RESEARCH
COUNCILS UK

DELIVERING THE UK'S E-INFRASTRUCTURE FOR RESEARCH AND INNOVATION

Report commissioned by the Department for Business Innovation and Skills

July 2010

E-INFRASTRUCTURE
FOR RESEARCH AND INNOVATION

Contents

'Continue funding policies that strongly encourage or require the creation and adoption of shared e-infrastructure'[1]

1. INTRODUCTION AND EXECUTIVE SUMMARY

1. The development of the UK's Research and Innovation e-infrastructure has arrived at a critical juncture. Do we build on the great success of the last few years and exploit the opportunities of the future or do we stand on the side lines and watch the UK's global advantage dissipate and fade? E-infrastructure is essential to high quality research and contributes to the international position of UK science. Rather than being a research by-product, it represents a critical national asset in its own right. This report recommends a clear way forward to secure e-infrastructure, vital to the UK as we come out of recession, and looks to the future.

2. In this report we define e-infrastructure to mean:

    - digitally-based technology (hardware and software),
    - resources (data, services, digital libraries),
    - communications (protocols, access rights and networks), and
    - the people and organisational structures needed to support modern, internationally leading collaborative research be it in the arts and humanities or the sciences.

    and the combination and interworking of all these.

3. The UK's Science and Innovation research base needs to further develop and manage its e-infrastructure to:

    - ensure targeted and co-ordinated investment that reduces ongoing costs through efficiency savings and facilitates the 'invest to save' approach;
    - sustain global competitiveness by maintaining momentum to retain our current edge as highlighted by the recent International Review of the UK's e-Science programme;
    - strategically develop and sustain core technologies, software, data and services that can be relied upon by our leading scientists, liberating them to become more effective and competitive;
    - drive greater adoption of existing world class e-infrastructure investments by our scientists in the field by bridging technology gaps, active promotion and reward, and training;
    - fully exploit and share the advances and expertise deployed in one area of research for the wider benefit of UK research through supporting interoperability and exchange between projects based upon common shared infrastructure. Disciplines can reduce wasteful reinvention and unnecessary repeated work. By increasing collaboration, sharing and reuse across the research community we drive gains in research productivity and innovation;

[1] Building a UK Foundation for the Transformative Enhancement of Research and Innovation: Report of the International Panel for the 2009 Review of the UK Research Councils e-Science Programme, February 2010

- respond to the ‘data deluge’ with new approaches to research. We need to support and promote new methods of working such as automated data collection and analysis pipelines and new technologies such as cloud and semantic computing;
- provide critical support to key national objectives, help facilitate interdisciplinary research, national and global challenges, better capitalise on investments in one area for the benefit of all and foster economic development and competitiveness of the UK.
- enable past, current and future investments in e-infrastructure to have the greatest possible impact both across and beyond the research base.

5. This report has been commissioned by the Department for Business, Innovation and Skills (BIS) with the aim of assessing the progress that the UK has made in driving forward the development of a world class e-infrastructure and to recommend what steps now must be taken to support a productive, globally competitive research base in the future. The report has been written against a backdrop of rapid technological change which includes, for example, the emergence of “cloud computing”, “green computing”, “Web 2.0”, “software as a service”, “open access repositories”, “crowd-sourcing”, “semantic linked open data” and “data.gov.uk”.

6. BIS invited RCUK to take the lead on the collection of new, and review of other recent, evidence which has formed the foundation of the report. Key amongst this evidence has been:

- a specially commissioned report seeking the views of key stakeholders involved in the funding, development and roll out of e-infrastructure, completed by the Research Information Network (RIN) in December 2009;
- The OSI report *Developing the UK’s e-infrastructure for science and innovation* (hereafter OSI report), published in 2007 http://www.nesc.ac.uk/documents/OSI/report.pdf
- the RCUK International Review of e-Science which reported in February 2010, http://www.epsrc.ac.uk/pubs/reports/Pages/internationalreviews.aspx

7. The RCUK e-Science review was conducted by an international panel of 15 experts in the latter part of 2009. It reviewed a substantial evidence base of documentation and conducted a week-long programme of presentations, interviews and visits covering 60 projects in December 2009.

