Supporting Collaborative Grid Application Development Within The E-Science Community

Supporting Collaboration within the e-Science Community

Cornelia Boldyreff, David Nutter & Stephen Rank

http://www.lincoln.ac.uk/faculties/computing/index.html

Introduction

Collaboration is at the heart of and key to the success of scientific endeavour, but much current Grid research until now has focussed on the Grid technology itself and the experiments themselves rather than on the people who must collaborate to make them happen!

- Background
- Vision
- Method
- Deployment and Evaluation

Background Work(1)

Representation and organisation of software artefacts including

- Specifications
- Design contracts/interfaces
- Implementations

is important for developing large distributed systems from software components and:

- Software re-use
- Program comprehension
- Software maintenance and evolution

Background Work(2)

At Durham two projects based on classic software decompositions studied this.
The AMES and Practitioner projects were environments supporting program comprehension, re-use and evolution:

- Traditional SEEs
- Organised artefacts, not people.
- Did not address dynamic aspects of SE (e.g. web services)
- Based on classic software decomposition theory (Goguen)

Background Work(3)

Following on from this work, the CARD project supported collaborative design in the steel industry

- A matrix containing constraints and roll design parameters
- Visible and changeable by all users

The CoDEEDS project extends this concept:

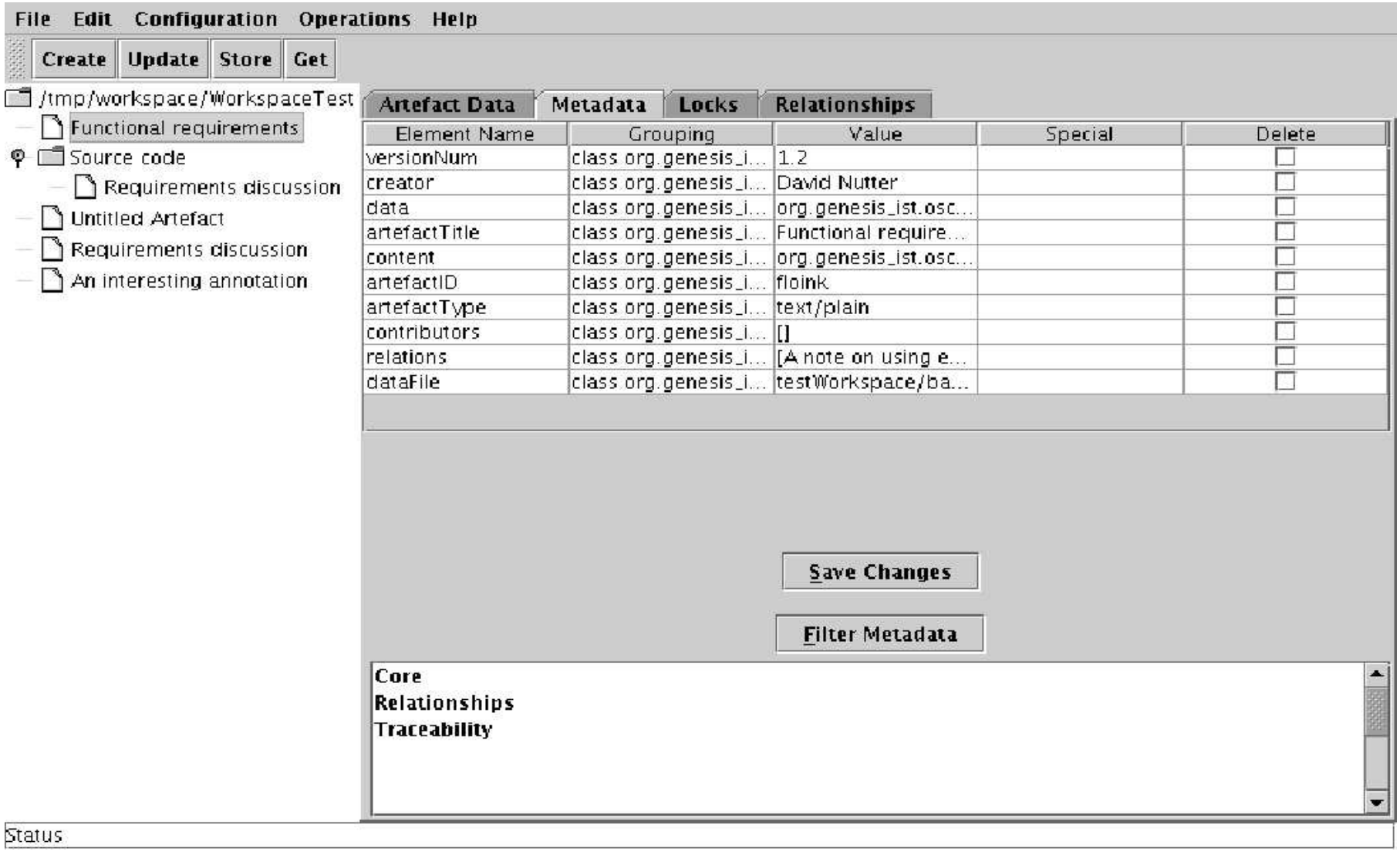
- Matrix contains software design constraints applied to compositions of software components
- Generic framework for matrix-based design spaces
- Employs OSCAR to store software artefacts
- Addresses dynamic composition not dealt with by AMES/Practitioner

