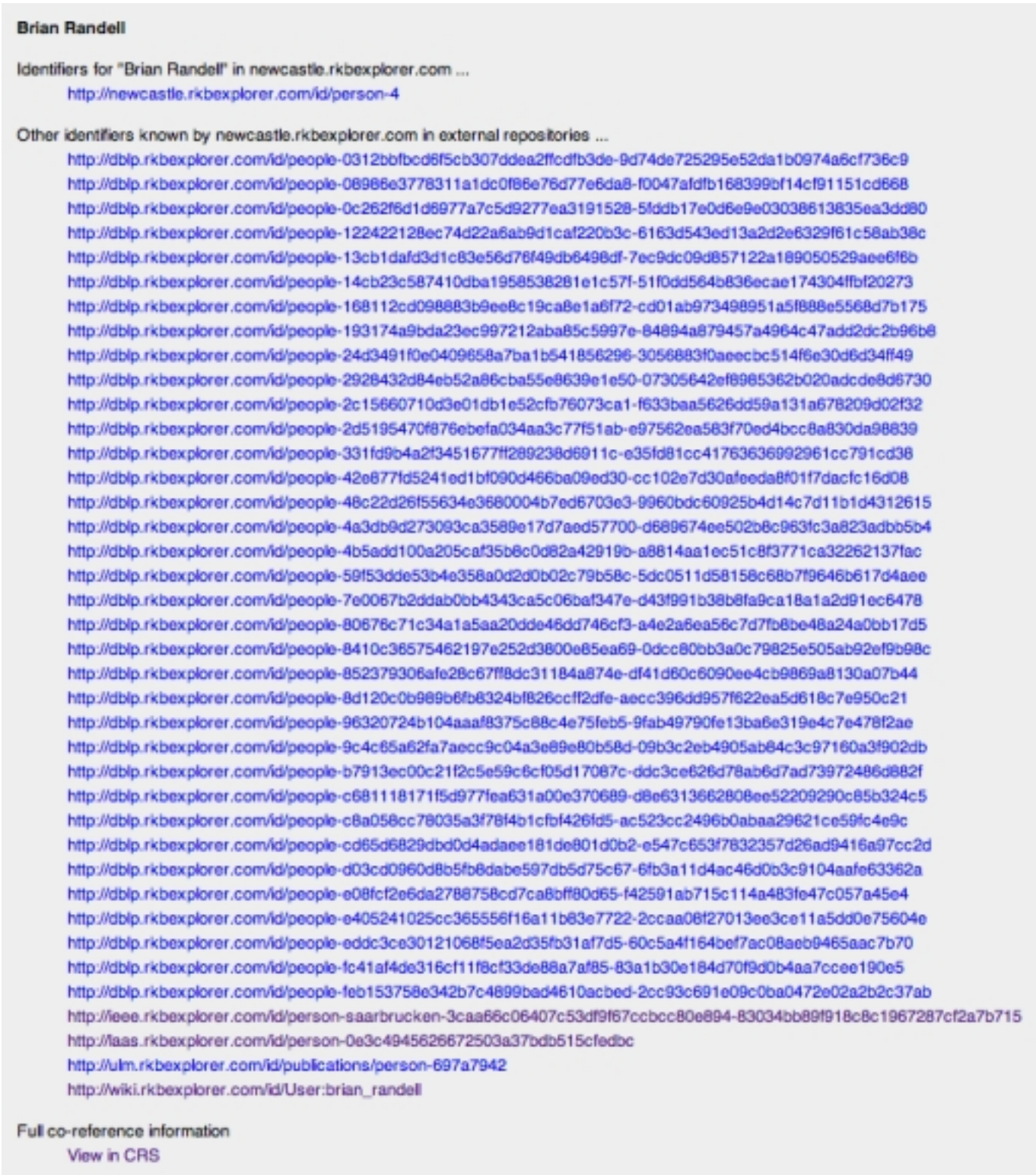


Figure 5 – CRS entry in newcastle.rkbexplorer.com for Brian Randell

3 - RKBExplorer

A revised version of the exploration user interface to the RKB, running on the newly-commissioned hardware, can be found at http://www.rkbexplorer.com/explorer/; details of this interface, in the form of a User Guide, were given in D10, and can be found at the "Help" link from that page. It was revised to conform to the new project visual identity, and minor changes were made in response to user comments.

The main change that can be seen from a user's point of view is that the data displayed in the interface has become more specialised towards the project subject domain, by the inclusion of ResEx Mechanisms.

