Chain ReAKTing: Collaborative Advanced Knowledge Technologies in the Combechem Grid

Michelle Bachler[1] Simon Buckingham Shum[1] Jessica Chen-Burger[2]
Jeff Dalton[2] **David De Roure**[3] Marc Eisenstadt[1] Jeremy Frey[3]
Jiri Komzak[1] Danius Michaelides[3] Kevin Page[3] Stephen Potter[2]
Nigel Shadbolt[3] Austin Tate[2]

[1] KMI, The Open University, UK
[2] AIAI, University of Edinburgh, UK
[3] ECS, University of Southampton, UK

Abstract

The CoAKTinG (Collaborative Advanced Knowledge Technologies in the Grid) project has developed a set of integrated tools to enhance collaboration between e-Scientists. As one of three case studies, these tools are being applied within the Combechem e-Science pilot project. Two levels of integration are being explored: straightforward deployment of generic CoAKTinG tools, and a "deep" integration between these tools and the Combechem grid. The deeper integration supports the publication at source research objective of Combechem, in which a digital record is maintained through the information processing chain that starts in the laboratory, supporting retrospective use in the e-Science process. In this paper we provide an overview of the tools and we focus in particular on the adaptation of one of the tools for the Combechem application.

1 Introduction

e-Science applications involve several dimensions of Grid computing, including the computational grid, data grid and collaborative grid. In this paper we focus on the collaborative grid. A collaborative grid infrastructure is not just an add-on but rather it has become part of the essence of Grid computing - it facilitates the creation and functioning of dynamic virtual organisations.

The importance of collaboration tools is exemplified by the tools in common use such as access grid and VRVS. In this paper we bring aspects of 'Semantic Grid'[4] thinking to the collaborative Grid infrastructure - our emphasis here is on interoperability and reuse.

2 CoAKTinG

The objective of CoAKTinG ('Collaborative Advanced Knowledge Technologies in the Grid') has been to advance the state of the art in collaborative mediated spaces for distributed e-Science through the novel application of advanced knowledge technologies. It comprises four tools: meeting capture and replay, instant messaging and presence notification (BuddySpace), graphical meeting and group memory capture (Compendium) and intelligent 'to-do' lists (Process Panels). These are integrated through exchanging and storing events according to a set of CoAKTinG ontologies. In its last phase, CoAKTinG has been conducting a number of case studies. One of these has been with Combechem, to put the tools in the hands of e-Scientists. The discussions between the two teams have led to two notions of integration: "shallow", where the tools are deployed as they are, and "deep" where the tools are more intimately integrated with the Combechem systems.

3 Combechem

The Combechem project aims to enhance structure property correlation and prediction by increasing the amount of knowledge about materials via synthesis and analysis of large compound libraries. Automation of the measurement and analysis is required in order to do this efficiently and reliably while ensuring that wide dissemination of the information occurs together with all the necessary associated background (raw) data that is needed to specify the provenance of the material. The project aims for a complete end-to-end connection between the laboratory bench and the intellectual chemical knowledge that is published as a result of the investigation; this necessitates that all steps in the process are enhanced by a suitable digital environment. Combechem has achieved many parts of this ambitious programme, e.g. the smart laboratory (smarttea.org), grid-enabled instrumentation, data tracking for analysis, methodology for publication@source, process and role based security and high throughput computation.

The CoAKTinG tools provide support for the e-Science process in Combechem and they also enable the digitisation of missing links in the processing chain which form part of the typical collaborative scientific processes that we are attempting to enhance using the grid infrastructure: support of the experimental process, tracking and awareness of people and machine states, capturing of the discussions about data as well as the traditional metadata, and enriched meta-data regarding these components to support interlinking.

4 CoAKTinG Tools

The CoAKTinG project has developed and integrated four tools, which are described in the following sections. The tools integrate at the data level via an ontology that describes meetings and events that occur. This ontology is part of a wider ontology that describes academic research [1] and so the tools are not limited to describing facts about meetings. Tools communicate using Jabber protocol to pass information and other control information, for example to synchronize the display of two tools.

4.1 BuddySpace

BuddySpace is an Instant Message client (based on the Jabber protocol) with features that enhance presence awareness. Specifically, it introduces the graphical visualization of people and the presence on a image or map, as can bee seen in the figure. This allows for multiple views of collaborative workgroups and the immediacy or “at a glance” nature gives users a snapshot of a virtual organisation.

