**Annex 2**

# RCUK REVIEW OF E-SCIENCE

## PUBLIC CALL FOR THE SUBMISSION OF EVIDENCE

Please complete and return this form to **e-sciencereview@epsrc.ac.uk** by **25th September 2009[1].** You must limit your submission to no more than 8 pages in length and no smaller than font size 11. You should address your comments to the issues flagged in the **evidence framework**. The headline questions only appear in this form.

All responses will be published on the website as part of the publication of evidence received by the Panel unless you state that there is confidential content for the Panel only.

| Name | Affiliation | Email Address | Website |
|---|---|---|---|
| Muffy Calder | UKCRC | muffy@dcs.gla.ac.uk | |

**Statement of interest (please indicate your reasons for making this submission - 200 words max.):**

The UK Computing Research Committee (UKCRC) was formed in November 2000. Its members are leading computing researchers from UK academia and industry. It gives scientific advice on computing research to the British Computer Society, the Institution of Engineering and Technology, the Committee of Professors and Heads of Computing; and it responds to appeals for evidence like this one.

The UK e-science programme has involved computer scientists in three roles. Firstly, computer scientists have engaged in the construction of the e-science framework and programme; secondly, computer scientists have collaborated in the majority of the projects within the e-Science programme, thereby assisting scientific research in other scientific disciplines; and thirdly they have tackled the general challenges of the collaborative exploitation of shared data and computational resources in modern science. Many of these challenges were pursued specifically in 'foundations of e-science' projects; their results are applicable generally over all branches of e-science; they are also relevant for broader commercial application. This direct experience of application has contributed to the development of Computer Science itself.

Our response concentrates on the experience of members of UKCRC who have been involved in the project, and on the future opportunities that may be opened up for the advancement of CS itself as a branch of e-science. We hope that the International review will find it helpful to take these points into account.

[1] Unfortunately no late submissions can be accepted.

**A. Did the UK e-Science Programme build a Platform which enables e-Science tools, infrastructure and practises to become incorporated into mainstream research in the UK?**

We believe that the application of Computer Science (CS) to the normal and now improved research practices of the other scientific disciplines (including medicine, engineering and sociology) has been as dramatically significant as the impact of CS on all other aspects of life of our society and of its members. Without the e-science programme, this development would surely have been delayed in UK.

The programme has certainly contributed widely to inter-disciplinary research between CS and other scientific disciplines. It has provided a model of collaborative tool construction and application which could well be adopted by researchers in the many other branches of CS. The UKCRC has been promoting such intra-disciplinary collaboration, as necessary to meet some of the basic scientific Grand Challenges of our subject, in particular the connection between computing theory and practice. Some of these challenges could well benefit from adopting and adapting the tools, infrastructure and practices pioneered by other scientific disciplines in the eScience Programme.

**B. How does UK e-Science activity compare globally?**

The UK is a world-leader in the application of Computer Science to the development of e-Science.

The UK has taken a broader view of e-Science than some other countries. The UK has developed production computational platforms such as the National Grid Service. We have taken a lead in the application of cloud computing to e-Science and application hosting (in particular the work of the Belfast and Newcastle e-Science Centres) and computational steering (for example in the EPSRC's RealityGrid pilot).

Furthermore, the UK has directed much of its research efforts to data management, knowledge management, and service management, addressing key challenges in data integration, service interoperation, workflow, semantic metadata, provenance, and secure, scalable data access. These challenges are now recognised as fundamental by the EU's European Strategy Forum on Research Infrastructures programmes and the NSF's CyberInfrastructure programme.

The UK can justifiably be said to have pioneered the application of semantic technologies and ontologies to e-Science and e-Infrastructure (the so called Semantic Grid). Mechanisms and standards for data access over Grids were led by the UK (the OGSA-DAI work). The UK's scientific workflow research has been technically influential, and the emerging software adopted by scientists world-wide (e.g. the Taverna from the EPSRC myGrid Pilot, and the Triana and P-GRADE systems) and transferred into commercial platforms (for example into the InforSense Knowledge Discovery Platform from the

Discovery Net and MESSAGE pilots).

**C. What has been the impact (accomplished and potential) of the UK e-Science Programme?**

We have evaluated the impact under ten headings, selecting just a few examples to illustrate each point.