8. This evidence has been marshalled and interrogated by an Expert Group made up of a range of stakeholders concerned with the strategic development of the UK’s e-infrastructure (membership attached at Annex 1). The Group has been chaired by Professor Carole Goble, University of Manchester. More detailed background information of the evidence underlying the report can be found at Annex 2.

9. The International Review of e-Science provides important evidence and recommendations which have informed this report. While closely aligned to the review, this report does not set out to provide a comprehensive implementation of its recommendations.

10. The agencies specifically addressed by the report are the Research Councils[2] and JISC[3]. The report also seeks to build a forum for cooperation and partnership with other stakeholders, notably: other major research funding agencies such as the Wellcome Trust and Technology Strategy Board (TSB); industry; government departments; the HEI-funding bodies; representatives of the scientific community and beacon individual HEIs and research institutes. These are referred to as “strategic partners” throughout the report and accompanying action plan.

11. The Expert Group acknowledges the rich existing landscape of e-infrastructure of all kinds currently supported by the Research Councils, JISC, HEIs, charities such as the Wellcome Trust, and European, international, industrial and private investments. Annex 5 gives a spending summary of each Research Council plus JISC on e-infrastructure, totalling over £170million per annum. JISC is particularly acknowledged as an important supplier of networking, data licensing, access management and digital repository infrastructure and a proactive programme of shared infrastructure provision. The Group is well aware that cross-European and international initiatives such as the ESFRI programmes are significant UK investments and are, or will make, significant contributions to the UK’s research infrastructure.

12. The Expert Group concluded that the UK’s e-infrastructure was amongst the best in the world, supported by sustained investment over recent years. The Group also noted that the UK’s competitive advantage is being eroded by major investment in e-infrastructure in Europe, Australia and notably the US through its cyber-infrastructure initiative and more recent stimulus packages. More needed to be done to harness and develop the strengths of the UK’s infrastructure in the face of these challenges. If steps are not taken then the UK would fail to fully capitalise in its own significant investment in e-infrastructure, and would undoubtedly fall behind the global leaders in producing the very best international research.

---

[2] Research Councils (there are currently seven Research Councils) are the public bodies charged with investing tax payer’s money in science and research in the UK in order to advance knowledge and generate new ideas which can be used to create wealth and drive improvements in quality of life. Each Research Council funds research and training activities in a different area of research ranging across the arts and humanities, social sciences, engineering and physical sciences and the medical and life sciences. The Councils support around 30,000 researchers, including 15,500 doctoral students in UK universities and in their own Research Institutes (www.rcuk.ac.uk).

[3] JISC inspires UK colleges and universities in the innovative use of digital technologies, helping to maintain the UK’s position as a global leader in education. JISC is funded by the UK HE and FE funding bodies to provide world-class leadership in the innovative use of ICT to support education and research. JISC manages and funds 193 projects within 29 Programmes. Outputs and lessons are made available to the HE and FE community. JISC also supports 29 Services that provide expertise, advice, guidance and resources to address the needs of all users in HE and FE (www.jics.ac.uk).

13. The Expert Group now makes the following recommendations that were agreed to by all its members:

- The UK's Research and Innovation e-infrastructure should be treated as a strategic National Facility. RCUK should ensure that major Research and Innovation infrastructure projects are included in the Large Facilities Roadmap, and considered for submission to the Large Facilities Capital Fund, where they meet the criteria for the Fund.

- The UK's Research and Innovation e-infrastructure needs to be led and driven to deliver a UK wide vision for research e-infrastructure, embedded in the international context essential to today's research challenges. The leadership must provide a multi-year perspective, identify best practice, coordinate stakeholder investment and champion relevant and fit for purpose cross-disciplinary standards to facilitate coordination.