The main change that is not clear to see is that it has been re-engineered to present a unified view of the many KBs that make up the RKB (rather then the single RKB previously). This means that it now also has the potential to include KBs that are from other sites and other projects.

Figure 6 shows the first screen of the latest version of RKBExplorer.

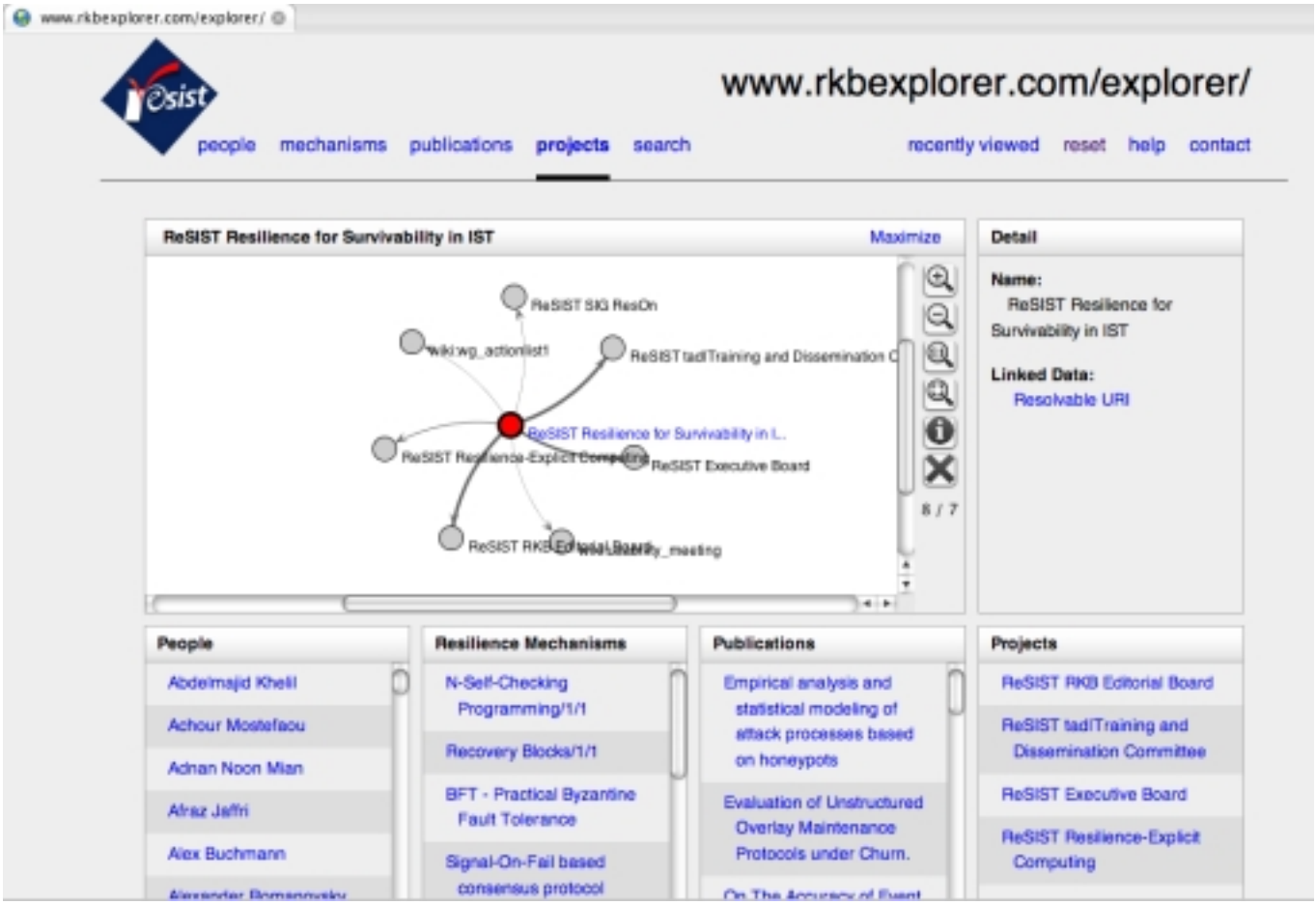


Figure 6 – RKBExplorer

4 – Conclusions

We have presented here the status of the RKB and associated mechanisms at the end of year two. The content has been maintained and expanded, in general terms and specifically, according to the needs of the project from time to time. Considerable work has been completed to improve the infrastructure so that those past and the future requirements can be met efficiently.

Future work will continue along the same lines so that the quality and coverage of the knowledge is improved, as well as the inferences that can be drawn from that knowledge.