

New tools to support collaboration and virtual organizations

BY I. FRAME¹, K. F. AUSTEN¹, M. CALLEJA², M. T. DOVE^{1,*},
T. O. H. WHITE¹ AND D. J. WILSON³

¹*Department of Earth Sciences, University of Cambridge, Downing Street,
Cambridge CB2 3EQ, UK*

²*Cambridge eScience Centre, Centre for Mathematical Sciences,
University of Cambridge, Wilberforce Road, Cambridge CB3 0WA, UK*

³*Institut für Geowissenschaften, Universität Frankfurt, Altenhöferallee 1,
60438 Frankfurt am Main, Germany*

We describe the use of new eScience tools to support collaboration, including the use of XML data representations to support shared viewing of the information content of data, metadata tools for documenting data and Web 2.0 social networking tools for documenting ideas and the collaboration process. This latter work has led to the development of the <http://SciSpace.net> Web resource.

Keywords: eScience; eMinerals; NIEeS; virtual organizations; collaboration; social networks

1. Introduction

eScience is often thought of in terms of grid computing or new approaches to data access and delivery, but a third plank is that of enabling collaboration. The term ‘virtual organization’ (VO) has been coined to show the way that technologies enable collaborators to work together in ways that parallel membership of a common institute. In practice, the concept of the VO is flexible. It might, for example, represent a formal entity associated with sharing of resources, with quality-of-service agreements and access control policies, but more generally the VO is an expression of a task-oriented collaboration between members of geographically distinct institutes. It is important to note that the tools to support a geographically dispersed VO should also operate at a local level, adding new value in supporting work between members and leaders of a research team. The eMinerals and NIEeS projects have had a long-standing interest in exploiting and evaluating tools to support the workings of VOs (Dove *et al.* 2004, 2005). Recently, we have seen the emergence of a new class of tools that, in modern parlance, are associated with what is called ‘Web 2.0’; in this brief review, we report on the work we have been doing in this area.

* Author for correspondence (mtd10@cam.ac.uk).

One contribution of 24 to a Discussion Meeting Issue ‘The environmental eScience revolution’.

2. Problems with traditional tools

The main tools commonly used to support collaboration are email, telephone calls and face-to-face meetings. In the present era, when collaborating teams are often geographically dispersed and span different research disciplines, these are woefully inadequate, and at best support only periodic updates of work carried out by individuals. Although email has several positive features, including its ease of use and ability to add attachments, it is at heart a *communication* tool rather than a *collaboration* tool. Most email client tools do not provide tools to organize email dialogues intelligently (sent and received messages, even if related, are held in different mail boxes), and even if I manage to organize my email one way there is no means to ensure that my collaborators do likewise. We also cite problems of being unable to edit sent messages, of there being no easy way to enable new people to view previous dialogues, and that it is too easy to accidentally omit some collaborators in an exchange.

Sharing data or documents also presents problems for some. Without a common data repository, collaborating researchers can email files to each other if they are small enough, with all the associated problems of email, or else they have to resort to a patchwork of individual data-sharing repositories (such as FTP servers; we note, however, that the digital security policies of many institutions mean that the researchers often have to rely on their own private data repositories for sharing files). Moreover, we are all familiar with collaborators sending us data in the form of badly documented output files or spreadsheets that require further emails before we can unlock the information content; in the forefront of the collaboration process is the need to exchange *information*, not just data.

3. New tools for collaboration

Here, we review the tools we have developed to support the emerging needs of collaborating researchers, and which overcome the problems outlined in §2. Moreover, our capacity to generate unprecedented quantities of new data means that the collaboration tools must enable researchers to quickly understand the information produced by their collaborators. The multidimensional collaboration process is represented in [figure 1](#), showing aspects of communication, data sharing and discovery and information delivery. The important new aspects of collaboration, which are shown by the arrows in the diagram, include the following.

