

# Case for Support: SOCA

Sister Project: SOuthampton Chicago Activity

*Luc Moreau, Michael Luck, Nick Jennings, David de Roure*

University of Southampton

*Ian Foster, Mike Wilde, Katarzyna Keahey*

Argonne National Laboratory/University of Chicago

## Part 1: Previous Research

**Southampton** The proposed research will be conducted in the *Intelligence, Agents, Multimedia* group (IAM) in the School of Electronics and Computer Science at the University of Southampton. The School was rated 5\* both in Computer Science and Electronics in the RAE 2001. IAM is the largest Computer Science research group in the School (with around 120 research staff), with an international reputation in the area of distributed and Grid computing, agent-based computing, distributed information management and knowledge technologies.

The Southampton investigators are involved in a number of national and international Grid/e-Science collaborations: myGrid is an EPSRC-funded pilot project that is targeted at developing open source high-level middleware to support personalised in silico experiments in biology on a Grid; it also involves U. of Manchester (lead), U. of Newcastle, U. Of Nottingham, U. Of Sheffield and the European Bioinformatics Institute. Comb-e-Chem is another EPSRC-funded pilot project that is working on Grid-enabled combinatorial chemistry, concentrating on crystallography and laser and surface chemistry. It involves Bristol University and several departments in Southampton (statistics, chemistry and ECS). Both PASOA (an EPSRC-funded e-Science project) and EU-PROVENANCE (EU funded) focus on “provenance” in the context of the Grid. PASOA is in collaboration with Cardiff, whereas EU-PROVENANCE involves IBM, Cardiff, German Aerospace, Budapest and Barcelona.

The School of Electronics and Computer Science also hosts the OMII, the Open Middleware Infrastructure Institute, whose mission is to provide Grid software distributions of choice for application scientists and industry wishing to either establish a grid infrastructure and/or develop applications and services for a grid infrastructure.

**Chicago** The proposed research will be conducted in the Distributed Systems Laboratory, a research and software development group within the Mathematics and Computer Science (MCS) Division at Argonne National Laboratory and the Department of Computer Science at the University of Chicago. Under the leadership of Dr. Ian Foster, the DSL hosts research and development activities designed to realize the potential of “Grids” for computational science and engineering. Collaboration with Carl Kesselman’s Center for Grid Technologies at the University of Southern California Information Sciences Institute set the foundation of the Globus Project, a highly-collaborative international and multidisciplinary effort to make Grid computing a reality. The Globus Alliance is now an international collaboration involving Argonne National Laboratory, U. of Chicago, U. of Southern Califor-

nia Information Sciences Institute, U. of Edinburgh and the Swedish Royal Institute of Technology.

The GriPhyN and the International Virtual Data Grid Laboratory (iVDGL) projects are large-scale collaborations of computer scientists, physicists and astronomers conducting Grid research applied to several national and international experiments with massive computational and data needs. Over twenty institutions are involved in these collaborations.

The Distributed Systems Laboratory hosts a series of projects, including the National Fusion Collaboratory, relevant to this proposal, which aims at achieving a reference implementation of the WS-agreement protocol, in order to negotiate and monitor resource allocation agreement.

**Previous Joint Activities** The project investigators have known each other for a number of years. Collaboration started by the design of a binding of the Nexus library for the Scheme language [15]. Recently, Luc Moreau, who was on sabbatical at Argonne National Laboratory and University of Chicago with the GriPhyN team, investigated some research issues in workflow-based management of data sets, which resulted in a joint research agenda, reflected in part in this proposal. Finally, Ian Foster, Nick Jennings and Carl Kesselman define a research agenda to apply agent-based computing to grids [8], and this proposal aims to develop some of its aspects.

**Researchers** Luc Moreau is a Professor of Computer Science in the school of Electronics and Computer Science at the University of Southampton. He has published over 80 articles in the domain of Grid computing, distributed systems, agent-based systems, and distributed information management. He is the principal investigator and co-investigator of a number of Grid and e-Science projects including PASOA, EU-PROVENANCE, GRIMOIRES, myGrid and Comb-e-chem. Luc Moreau spent time as a visiting researcher with the GRIPHYN team at the University of Chicago and Argonne National Laboratory, at the Institute for Human and Machine Interaction, Pensacola, Florida, at École Normale Supérieure, Lyon, France, at the University of Linz, Austria, at the University of Edinburgh, and at INRIA-Rocquencourt. He serves on the program committees of multiple international conferences and workshops on grid and distributed computing, multi-agent systems and peer-to-peer computing. He is programme co-chair for CCGrid’05.