In a meeting, the instant message capabilities of BuddySpace naturally provide a “back-channel” to the meeting, for example, conveying URLs of documents discussed or as a non-disrupting communication. For distributed meetings, such Access Grid meetings, the presence of individuals gives an extra indication of co-location (especially if the videoconferencing technology is failing). The back-channel can also be used for meeting control tasks, such as queuing of speakers and voting on issues. For meeting capture purposes, logs of the channel conversations are made. Individual messages are timestamped and possibly examined to see if they control meeting specific messages.

For e-Scientists, like another other distributed group of collaborators, the awareness of availability and activity of members of the team is crucial. This is especially true in the Combechem scenario where researchers work in a number of different locations, with varying levels of availability, for example in the “wet” lab at a crucial step in a synthesis experiment.

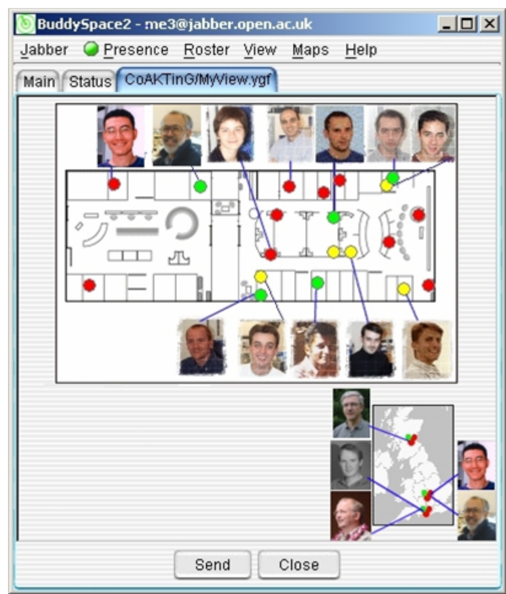

Figure 1: BuddySpace showing a virtual organisation and presence indicators

4.2 I-X Process Panels

I-X is a suite of tools[10] whose function is to aid in processes which create or modify one or more “product” (such as a document, a physical entity or even some desired changes in the world state). The main interface is the I-X Process Panel (I-P2) which, in its simplest form, acts like an intelligent “to do” list. The panel shows users their current issues and activities, on which Standard Operating Procedures can be applied to manage complex and long-running processes. I-X also has a collaborative element to it, in that issues and activities can be passed between different process panels to enact a workflow across an organisation. Web services can be called to automatically enact steps of the processes involved. Progress and completion reporting between panels and external services is possible. The underlying model on which I-X is based is the <I-N-C-A> Constraints Model[9]. In a meeting scenario, actions raised in a meeting have a direct mapping to <I-N-C-A> activities. Actions created in a meeting specific I-X panel are passed onto the relevant user panel's for individuals, which, on completion report back.

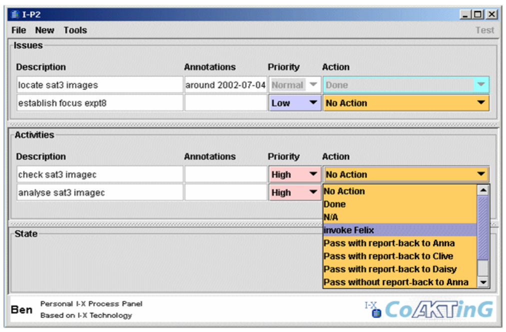

Figure 2: A I-X Process Panel showing pending issues and activities

4.3 Compendium

Compendium, first developed in 1993 as an approach to aid cross-functional business process redesign (BPR) teams, has been applied in several dozen projects in both industry and academic settings [2]. Its origins lie in the problem of creating shared understanding between the team members, typical of those attending teams working over weeks or months to design business processes: keeping track of the plethora of ideas, issues, and conceptual interrelationships without needing to sift through piles of easel sheets, surfacing and tracking design rationale, and staying on track and "bought-in" to the project's overall structure and goals [6]. The key feature of the early approach was the combination of an Issue-Based Information System (IBIS) concept-mapping tool [3], which supported informal and exploratory conversation and facilitation, with a structured modelling approach [5]. This allowed teams to move along the spectra of *formal to informal representation*, and *prescribed to spontaneous approaches*, as their needs dictated. It also let them incrementally formalise data [8] over the life of the project. As the approach was tested and refined over the course of several years, additional modelling methods were added, plus tools to transform Compendium's hypertext models into established organisational document forms, and vice-versa [7].