**UK Centres of excellence**
Prestigious centres of e-Science research have been set up at most of the leading CS Departments in UK. They include Oxford, UCL, Newcastle, Manchester, Edinburgh, Nottingham, Imperial, Belfast and Southampton, most of which are major nodes in the UK's National Grid Service. The programme also established novel institutions that have developed close links with CS, such as the National Centre for e-Social Science and the first national Digital Curation Centre.

The UK has had the foresight to recognise that research software must be hardened, supported and sustained if it is to have a life time beyond prototypes and impact beyond its origins. The National Grid Service and the Open Middleware Infrastructure Institute-UK are important steps to the achievement of software sustainability.

**Digital Economy**
The eScience programme has been excellent grounding for the cross council Digital Economies programme, building on experience of common research challenges, inter disciplinary methods and intra-disciplinary mix of theory and applied computing. All three funded hubs of the Digital Economy Programme have their roots in the e-Science programme and are located at leading e-Science and e-Social Science centres driven by CS researchers (Watson at Newcastle, Rodden at Nottingham, Edwards at Aberdeen).

**Industrial connections**
Over 70 companies were involved in the first phase of the initiative; and some are funding follow-on work. Notably Microsoft Corp at Redmond, are still heavily engaged on novel publishing methods emerging from the EPSRC CombeChem pilot at Southampton; and they provide financially to support workat Newcastle on cloud computing and Manchester on data mining.

The Belfast e-Science Centre, in conjunction with the BBC and BT, have developed infrastructure to handle petabytes of personal media data and are pioneering mechanisms to extend the BBC iPlayer. Belfast developed the UK's first financial services grid with First Derivatives plc integrating the technology and the approach within First Derivatives product line.

The North East Regional e-Science Centre's approach to "cloud-based" e-Science garnered the following statement from the CEO of JBoss "your work has influenced vendor strategies as much as it has been influenced by those strategies, which is refreshing".

**Spin-out companies**
The MESSAGE and DiscoveryNet Pilots had direct exploitation routes into Inforsense, a pre-existing company at Imperial recently bought by IDBS. E-Science spin-outs include Cybula at York arising from the DAME project; Inkspot from the North East Regional e-Science Centre and EoverI from the Belfast e-Science Centre. The myGrid project based at Manchester is partnering with an established biotech support company Eagle Genomics.

**International leadership**
The Digital Curation Centre has combined theoretical computing research with practice. It is the model for EU's PARSE programme.
The Protégé-OWL ontology development tool is currently being used to develop the next generation of the International Classification of Diseases by the World Health Organisation.

The National e-Science Centre and the e-Science Institute have taken an international lead in education and dissemination, bringing together computer scientists with natural scientists. They are now models for countries throughout Europe and Australia.

The UK showed significant leadership on moving Grid standards to Web Services, which was strongly endorsed by industry, including Microsoft and IBM, who committed resources (people and project funding) to facilitate this.

**International standardisation**
Leading UK CS researchers have been instrumental in leading international standards, for example in: Open/Global Grid Forum including architecture (OSGI), Data access (OGSA-DAIS), Job submission (JSDL), Grid programming (SAGA), and Semantic Grid; and W3C Semantic Web languages RDF, SKOS and OWL.

The W3C OWL 2.0 specification, driven by realistic life science case studies, is being incorporated into the chEBI chemical compounds ontology by the European Bioinformatics Institute.

Research into provenance of data and workflows has lead to new mechanisms for data representation in large scale astronomy and biological archive data. The international Open Provenance Model movement is now forming part of a W3C incubator in provenance representation. OPM has been adopted by Microsoft Corp in their Trident scientific workflow system.

**Marks of international distinction**
UK CS researchers have spoken at every major international conference in the field and chaired most of them. Our annual All Hands Meeting attracts over 600 delegates including significant numbers from overseas and from industry. UK CS researchers have also advised on NSF, EU and OECD e-Science and e-Infrastructure policy. The technical manager of the EU's EGEE III pan-European e-Infrastructure emerged from the UK (Newhouse).