Nick Jennings is Professor of Computer Science in the School of Electronics and Computer Science where he carries out basic and applied research in agent-based computing. Professor Jennings helped pioneer the use of agent-based techniques for real-world applications; developing systems in the domains of: process control, business process management, e-commerce, telecommunications net-

work management, virtual laboratories, and scientific data interpretation. He has also been instrumental in shaping the research agenda for the Semantic Grid. He has also made foundational contributions in the areas of automated negotiation and auctions, and cooperative problem solving. He has published over 200 articles and 6 books on various facets of agent-based computing and holds 2 patents (3 more pending). He has received a number of awards for his research: the Computers and Thought Award in 1999, an IEE Achievement Medal in 2000, and the ACM Autonomous Agents Research Award in 2003. He is a Fellow of the British Computer Society, the Institution of Electrical Engineers, and the European artificial intelligence association (ECCAI) and a member of the UK Computing Research Committee (UKCRC).

Michael Luck is a Professor of Computer Science in the Department of Electronics and Computer Science at the University of Southampton, UK, and has worked in the field of agent technology and multi-agent systems for over ten years. Prof. Luck is a co-founder of the UK Special Interest Group on Multi-Agent Systems, and of the UK and European Workshops on Multi-Agent Systems. He is currently Chair of the UKMAS Steering Committee, and a member of the Advisory Boards of FIPA (the agent standards body), MAAMAW (the European agent conference) and CEEMAS (the Central and Eastern European agent conference). He is principal or co-investigator on several e-Science and Grid-related projects, including CONOISE-G, CONOISE, myGrid and Comb-e-chem. From 2000 to 2003, he was Director of AgentLink II, the European Network of Excellence for Agent-Based Computing, and since 2004 has been Executive Director of AgentLink III. He has served on numerous organising and programme committees for agent conferences and workshops, has contributed to policy making forums for national and European agencies, and has published extensively in this area (with over 100 articles and 8 books).

Professor David De Roure heads Grid and Pervasive Computing in the School of Electronics and Computer Science. He was a founding member of the School's Intelligence, Agents, Multimedia Group and his current research focus is the application of advanced knowledge technologies to both Grid and pervasive computing. He is involved in the UK e-Science programme as Co-Director of the Southampton Regional e-Science Centre, a grantholder of the Open Middleware Infrastructure Institute and a member of the Architecture Task Force, as well as leading the CoAKTinG project and the CS team in the Comb-e-Chem pilot. He is Director of the DTI-sponsored Centre for Pervasive Computing in the Environment. Internationally, he is chair of the Semantic Grid Research Group in the Global Grid Forum and a member of the GGF Steering Group, has participated in the EU Next Generation Grids experts' group, and is leading the Grid activity in the Worldwide Universities Network. He obtained his PhD in 1990 in the design of distributed programming languages and subsequently conducted research in both distributed computation and distributed information systems. He is a WorldWide Web Consortium Advisory Committee member and actively involved in the international World

Wide Web conference series. He sits on national committees for JISC and the Arts and Humanities Data Service.

Ian Foster is Associate Director of the Mathematics and Computer Science Division of Argonne National Laboratory and the Arthur Holly Compton Professor of Computer Science at the University of Chicago. His research interests are in distributed and parallel computing and computational science, and he has published six books and over 200 articles and technical reports on these and related topics. He is an internationally recognized researcher and leader in Grid computing, and leads projects applying Grid technologies to scientific and engineering problems, in such fields as high energy physics, climate data analysis, and earthquake engineering. Foster is a fellow of the American Association for the Advancement of Science and the British Computer Society. His awards include the British Computer Society's award for technical innovation, the Global Information Infrastructure (GII) Next Generation award, the British Computer Society's Lovelace Medal, and Research and Development Magazine's Innovator of the Year.