In our experience, Compendium introduces a distinctive element to the design space of knowledge technologies, namely, making meetings into true events for group knowledge creation which leave a *trace* - a structured, collectively owned, searchable group memory that is generated in real time as a product of a meeting. Effective, on-the-fly construction of knowledge resources does not come "for free" - the lower the effort invested at the capture stage (e.g. simply video recording all meetings, or taking conventional minutes), the more work is required for collective reuse and computational support. Naturally, we want quality knowledge resources for minimal effort, and while smart analysis technologies will continue to push the boundaries, there are pragmatic factors to consider: what is possible *now*? Compendium tackles the capture bottleneck that any knowledge construction effort must confront, by investing effort in real time quality capture by a facilitator, mediated and validated by those at the meeting.

Compendium is a key component in the publication@source pipeline, allowing scientists to capture their discourse about a particular experiment and its results. This is crucial link between the act of performing an experiment and its inclusion in a paper, which a reader could follow to better understand the author's intentions.

4.4 Meeting Replay

Once a meeting has taken place it can be useful to revisit the ideas and topics discussed. Traditionally, formal minutes are taken to record the salient points, but often these are too brief to be more than a simple aide memoire; in the typical CoAKTinG scenario (such as an Access Grid node) full audio and video logs are available, but conversely these are too verbose to be of practical use. We require the ability to select high-level points of reference from the meeting, then "zoom in" to view detailed records. e.g. a user sees from Compendium notes that a decision was made, but to understand the subtle reasoning behind that outcome wishes to view the video of discussion between participants. Each meeting is described using RDF conforming to the OWL meeting ontology; this represents resources such as: the meeting time, location, attendees, audio/video recordings, any presentations given (and associated web viewable versions), and argumentation annotation from Compendium. The Event / has-sub-event structure held within the RDF is mapped onto a more conventional time-line, which is automatically published using HTML and Javascript on a web site (figure 4.4). The user can navigate the meeting using the video timeline, or jump to a different point in the meeting by selecting a particular event, such as a slide being presented, or a Compendium node being created. By using the shared AKT reference ontology, we can also link to further information about resources held in other knowledge bases, e.g. when a person is referenced we link to

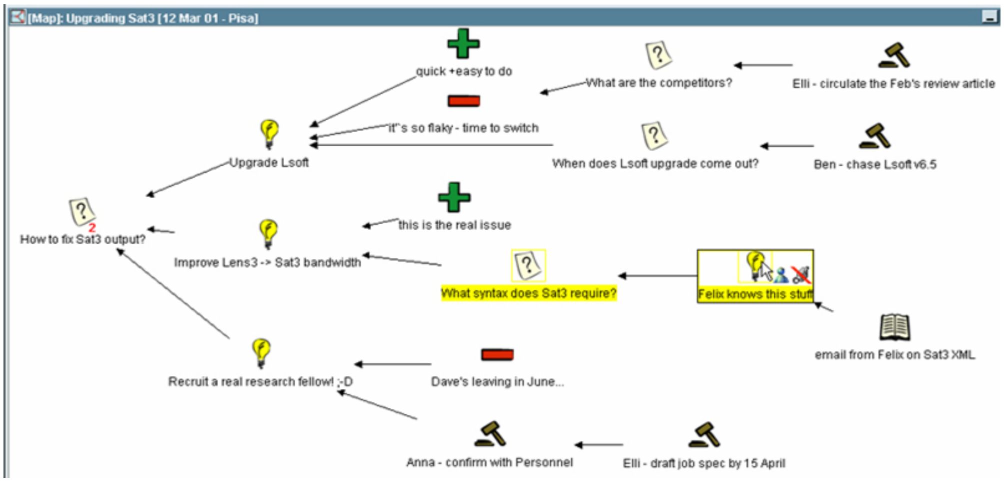

Figure 3: A Compendium map showing various node types and links

information about them in the populated AKT triple store. We populate the timeline with any temporally annotated information about the meeting that would aid the user in navigation.

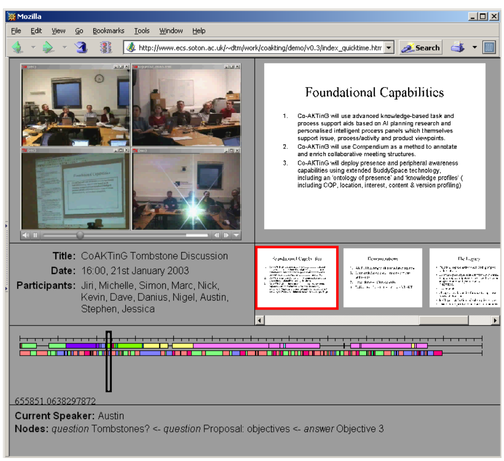

Figure 4: The meeting replay tool

5 Deep Integration

The BuddySpace systems can be adapted to show and track the interactions between the staff and equipment using the National Crystallographic Service (NCS), providing information to their users about the state of the service. Compendium provides the harness to ensure more adequate capture of the discussions in analysis, while Process Panels provide the means to initiate and track key tasks and issues. Additionally the ideas from CoAKTinG provide different techniques to achieve the necessary multi-user interaction in real time over the network and give Combechem the opportunity to implement the "video interaction" collaboration part of Combechem using our event based ontologies to annotate real time streaming media and content.