Background Work(4)

OSCAR (Open Source Component Artefact Repository) is a component of the GENESIS platform:

- General process-aware co-operative SEE
- OSCAR stores artefacts from the Workflow and Resource management components
- Extensible artefact metadata supports recording of:
  - Relationships to other artefacts
  - Relationships to developers, tools, process etc
  - Changes to the artefacts
- Artefacts actively inform each other and users of changes.
- Awareness support is being introduced

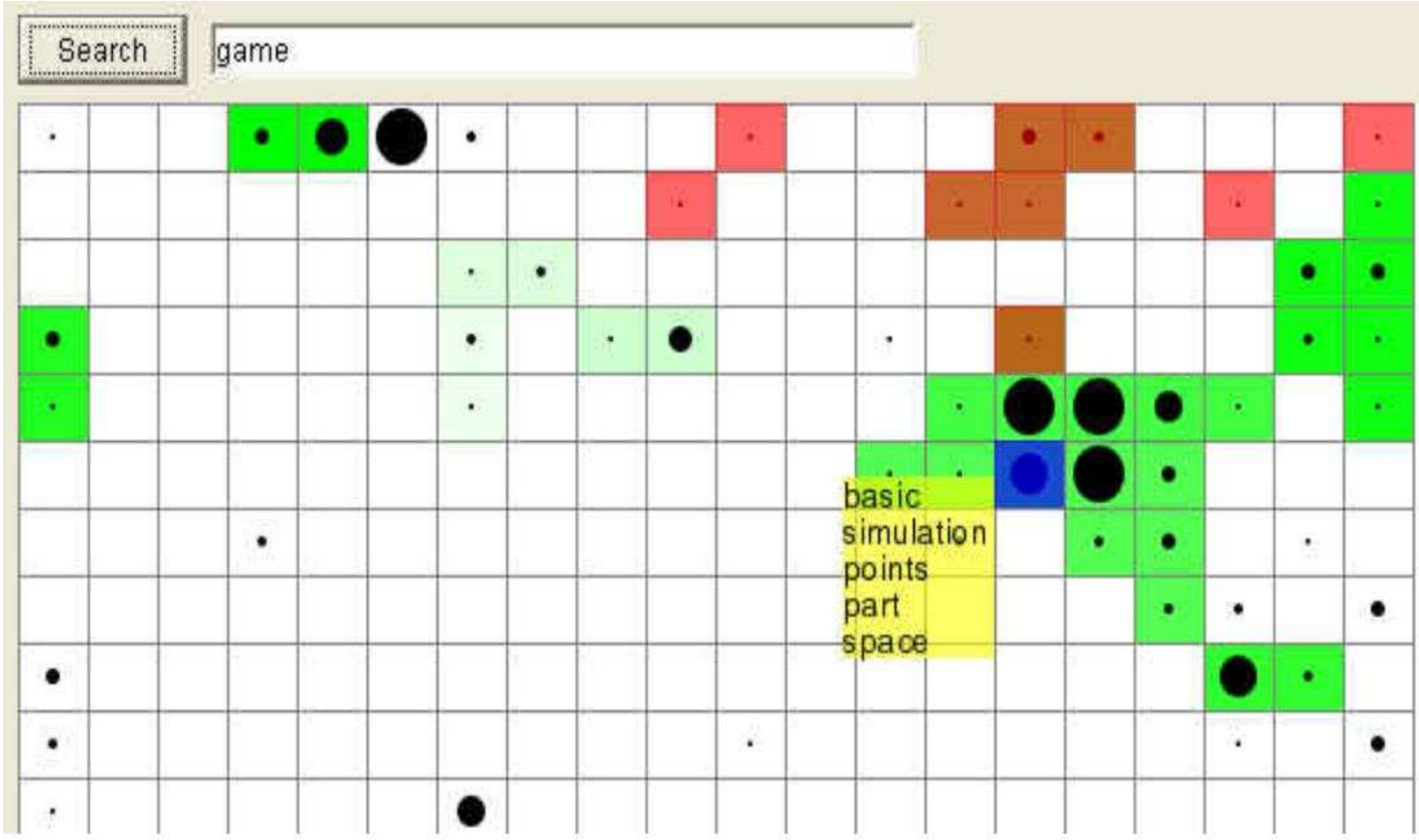
OSCAR Client Screenshot



Background Work(5)

- Search support for OSCAR based on extensions to Kohonen's WebSOM
- Uses a Growing Heirarchical Self Organising Map (GHSOM) to organise arbitrary collections of software artefacts.
- Test population drawn from Debian Packages data and own Java source code.
- Solves the problem of incrementally and automatically organising large collections of data
- Speculatively:
  - Using SOM techniques to organise networks of OSCAR clients to distribute awareness information
  - Organising the artefacts themselves according to some map.

GENISOM Screenshot



Vision(1)

To extend the support offered by OSCAR to Software Engineering to general scientific work.

- Semantic interoperability between disparate grid artefacts including:
  - computing resource descriptions
  - data sources and processing tools
  - profiles of scientists
  - experiment designs and rationale
- Facilitate collaboration and discourse involving artefacts
- Handle “provenance” of artefacts for accountability

Vision(2)

The main aims are:

- Develop a model of online collaboration in e-Science
- Develop environment(s) implementing this model and deploy them in practice in order to refine it.

Other objectives contingent on the above:

- Handle legacy data and applications
- Record design decisions & rationale for experiments and collaborations.
- Handle existing forms of collaboration
- Accountability, e.g facilitate replication of experiments

Method(1)

Speculative elements of the proposed research include:

- Developing implementing new collaboration techniques
- Developing semantic interoperability for new, scientific, artefact types. This will necessarily build on our types developed for Software Engineering.

However, we will also incrementally improve our previous work:

- Use of design spaces, a concept proven by earlier research.