Mike Wilde is a member of the Mathematics and Computer Science division of Argonne National Laboratory, where he serves as Software Architect in the Distributed Systems Laboratory, and of the Computation Institute of the University of Chicago. Since 2001 he has been the coordinator of GriPhyN - the Grid Physics Network - an NSF-funded project that brings together four large-scale scientific experiments with a large national team of computer science researchers, to solve problems in data intensive computing on the Grid. GriPhyN has developed and applied a widely-used set of tools which use the concept of "virtual data" to manage large-scale scientific workflows and to track data provenance. Wilde previously held several consulting positions within AT&T Bell Laboratories, where his research and development efforts included mainframe multiprocessor UNIX systems, network communication processors, and systems integration languages and environments. He also served as Chief Technology Officer for coolsavings.com, a successful Chicago-based internet marketing company, and managed the development of mass storage management technologies.

Katarzyna Keahey is an Assistant Computer Scientist in the Distributed Systems Laboratory in the Mathematics and Computer Science Division at Argonne National Laboratory and a fellow of the Computation Institute at the University of Chicago. She has worked on parallel extensions to the CORBA standard and component technologies and became a founding member of the Common Component Architecture (CCA) forum. She initiated her involvement with Grid computing nine years ago by winning an award for the best integration of heterogeneous applications during the I-WAY experiment. Her current interests focus on distributed and Grid computing, virtualization, and policy-driven resource management. She actively participates in the Global Grid Forum (GGF) where she has chaired the Grid Economic Services Architecture (GESA) working group and co-authored the WS-Agreement specification allowing a client to dynamically negotiate agreements in a Grid environment.

## Part 2: Proposed Research and its Context

### A. Background

In their respective activities, researchers in Southampton and Chicago investigate a set of similar issues, pertaining to the fundamentals of Grids in the context of e-Science. In this section, we overview some of the projects and issues that are being investigated by each team, and then we present the case for a joint sister project that aims to bridge these activities.

**Southampton Background** In the UK, myGrid ([www.mygrid.org.uk](http://www.mygrid.org.uk)) is an EPSRC e-Science pilot aiming at designing and building high level services to form personalised in silico experiments in biology [21]. Within this context, the Southampton team has designed a *registry* capable of hosting descriptions of services and workflows [13]; in addition, the myGrid registry supports third-party attachment of metadata about published services and workflows [14]. It allows publishers but also third-party users to publish information about services, such as functionality, semantic description, quality of service and trust metrics. As part of the OMII-funded GRIMOIRES project, the myGrid registry is going to be refactored and made secure, so as to become the registry of choice of the OMII distribution.

In the myGrid project, Southampton has developed a notification service [10] that mediates between information publishers and consumers that wish to subscribe to periodic updates. Since there is commonly a mismatch between the dissemination of these updates and the delivery preferences of the consumer (in terms of frequency of delivery, quality, etc), an automated *negotiation engine* was designed to identify quality of service that is mutually acceptable between the different parties [11]. A novelty of the proposed approach is that it attempts to reach negotiated solutions when multiple intermediaries are involved in the search for quality of service [12].

Two new key projects, PASOA (EPSRC-funded project) and EU-PROVENANCE (STREP funded by the sixth European framework) aim at the design, implementation and deployment of a provenance architecture for grids. The former project tends to focus on the fundamental aspect of provenance and its algorithmic solutions [20, 7], whereas the latter will investigate an industrial-strength solution, which must offer security and scalability. PASOA adopts a bioinformatics use case to drive the research, whereas EU-PROVENANCE concentrates on grid applications in the medical and aerospace domains.

In the CONOISE project, Southampton investigated different ways in which agents can negotiate to form virtual organisations. In CONOISE-G, Southampton extended this to consider the interactions between different agents in forming virtual organisations in a Grid context to address issues of trust and reputation. Most previous work assumed complete trust, avoiding the issue, but with new efforts in e-Science and e-Business introducing large-scale distributed interactions among large numbers of agents, the problems of trust take on new significance. Questions of deception and fraud in communication and interaction,

of assurance and reputation, and of risk and confidence, are particularly significant, especially where interactions take place with new partners. CONOISE-G Southampton is thus addressing the issues involved in modelling and using trust and reputation in support of effective and assured interaction in large-scale agent-based open computing environments like the Grid.