- Data and documents are shared using a data grid or shared data repository that is designed to easily expand access for the new collaborators (top arrow in [figure 1](#)). Our grid job submission system ([Walker *et al.* 2009](#)) will automatically generate complete archives of all files (input and output) associated with a given simulation job that can be easily shared within a collaborating team.
- The problems associated with understanding the data and documents prepared by the collaborators can be overcome by using a common format. We have found XML to be particularly useful. It allows the information content to be viewed using information-centric tools such as the *eMinerals* *ccViz* tool; this transforms XML files into easily readable XHTML format, with tables of data

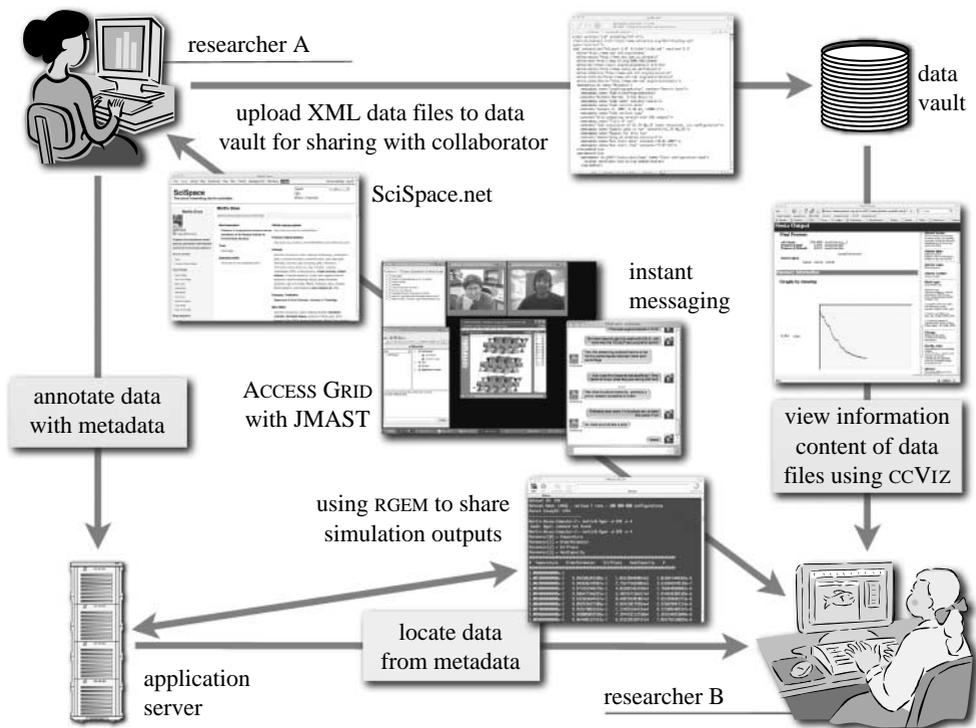


Figure 1. The collaboration process and tools described in §3.

converted to graphs on-the-fly, and all quantities given with dictionary definitions to make the output easily understandable (figure 1, right arrow; White *et al.* 2006, 2009).

- The use of metadata is critical for data discovery, and the job submission system that we have developed allows metadata to be automatically harvested from the job environment and the main output files. Data generated by collaborators can then be discovered from the metadata catalogue (figure 1, left and bottom arrows; Tyler *et al.* 2007).
- By harvesting core output data as metadata in addition to the standard types of data descriptions, we have shown that the metadata tools can provide a useful interface to data, e.g. using our *RGEM* tool researchers can collate output quantities from many jobs run by their collaborators (figure 1, left and bottom arrows; Tyler *et al.* 2007).
- Collaborators can communicate using video conferencing tools (such as the ACCESS GRID with application-sharing tools such as the *eMinerals* JMAST tool) or instant messaging. In fact, our experience has found that instant messaging is the most useful tool for immediate communication.
- Web 2.0 social networking tools (e.g. our *SciSpace.net* tool) enable collaborators to share and document ideas, dialogues, images, dialogues and reports; we discuss this in more detail next.