These various components are valuable complements to Combechem individually but jointly are even more powerful. For example, Process Panels can exploit the presence information derived from Buddy Space with respect to instrument status and operator availability to offer more informed task delegation options. This completes the chain of digital support and capture, maximising the potential for re-use of the digital information in support of the scientific process.

6 Process panels

Here we focus on one particular aspect of the deep integration the application of the Process Panel tool to the laboratory, building on the process capture work of Combechem's Smart Tea team.

Figure 5 shows a screen capture of an IX Process Panel and its Map Tool resulting from our initial experiment. The Map Tool depicts a real Chemistry lab where both fixed and mobile entities are represented. The positions of mobile entities such as movable equipment and technicians are updated automatically through the (World) State sub-panel. By sharing information with BuddySpace, (dynamic) properties of devices are also described in the same panel.

The panel describes a process illustrating steps for composing a chemical compound called CB0110BF4. This process contains several smaller steps and has been described as sub-processes in the I-X process panel. At this particular point in time, it shows Technician-2 is in front of the Rotary Evaporator and about to carry out the sub-process "Re-

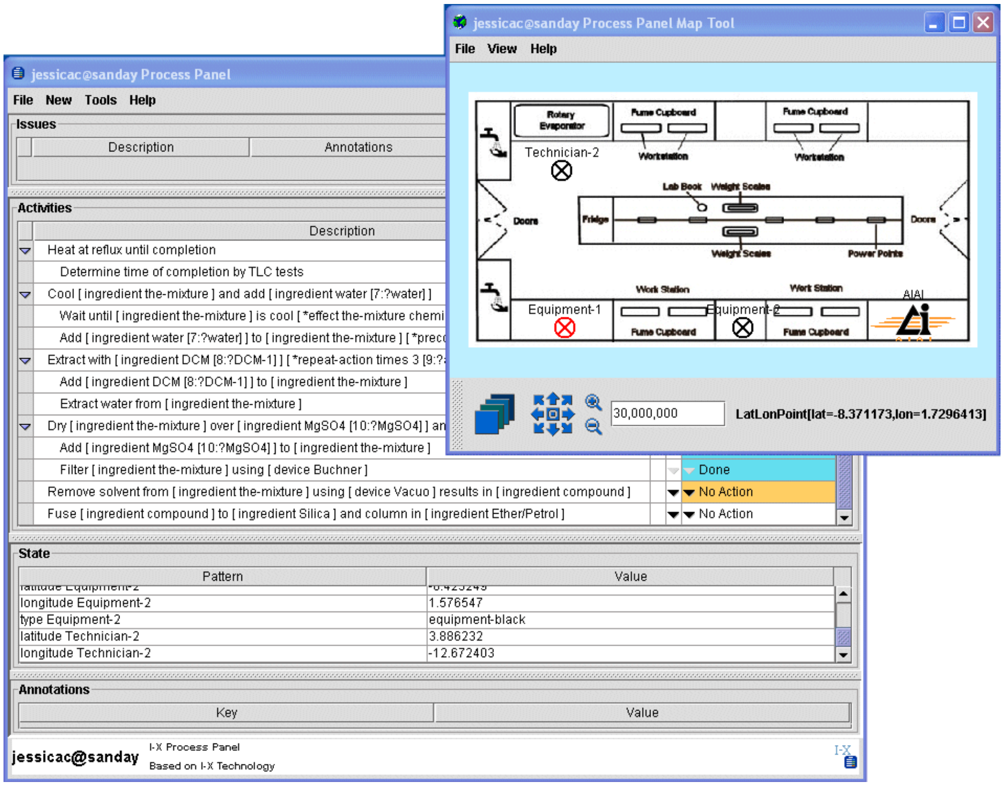

Figure 5: I-X Process Panel configured for e-Chemists

move solvent from the-mixture using Vacuo results in Compound", having completed the previous steps in this process. In additional to its text description, such as "Done", "No Action", the status of a process is indicated by colour code. For instance, the next active (sub-)processes are indicated by an orange-yellow highlight, where as finished processes are blue colour coded, as shown in the figure. Statuses of processes are shared between participants in a collaborative task to help them get instant update of current state of affairs and improve synchronisation between them.