The EPSRC project “Virtual Organisations for e-Science” seeks to carry out the fundamental computer science research that is necessary to support the entire virtual organisation (VO) lifecycle as it exists for e-Science and the Grid. Specifically, this involves research into how services can be composed in order to achieve a particular objective, how to recruit an appropriate set of agents to meet service specifications, how to assign individual duties to the constituent agents, how to monitor the performance of the VO, and how to reconfigure the VO in face of changing circumstances.

**Chicago Background** The GriPhyN project ([www.griphyn.org](http://www.griphyn.org)) is developing Grid technologies for scientific and engineering projects that must collect and analyze distributed, petabyte-scale datasets. GriPhyN research enables the development of Petascale Virtual Data Grids [4] through its Virtual Data Toolkit. It is observed that much scientific data is not obtained from measurements but rather derived from other data by the application of computational procedures. The explicit representation of these procedures can enable documentation of data provenance, discovery of available methods, and on-demand data generation (so-called “virtual data”). To explore this idea, the Chimera Virtual Data System [3] combines a *virtual data catalog*, for representing data derivation procedures and derived data, with a virtual data language interpreter that translates user requests into data definition and query operations on the database.

During his sabbatical stay at the University of Chicago and Argonne National Laboratory (April to July 2004), Luc Moreau worked in close collaboration with Ian Foster, Mike Wilde and the GriPhyN team on the problem of accessing diverse data stored in many different formats. This work resulted in the design of XDGM, the *XML Dataset Type and Mapping* system that provides a separation of concern between the logical structure of datasets and its mapping to a physical representation [17]. XDGM allows a workflow language to specify operations over complex datasets, independently of their physical representation.

The iVDGL ([www.ivdgl.org](http://www.ivdgl.org)) is a global Data Grid that serves forefront experiments in physics and astronomy. Its computing, storage and networking resources in the U.S., Europe, Asia and South America provide a unique laboratory to test and validate Grid technologies at international and global scales. Within this context, Grid3 is an operational Grid, involving 28 sites with about 2500 CPUs, with 400-1300 concurrent jobs running on average, and 7 substantial applications. Each application is associated with a virtual organisation (VO), which specifies acceptable use policies, allocation of resources, membership and available services. Grid3 is evolving to the Open Science Grid,

to be launched in Feb 2005 as the first project of the Open Science Grid consortium.

The National Fusion Collaboratory [19] focuses on enabling fusion scientists to explore Grid capabilities in support of experimental science. Fusion experiments are structured as a series of plasma pulses initiated roughly every 20 minutes. In the between-pulse intervals, scientists perform data analysis and discuss results to reach decisions affecting changes to the next plasma pulse. This interaction can be made more efficient by performing more analysis and engaging more expertise from a geographically distributed team of scientists and resources. In order to ensure that fusion applications meet the between-pulse deadline, it is necessary to develop not only resource management strategies but also protocols enabling structured negotiation for the use of resources. Such protocols allow the creation of policies dynamically and automatically, broadening the scope of Grid applications. Fusion developed such prototype system [9] and contributed its experiences to the WS-Agreement specification developed at GGF [1]. So far, the focus of the project has been on defining the terms of an agreement for grid resources and on monitoring such an agreement. Negotiation of such an agreement has only preliminarily been investigated, and solutions are required that offer more optimal results.

**Common set of issues** This overview of Southampton and Chicago projects has identified several problems that are investigated by both teams in parallel, and would benefit from a structured collaboration. We discuss such problems here.

The virtual data catalog of the Chimera system combines the functionality of several services investigated by Southampton researchers: the *registry* for publishing services, workflow and legacy programs and the *provenance* service for recording provenance data and reasoning over it. Use cases should be shared between teams in order to identify common requirements and design common interfaces. As a result services could be shared between projects.

The Southampton negotiation engine is a natural complement to the architecture of the National Fusion Collaboratory project, which assumes the existence of a negotiated agreement, monitors and enforces it. In order for these systems to interact usefully, common interfaces and data model need to be identified.

Following the work initiated in [17], further work is required to specify a workflow language that takes advantage of such a new infrastructure. It is expected that an enactor for such a workflow language will also have to contribute provenance information about the processes that it executes, and therefore will have to interact with the provenance service and virtual data catalog investigated by each team.