UK projects won top awards at the leading International SuperComputing Conference in 2002, 2005 and 2006. The inaugural winner of the Microsoft Jim Gray award for outstanding contribution to eScience went to Carole Goble – a UK Computer Scientist (2008). The UK e-Science Programme won the first ”peoples award” at GridWorld (2006).

**Interdisciplinary connections**

The programme has activated a genuine appreciation from many scientists for the methods of software engineering which deliver benefit not only in the original implementation of new tools, but also in subsequent tool evolution, driven by the needs of scientific researchers.

The Taverna workflow system, is used by over 350 organisations world-wide including universities, public and private scientific research institutes national and international, and private industry (including 25 SMEs and 12 major R&D Biotechs/Pharmaceutical companies). Moreover, although it was originally developed for the Life Sciences, it has subsequently been adopted by astronomy, chemistry, social sciences, climate modelling and notably by the US NCI caBIG cancer cyberinfrastructure and the EU NordicGrids.

The DAME distributed decision support system for real-time monitoring of airplane engines has been adopted by Rolls-Royce; moreover the techniques developed have been transferred to a virtual laboratory (CARMEN) for neurophysiology where the signals are neural activity recordings rather than engine instrument readings.

**Educational benefits in UK**

Within CS itself, e-Science topics such as distributed and grid computing have established themselves with real case studies in undergraduate and postgraduate curricula.

CS PhD students working in the e-Science programme include the EuroSys Roger Needham PhD Award winner (Cook, 2007) and the runner up of the CPHC/BCS Distinguished Dissertation competition (Missier, 2008).

**Computer Science Research**

The e-Science programme stimulated fundamental and applied computer science research into problems prevalent in e-Science applications and large scale infrastructure through investments in established CS Interdisciplinary Research Centres, an EPSRC Fundamental of CS in e-Science special programme and by bringing together CS researchers and scientists in pilots and centres.

Work on the use of description logics and ontologies for the representation of biomedical knowledge by Oxford and Manchester led to new ontology development tools (Protégé-OWL) adopted by the World Health Organisation amongst many others, and fundamental research into ontology modularisation, explanation and the representation and reasoning over structures.

**D. What are the future opportunities for UK e-Science?**

The programme successfully stimulated working collaborations between natural scientists and computer scientists. It also established strong international collaborations. It is important that these established collaborations continue and be allowed to flourish, and that we maintain the research environment that has made such collaborations possible. This will require continued coordination between the research councils to enable mutually beneficial inter-disciplinary research.

UK CS research could benefit from using some of the e-Science technologies to better facilitate collaboration, sharing test data sets and application repositories. (e.g. for machine learning research, software verification research and other large-scale scientific challenges) There is a great opportunity for the EPSRC ITC and Digital Economy programmes to explicitly support such exploitation.

The substantial legacy of applications and data arising from the programme should not be allowed to whither as the underlying technologies evolve. We strongly support initiatives that release academia from a perpetual process of rewriting and reinvention. Technical approaches for the preservation of data and environments, such as virtualisation should also be pursued alongside initiatives that support long term data conservation and software sustainability.

**E. How did the Programme Strategy (having a Core and individual Research Council Programmes, developing tools and applications in parallel) affect the outputs from UK e-Science?**

The programme needed to rapidly establish and coordinate a cross-council core of activity to establish key infrastructure, international and national strategy, cross-council links (for example with JISC and the DTI), and cross-council centres such as the National Data Curation Centre, the National Grid Service and the Open Middleware Infrastructure Institute-UK. The Core programme was crucial in this.

The research councils' individual programmes were required to build programmes relevant to their disciplines and in some cases specialist centres such as the National e-Social Science Centre and the NERC e-Science Centre.

**Any other comments**

At the outset of the UK e-Science programme there was a concern in many quarters of the CS community that their role would be simply to provide programming services to science projects. As this response indicates these fears were unfounded and significant computer science was embedded in the programme. In part this came from leadership within the CS community and

in part from vigilance by the funding agencies (predominantly EPSRC): including a core CS research component in the programme was critical. It is important that e-Science continues the outward transfer of computing expertise to other disciplines, and at the same time core computer science has the resources to pursue its own scientific agenda.

Please complete and return this form to **e-sciencereview@epsrc.ac.uk** by **25th September 2009**.