This set of tools enables researchers to easily share their data and information with their collaborators, and similarly enable their collaborators to locate and understand the information. They effectively bypass many of the time-consuming

problems faced within the collaboration process, such as the physical processes of sharing data and the processes associated with extracting and understanding the information content of data. In the best use cases, it is possible for someone to locate and understand the information generated by a collaborator without having had to be provided with details and without the collaborator having to put in a lot of effort. In such a use case, the communication tools can be put to best use, namely to discuss the scientific advances.

4. Social network tools: **SciSpace.net**

The rise of social networking via the Web over the past 4 years or so is the Internet phenomenon of the current era. Social networking sites integrate many tools that are very useful for collaboration, including blogs (with comments for dialogue), wikis, comment walls, document and image storage, RSS/atom feeds and related tools for aggregation of information feeds (both internal and external), shared bookmarks, networks and communities, user and community profiles, messages, tagging, person/community discovery, knowledge/expertise discovery. Unfortunately it is not practical to use many of the popular social network sites such as Facebook, or even Nature Network (a social networking site for scientists) for scientific collaboration. In particular, we note that most of the social network sites are geared towards publication in its widest sense, whether as a vehicle for publishing people's thoughts through the blog tools, personal information through profiles or specific tools (such as for marking cities you have visited on a map), or information about your day-to-day activities. Collaboration requires a diametrically opposite situation, namely one in which privacy is all-important and that all content created or uploaded must be subject to very tight access controls. We also note the parallel development of eLearning Web tools with similar functionality to social networking sites but tuned specifically for education. Here again, confidentiality is important, but implementation is targeted to a situation in which participants have very different roles from those in a research collaboration. So, although eLearning Web tools cannot easily be used out-of-the-box to support collaboration, they do suggest that the underpinning technologies can be tuned for that purpose.

We have created <http://SciSpace.net> (figure 2) as a prototype Web-based collaborative tool for researchers, building on the technologies of social networking and eLearning tools and including the features they provide (listed above). SciSpace.net is built on the open-source ELGG software (<http://elgg.org/>), and is free for any researcher to join. We note, however, that to obtain maximum benefit people need to join with their collaborators; although a tour through SciSpace.net might be interesting, much of its value will be hidden from the sight of the casual observer (which is exactly how it should be). The software underpinning SciSpace.net provides the fine-scale access control over the content required by collaborating researchers. For example, a user can identify other users as 'friends' and can form access control groups that include named friends. Access to any item created or uploaded by a user can be controlled separately. Thus, as an illustration, a user can post blog entries that can be read and commented on only by other members of the research team, or by a group of collaborators in other universities. Conversely, the same user may want other content to be seen by any

The screenshot shows the SciSpace website interface. At the top, there is a navigation bar with links like 'Help', 'Home', 'Activity', 'Blog', 'Bookmarks', 'Files', 'Wiki', 'Friends', 'Messages (20)', 'RSS feeds', and 'Profile'. On the right, there are links for 'Account settings' and 'Log off'. The main header features the SciSpace logo and the tagline 'The social networking site for scientists'. A search bar is located on the right side of the header.

The main content area is titled 'Negative thermal expansion'. It includes a sidebar on the left with a community profile picture, a description, and links for 'Tags', 'RSS Feeds', and 'Send Message'. The main content area is divided into several sections:

- Brief description:** 'Negative Thermal Expansion: collaboration on experiments and simulations'.
- Interests:** 'ab initio, atomistic simulation, CASTEP, computer simulation, Negative thermal expansion, neutron scattering, neutron total scattering, reverse monte carlo, rigid unit modes, RMC, SIESTA, total scattering, ZrW208'.
- Extended profile:** 'Click here to view extended profile'.
- Introduction:** 'This is a community for work on Negative Thermal Expansion. It has a request-approval membership to enable it to be used for discussion on work in progress.'
- Weblog:** A post titled '[Restricted] Simulations of high-pressure phase' by Martin Dove, including a code snippet for a CASTEP file:


```
%BLOCK LATTICE_ABC
6.625 6.625 6.622
95.829 84.171 60.345
%ENDBLOCK LATTICE_ABC

%BLOCK POSITIONS_FRAC
Ag .50000 .50000 .50000
Ag .23662 .23662 .50000
```
- What are we doing now?:** 'Writing papers and trying to understand the differences between Ag₃Co(CN)₆ and H₃Co(CN)₆'.
- Comment Wall:** A comment from Martin Dove at 9:10 am on February 9 says...

Figure 2. Example home page for a collaborative group on SciSpace.net.

other member of SciSpace.net or indeed by the whole world. Within SciSpace.net, it is also possible and useful to form ‘communities’ of users, which have access to the same tools and access controls as individuals. Membership of such communities can be tightly restricted or completely open. For example, the core team that is creating SciSpace.net has its own private community, in which progress, problems and new ideas are documented. The authors have established a number of private communities associated with their own scientific studies. At the same time, communities can have content publicly displayed where this meets other needs; one example is an open community devoted to publicizing advances in the use of Web technologies for geospatial data representation.

We now have a number of use cases for SciSpace.net, which will be documented in detail elsewhere. These include collaborators working together within a secure environment to share data, develop thoughts and discuss ideas, with the research process and progression being fully documented in a way that is easy for new collaborators to understand. In several cases, these have led to scientific papers (e.g. Calleja *et al.* 2008) and research proposals. Another use case is that of researchers using SciSpace.net as a virtual logbook to document their research processes and results, something that is particularly and uniquely useful for research supervisors.

We acknowledge support from NERC under the eScience thematic programme.

References

- Calleja, M., Goodwin, A. L. & Dove, M. T. 2008 Origin of the colossal positive and negative thermal expansion in $\text{Ag}_3[\text{Co}(\text{CN})_6]$: an ab initio density functional theory study. *J. Phys. Condens. Matter* **20**, 255–226. (doi:10.1088/0953-8984/20/25/255226)
- Dove, M. T. *et al.* 2004 Collaborative tools in support of the eMinerals virtual organisation. In *Proc. UK e-Science All Hands Meeting, 2004*, pp. 127–134. See <http://www.allhands.org.uk/archive/index.cfm>.
- Dove, M. T. *et al.* 2005 The eMinerals collaboratory: tools and experience. *Mol. Simul.* **31**, 329–337. (doi:10.1080/08927020500066163)
- Tyer, R. P. *et al.* 2007 Metadata management and grid computing within the eMinerals project. In *Proc. UK e-Science All Hands Meeting, 2007*, pp. 415–422. See <http://www.allhands.org.uk/archive/index.cfm>.
- Walker, A. M., Bruin, R. P., Dove, M. T., White, T. O. H., van Dam, K. K. & Tyer, R. P. 2009 Integrating computing, data and collaboration grids: the RMCS tool. *Phil. Trans. R. Soc. A* **367**, 1047–1050. (doi:10.1098/rsta.2008.0159)
- White, T. O. H., Murray-Rust, P., Couch, P. A., Tyer, R. P., Bruin, R. P., Todorov, I. T., Wilson, D. J., Dove, M. T. & Austen, K. F. 2006 Application and uses of CML within the eMinerals project. In *Proc. UK e-Science All Hands Meeting, 2006*, pp. 606–613. See <http://www.allhands.org.uk/archive/index.cfm>.
- White, T. O. H., Bruin, R. P., Chiang, G.-T. & Dove, M. T. 2009 Lessons in scientific data interoperability: XML and the eMinerals project. *Phil. Trans. R. Soc. A* **367**, 1041–1046. (doi:10.1098/rsta.2008.0175)