Currently, virtual organisations in Grid3 are mostly set up manually, though the infrastructure provides some support such as VO-based discovery or membership services. More automation is required in order to support the autonomous VO formation. To do so, it is necessary to determine policy (negotiation, trust management),

to determine membership and role (coalition formation), to create VO wide services (global behaviors), to managing work (collaborative problem solving, workflow management) and to deal with evolution. Some of these issues are being investigated by the Southampton CONOISE-G and “Virtual Organisations for e-Science”.

## B. Programme and Methodology

By a close collaboration between Southampton and Chicago researchers, the aim of the SOCA project is to investigate some fundamental issues in the design, development and deployment of grids for e-Science, through the sharing of uses cases, the design of common systems and associated interfaces, and the practical integration of services in deployed grids. We aim to demonstrate the integration of registry and provenance services in the Chimera system, through their use in scientific applications. Specifically, the scientific objectives of the project are to study the following fundamental issues:

1. *Provenance*: understand the principles of provenance recording in the context of workflow-based grid systems, specify common interfaces, and integrate components.
2. *Registry*: adopt common models for describing services, workflows and legacy programs and specify common interfaces for registering and discovering them, and integrate components.
3. *Workflow*: pursue the design of the second generation virtual data language, which focuses on a capability to manipulate data sets in an abstract manner.
4. *Negotiation*: investigate how agent-based negotiation protocols can be integrated into the WS-agreement framework.
5. *Virtual Organisations*: analyse existing operational virtual organisations such as the ones supported by Grid3, and compare with methodologies and algorithms conceived in the context of the Virtual Organisations for e-Science and CONOISE-G projects.

The SOCA project will build bridges between the US and UK communities, since both Southampton and Chicago are actively involved in numerous collaborations in their respective countries. In addition, SOCA will establish links between several communities. The themes of virtual organisations and negotiations are actively researched by the agent community, and this project will seek to bring agents techniques to the grid community and vice-versa, as already pioneered and advocated by some of the investigators [18, 16, 8].

The five fundamental issues to be studied are all aspects of the Semantic Grid research agenda. This was first defined by De Roure, Jennings and Shadbolt in 2001 [2] and emphasises the role of machine processable service and data descriptions within the grid middleware. Semantic Grid research has attracted increasing attention through the e-Science programme, particularly in the myGrid and CombeChem projects, as well as influencing international

research projects, and is the subject of ongoing development through the Semantic Grid Research Group in the Global Grid Forum.

Funding the SOCA project is important since it will help structure a fruitful cooperation between the Southampton and Chicago teams. First, both teams have complementary expertise in a number of common domains and are keen to exploit such a complementarity in the context a structured cooperation. Second, the proposed project will build up on previous work already undertaken by the investigators (Luc Moreau's sabbatical at University of Chicago/Argonne National Laboratory and Ian Foster and Nicholas Jennings' paper on agents and the grid), who have identified a common agenda of future research that is reflected in this proposal. Third, both Chicago and Southampton are involved in multiple national collaborations in their respective countries, which would facilitate the dissemination of results.

**Detail** As far as provenance is concerned, our work will investigate the mechanisms by which granularity of provenance can be specified by users, and how such levels of provenance can be recorded. Our investigation will consider the relationship of such a notion of provenance with the notions of "where" and "why" provenance introduced by P. Bunemann *et al.* for databases. Finally, we will study the semantics of provenance with respect to workflow languages. Preliminary work regarding these issues has already been undertaken during Luc Moreau's sabbatical in Chicago [17].

The challenge in integrating the myGrid/GRIMOIRES registry with the virtual data catalog is that they use different underlying models for describing services. The myGrid/GRIMOIRES registry adopts the UDDI service description as its underlying representation, but also supports myGrid-specific profiles that contain semantic descriptions of services. On the other hand, the virtual data catalog uses its own metadata schema, and adopts its own representation for describing legacy programs. By designing a model that can support these different representations, we will obtain a very powerful representation technique, which will be applicable to many application domains.

While preliminary work [17] has identified a methodology for designing a workflow language that is capable of manipulating data sets in an abstract manner, independently of their physical representation, a number of issues still need to be addressed in order to make such a language practical and grid-executable. For instance, (i) one needs to investigate how both Web/Grid services and legacy applications can be invoked; (ii) one needs to identify iteration constructs that are capable of performing repetitive behaviour over elements of a set, but that can be executed over a Grid efficiently; (iii) one needs to study the relationship between workflow variables used to hold (intermediate) results and logical file descriptors identifying where such results are stored in a file system.