In our investigation, the process decomposition facility of the IX Activity sub-panel supports views of different levels of abstraction that fits nicely with the different practice of chemists and labs. For the same chemical process, chemists of different level of expertises may wish to view an elaborated or concise version of it. Similarly, different chemical labs may employ different procedures for the same process. I-X allows a process to be described in different levels of elaboration, but does not restrict the user to fix on a certain view. The user may wish to explore details of any part of the process and execute them according, but only view the coarse grain description of the rest. I-X also allows alternative (sub-)processes to be associated with a give task enabling flexible practice according to requirements. Different (lab) practices therefore may be recorded in I-X at the same time for a given task. The actual decision as which procedures are chosen for deployment is made at run time.

I-X primitives, such as activities, issues, annotations and constraints may be recorded directly or via Compendium where in-depth discussion has taken place. Static and dynamic process editing provide great flexibility as processes are modifiable at run-time in response to unexpected changes. The ability to store, retrieve and refine process models is important in the Chemistry domain where existing processes are constantly reviewed and modified to discover or synthesise new chemical compounds. This facility alone makes I-X a valuable back-end component for integration with the existing Combechem Grid.

7 Conclusions

Grid computing is fundamentally about joining computational and data resources together. Semantic Web is also about joining things up. In this paper we have demonstrated a set of tools which help join people together too. Furthermore we have made the case for digital capture of the rich information flows in the scientific process, including the discourse between scientists, in a way which supports both the anticipated and unanticipated reuse of this information.

Acknowledgments

This work is funded by the UK Engineering and Physical Sciences Research Council under grant number GR/R85143/01, in association with the Advanced Knowledge Technologies Interdisciplinary Research Collaboration. The AKT IRC research partners and sponsors are authorized to reproduce and distribute reprints and on-line copies for their purposes notwithstanding any copyright annotation hereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of other parties. We are grateful to other members of the Combechem team for their assistance with the case study: Gareth Hughes, Hugo Mills, monica schraefel and Graham Smith. Combechem is also funded by the EPSRC under grant number GR/R67729/01.

References

[1] The AKT Reference Ontology. http://www.aktors.org/ontology/.

[2] J. Conklin, A. Selvin, S. Buckingham Shum, and M. Sierhuis. Facilitated Hypertext for Collective Sensemaking: 15 Years on from gIBIS. In *Proceedings The Twelfth ACM Conference on Hypertext and Hypermedia (Hypertext '01)*, pages 123–124, 2001.

[3] J. Conklin and K. C. Burgess Yakemovic. A Process-Oriented Approach to Design Rationale. *Human-Computer Interaction*, 6(3&4):357–391, 1991.

[4] D. De Roure, N. Jennings, and N. Shadbolt. Research Agenda for the Semantic Grid: A Future e-Science Infrastructure. Technical Report UKeS-2002-2, National e-Science Centre, Edinburgh, UK, 2001.

[5] A. Selvin. Supporting Collaborative Analysis and Design with Hypertext Functionality. *Journal of Digital Information*, 1(4), 1999.

[6] A. Selvin and M. Sierhuis. Towards a Framework for Collaborative Modelling and Simulation. In *Workshop on Strategies for Collaborative Modelling and Simulation, CSCW'96: ACM Conference on Computer-Supported Cooperative Work*, Boston, MA, USA, 1996.

[7] A.M. Selvin and S.J. Buckingham Shum. Repurposing Requirements: Improving Collaborative Sensemaking over the Lifecycle. In *Profess'99: International Conference on Product Focused Software Process Improvement, Oulu, Finland*, pages 539–559, June 1999.

[8] F.M. Shipman and R. McCall. Supporting Knowledge-Base Evolution with Incremental Formalization. In *Proc. ACM CHI'94: Human Factors in Computing Systems, Boston, Mass.*, pages 285–291. ACM Press: New York, 1994.

[9] A. Tate. <I-N-C-A>: an Ontology for Mixed-initiative Synthesis Tasks. In *Proceedings of the Workshop on Mixed-Initiative Intelligent Systems (MIIS) at the International Joint Conference on Artificial Intelligence (IJCAI-03)*, Acapulco, Mexico, August 2003.

[10] A. Tate, J. Dalton, and J J. Stader. I-P2 - Intelligent Process Panels to Support Coalition Operations. In *Proceedings of the Second International Conference on Knowledge Systems for Coalition Operations (KSCO-2002)*, Toulouse, France, April 2002.