- Employ existing research prototypes (e.g. existing integration of CoDEEDS and OSCAR)
- Obtain feedback at all stages of development.

Method(2)

- Our previous research has been very domain-specific. This research attempts to elicit general models for collaboration in artefact-management environments which may be engaged in eScience, software engineering or other tasks.
- Our implementations will be based on existing environments, notably GENESIS and CoDEEDS.
- To this end we will use Industry-As-Laboratory to evaluate our work.

Industry as Laboratory(1)

CARD and GENESIS have both involved industrial partners with real needs addressed by the software.
Key points:

- Feedback incorporated into future versions of the software
- Software released as Open Source to facilitate adoption and growth of user communities (OSCAR/GENESIS)
- Evolutionary development of tools and theory using XP/Agile methods.

Industry as Laboratory(2)

We intend to work with researchers in Lincoln's Faculties of Health and Life Sciences to rapidly develop prototypes. The process will include:

- Rapid release of prototypes, eventually as Open Source
- Parallel evolution of models and corresponding software
- Use of existing Grid architecture and services
- Regular meetings with clients to discuss requirements
- Propogation of results via All Hands eScience meetings and related workshops.

These techniques have all been used with success in our previous projects.

Evaluation(1)

Aside from feedback from the eScience community and the direct testing against current requirements of our prototype we will evaluate our model and implementation as follows:

- Initially, confirm that Design Spaces can be applied to the domain of scientific experiments
- Deploy the resulting prototype to gain feedback
- Develop models of collaboration by observing collaboration between scientists
- Evaluation of the resulting applications using the Ramage Onion Model and the extensions extensions suited to web-based groupware.

Conclusion

- The maintainability of the system must be evaluated alongside usability. An system that can be maintained easily has a much better chance of avoiding the “software tools graveyard” and being widely adopted .
- The recent development of grid and grid-related software as focused on the core technology
- We propose a complimentary track of research focused on support for the collaborating necessary to ensure effective long term grid applications within the scientific community.