Negotiation can generally be seen as a form of search for an optimal solution in a multi-dimensional space. In the context of grid computing, it is required to understand

what an optimal solution is and what impact on performance such a search may have. It is also important to be able to decide which negotiation protocol may be more desirable for the prevailing circumstances. In order to make such negotiation protocols usable in practice, they need to be integrated with the WS-agreement specification.

Work on virtual organisations has largely adopted different perspectives. On the one hand, it is a natural consequence of the agent approach to computing, where different agents come together to achieve some overarching objectives, and this is a stance that has been studied by Southampton. On the other hand, the starting point from Grid computing is influenced by underlying infrastructural concerns. At the same time, there are still open questions of what constitutes a VO, when to form a VO, who should be in it, and so on. We will aim to integrate the two views, drawing on experience from the different base projects, to provide a coherent set of answers to these questions.

**Programme** In order to achieve those objectives, we have defined a work programme articulated around a set of meetings. Part III presents a tentative schedule for the purpose of illustration; the schedule will be adapted according to the dates of external events in order to limit the cost of meetings, but also according to the priorities of the different projects in order to maximise research outputs.

The core of the collaboration will take place in three workshops where many researchers of the Southampton and Chicago teams will meet to discuss issues pre-defined in well-defined agendas. Such workshops will be composed of plenary and break-out sessions: the former will consist of technical presentations of relevant projects and reporting of different working groups, whereas the break-out sessions will consist of small working groups investigating specific research objectives of the SOCA project.

Each major workshop will be prepared at a face-to-face meeting co-located with another event in the Grid/e-Science community. For the purpose of project planning, we have used the Global Grid Forum, regularly attended by many of the project participants, as such an external event. Alternate major events include UK e-Science meetings or conferences such as CCGRID and HPDC.

We expect the working groups to sketch solutions, identify directions of investigation, or outline report structures. Concrete project deliverables will be realised through research exchanges: in total, we expect 3 researchers of each team to spend a week at the sister location.

Finally, monthly access grid meetings or teleconferences will be used to track progress of the different activities.

**Project deliverables** Both teams are familiar with twiki, which we propose to use as a collaborative tool to make the following project deliverables available:

1. Meetings reports.
2. Common interface specifications (registry, provenance, negotiation).

3. Software demonstrator: registry and provenance services integrated with Chimera system and used in a scientific application.
4. Report on workflow language to support virtual data.
5. Report on virtual organisation requirements in practical grid deployments.

### C. Relevance to Beneficiaries

The proposed work is beneficial for the e-Science and Grid communities in the UK and the US, since it addresses *fundamental issues* for building grids for e-Science applications, namely virtual organisations, negotiation, resource discovery, workflows and provenance. The SOCA project will help researchers to share common use cases, address more complex requirements, and find solutions that are more generic and that can be used in multiple application domains. The design of common interfaces will also promote systems inter-operability and components reuse.

Specifically, EPSRC-funded PASOA, EU-PROVENANCE and OMII GRIMOIRES will benefit from the collaboration since it will foster the development of generic interfaces, and ensure the generality of the services these projects will deliver; as a result, provenance and registry services for example will be reusable with other systems such as Chimera. Likewise, EPSRC-funded “Virtual Organisations for e-Science” project and DTI/Welsh e-Science center CONOISE-G will benefit of practical use cases from deployed virtual organisations. Finally, the collaboration will offer new dissemination routes for existing projects; for instance, the negotiation capability of the myGrid notification service will be reusable in the Argonne National Laboratory project National Fusion Collaboratory.

This effort towards making services more generic and more reusable not only benefits the aforementioned projects directly, but also the whole UK e-Science programme and the e-Science community in general, the industrial partners of these projects, including IBM, Sun Microsystems, BT, GSK, AstraZeneca, Merck KGaA, German Aerospace and the Global Grid Forum, in which the different collaborators are actively involved.

### D. Dissemination and Exploitation

We expect the results of this work to be disseminated by publication at appropriate national and international conferences and in international journals. We shall maintain an up-to-date online archive of technical reports and published papers. The Global Grid Forum, the World Wide Web Consortium, and the new EU exchange forum are also ideal forums to present our results.

### E. Justification of Resources

We seek funds for travel and subsistence to support meetings and exchange of researchers. Specifically, for the 2 year project duration, we will hold 3 workshops, we will organise some researcher exchanges, and setup meetings co-located with other events.

1. **Three workshops:** we will organise three workshops for which travel and subsistence is costed at £8k; in order to limit the cost, such workshops will be located at the University of Southampton, at the University of Chicago, or Argonne National Laboratory.
2. **Six Researcher Exchanges:** during the two year project duration, we budget for visits of six researchers from each participating institution to visit sister institutions for a duration of one week. We budget each exchange at £2.3K.
3. **Meetings co-located with Other Events:** we seek funding to organise face to face meetings peripheral to major events (such as GGF, CCGRID, HPDC), likely to be attended by many project participants. We allocate £2K for such meetings.

Three workshops	24000
Six exchanges	14000
Co-located meetings	2000
<b>Total</b>	<b>40000</b>

## References

- [1] A. Andrieux, K. Czajkowski, A. Dan, K. Keahey, H. Ludwig, Jim Pruyne, J. Rofrano an S. Tuecke, and M. Xu. Web services agreement specification (ws-agreement). Technical report, Global Grid Forum, 2004.
- [2] David de Roure, Nicholas R. Jennings, and Nigel Shadbolt. The Semantic Grid: A future e-Science infrastructure. In Fran Berman, Anthony J.G. Hey, and Geoffrey Fox, editors, *Grid Computing: Making The Global Infrastructure a Reality*, pages 437–470. John Wiley & Sons, April 2003.
- [3] I. Foster, J. Voeckler, M. Wilde, and Y. Zhao. Chimera: A virtual data system for representing, querying and automating data derivation. In *Proceedings of the 14th Conference on Scientific and Statistical Database Management*, Edinburgh, Scotland, July 2002.
- [4] I. Foster, J. Voeckler, M. Wilde, and Y. Zhao. The virtual data grid: a new model and architecture for data-intensive collaboration. In *Proceedings of the Conference on Innovative Data Systems Research*, Asilomar, CA, January 2003.
- [5] Ian Foster, Carl Kesselman, Jeffrey M. Nick, and Steven Tuecke. The Physiology of the Grid — An Open Grid Services Architecture for Distributed Systems Integration. Technical report, Argonne National Laboratory, 2002.
- [6] Ian Foster, Carl Kesselman, and Steve Tuecke. The Anatomy of the Grid. Enabling Scalable Virtual Organizations. *International Journal of Supercomputer Applications*, 2001.
- [7] Paul Groth, Michael Luck, and Luc Moreau. Formalising a protocol for recording provenance in grids. In *Proceedings of the UK OST e-Science second All Hands Meeting 2004 (AHM'04)*, Nottingham, UK, September 2004.
- [8] Nicholas R. Jennings Ian Foster and Carl Kesselman. Brain meets brawn: Why grid and agents need each other. In *Proc. 3rd Int. Conf. on Autonomous Agents and Multi-Agent Systems*, pages 8–15, New York, USA, July 2004.
- [9] Katarzyna Keahey, Takuya Araki, and Peter Lane. Agreement-based interactions for experimental science. In *Tenth International Europar Conference (EURO-PAR'04)*, Pisa, Italy, August 2004.
- [10] Ananth Krishna, Victor Tan, Richard Lawley, Simon Miles, and Luc Moreau. The mygrid notification service. In *Proceedings of the UK OST e-Science second All Hands Meeting 2003 (AHM'03)*, pages 475–482, Nottingham, UK, September 2003.

- [11] Richard Lawley, Keith Decker, Mike Luck, Terry Payne, and Luc Moreau. Automated negotiation between publishers and consumers of grid notifications. *Parallel Processing Letter*, 13(4), 2003.
- [12] Richard Lawley, Michael Luck, and Luc Moreau. Chained negotiation for distributed notification services. In *Proceedings of the UK OST e-Science second All Hands Meeting 2004 (AHM'04)*, Nottingham, UK, September 2004.
- [13] Simon Miles, Juri Papay, Vijay Dialani, Michael Luck, Keith Decker, Terry Payne, and Luc Moreau. Personalised grid service discovery. *IEE Proceedings Software: Special Issue on Performance Engineering*, 150(4):252–256, August 2003.
- [14] Simon Miles, Juri Papay, Terry Payne, Keith Decker, and Luc Moreau. Towards a protocol for the attachment of semantic descriptions to grid services. In *The Second European across Grids Conference*, page 10, Nicosia, Cyprus, January 2004.
- [15] Luc Moreau, David DeRoure, and Ian Foster. NeXeme: a Distributed Scheme Based on Nexus. In *Third International European Conference (EURO-PAR'97)*, volume 1300 of *Lecture Notes in Computer Science*, pages 581–590, Passau, Germany, August 1997. Springer-Verlag.
- [16] Luc Moreau, Simon Miles, Carole Goble, Mark Greenwood, Vijay Dialani, Matthew Addis, Nedim Alpdemir, Rich Cawley, David De Roure, Justin Ferris, Rob Gaizauskas, Kevin Glover, Chris Greenhalgh, Peter Li, Xiaojian Liu, Phillip Lord, Michael Luck, Darren Marvin, Tom Oinn, Norman Paton, Stephen Pettifer, Milena V Radenkovic, Angus Roberts, Alan Robinson, Tom Rodden, Martin Senger, Nick Sharman, Robert Stevens, Brian Warboys, Anil Wipat, and Chris Wroe. On the Use of Agents in a BioInformatics Grid. In Sangsan Lee, Satoshi Sekguchi, Satoshi Matsuoka, and Mitsuhsa Sato, editors, *Proceedings of the Third IEEE/ACM CCGRID'2003 Workshop on Agent Based Cluster and Grid Computing*, pages 653–661, Tokyo, Japan, May 2003.
- [17] Luc Moreau, Yong Zhao, Ian Foster, Jens Voeckler, and Michael Wilde. Specifying and iterating over virtual datasets. Technical report, University of Southampton, 2004.
- [18] Omer F. Rana and Luc Moreau. Issues in Building Agent based Computational Grids. In *Third Workshop of the UK Special Interest Group on Multi-Agent Systems (UKMAS'2000)*, page 11, Oxford, UK, December 2000.
- [19] D.P. Schissel, K. Keahey, T. Araki, J.R. Burruss, E. Feibush, S.M. Flanagan, T.W. Fredian, M.J. Greenwald, S.A. Klasky, T. Leggett, K. Li, D.C. McCune, P. Lane, M.E. Papka, Q. Peng, L. Randerson, A. Sanderson, J. Stillerman, M.R. Thompson, and G. Wallace. The national fusion collaborative project: Applying grid technology for magnetic fusion research. In *Workshop on Case Studies for Grid Applications (GGF'10)*, Berlin, Germany, March 2004.
- [20] Martin Szomszor and Luc Moreau. Recording and reasoning over data provenance in web and grid services. In *International Conference on Ontologies, Databases and Applications of SEMantics (ODBASE'03)*, volume 2888 of *Lecture Notes in Computer Science*, pages 603–620, Catania, Sicily, Italy, November 2003.
- [21] Chris Wroe, Carole Goble, Mark Greenwood, Phillip Lord, Simon Miles, Juri Papay, Terry Payne, and Luc Moreau. Automating experiments using semantic data on a bioinformatics grid. *IEEE Intelligent Systems*, pages 48–55, 2004.

### Part 3: Diagrammatic Project Plan

Time	SOCA Activities	External Events	Related Projects
3/04			PASOA start
3/04			VO start
10/04			EU-PROVENANCE start
11/04			GRIMOIRES start
01/05	<b>SOCA start</b>		
03/05	<b>co-located meeting 1</b>	GGF 13 eScience pilots	
06/05	<b>workshop 1</b>		myGrid end
07/05	<b>Researcher exchange 1</b>	GGF 14	
09/05	<b>co-located meeting 2</b>	AHM GGF 15	
12/05	<b>workshop 2</b>		GriPhyN end
01/06	<b>Researcher exchange 2</b>		
03/06	<b>co-located meeting 3</b>	GGF 16 eScience pilots	
04/06			GRIMOIRES end
06/06	<b>workshop 3</b>		
07/06		GGF 17	
09/06	<b>Researcher exchange 3</b>		EU-PROVENANCE end
10/06		AHM GGF18	
12/06	<b>SOCA end</b>		iVDGL end
02/07			PASOA end VO end