- The leadership of the UK's Research and Innovation e-infrastructure should be delivered by a Director of e-infrastructure, overseen by a Strategic Board with members drawn from areas closely related to e-infrastructure, such as software specialists, research technologists, and specialists in data management. The Directorship would run for five years (subject to approval of a business plan prepared by the Director during the first year) and comprise of a director, a deputy and support staff.

- The Director would develop a business plan to deliver the strategic vision. This will include a detailed implementation strategy and a fully costed budget statement. The business plan would be considered and agreed by the Strategic Board.

14. These recommendations clearly recognise that different disciplines have different needs and do not represent an attempt to impose a uniform e-infrastructure across the research base. One size does not fit all. The recommendations are, rather, geared more to building upon and more fully exploiting existing strengths particularly where e-infrastructure can be effectively generalised for mutual benefits and scientific gain as well as delivering significant medium and long term efficiency savings.

15. In the absence of stronger strategic leadership and co-ordination all the evidence points to the real danger of duplication and significant wasted investment. What this report calls for is the fuller exploitation of infrastructure and expertise deployed in one domain for the wider benefit of UK research. A far greater sharing of e-infrastructure also offers up the potential to fuel exciting new multi-disciplinary research which it is widely recognised is now critical to address the most pressing scientific challenges.

16. This report is not arguing for massive new investment in e-infrastructure at a time of scarce public funding. Nor is it seeking to interfere with e-infrastructure that is already successfully deployed or tamper with domain-specific e-infrastructure. What it is arguing for is the more efficient deployment and strategic development of existing e-infrastructure through a

targeted programme of investment steered through light touch leadership and co-ordination. By cutting out waste and duplication much of the potential investment suggested in this report will be recouped through efficiency savings.

17. The compelling need to develop both co-ordinated and cost efficient e-infrastructure fit for a knowledge-driven research base is not just a UK one. The USA has invested in data-driven science through the coordinated Data Net programme. The Obama stimulus package has provided a further $111 billion[4] for infrastructure and science including strategic co-ordinated investments in infrastructure to support researchers in HEI, for example vivoweb project[5] and software sustainability. This largescale programme of 'invest to save' has helped to move the US away from what was previously a 'balkanised' infrastructure which was costly, duplicative and inefficient.

18. Recent evidence also highlights the success of the Dutch e-infrastructure programme that to a large extent has been developed and delivered through 'lean' investment coupled with effective national planning and coordination.[6] In Europe the e-IRG (e-Infrastructure Reflection Group)[7] roadmap stresses the critical need for co-ordinated strategic investment arguing that this is the most cost effective way to develop e-Infrastructure services that will be essential for 40 million users in research and academia in Europe. Through its National Collaborative Research Infrastructure Strategy[8] the Australian Government is providing $542 million over 2005-2011 to provide researchers with major facilities, supporting infrastructure and networks necessary for world-class research. As part of the Strategy Australia, for example, has invested in coordinated e-infrastructures such as the Australian National Data Service[9].

19. The report suggests there are five broad areas (as shown in the Figure 1) where further strategic investment could add particular value to the UK's existing e-infrastructure. These are:

- Sustainability of new and existing mission critical e-Infrastructure supported by the research funding agencies and JISC;
- Interoperability between new and existing e-Infrastructure supported by the different research funding agencies and JISC;

4 http://www.stimuluspackagedetails.com/details.html
5 http://www.vivoweb.org
6 Gordon Cook, *Building a National Knowledge Infrastructure. How Dutch Pragmatism Nurtures a 21st Century Economy. The Cook Report in Internet Protocol.* SURF Utrecht, the Netherlands, 2010, http://www.cookreport.com/knowledge.pdf
7 The e-Infrastructure Reflection Group was founded to define and recommend best practices for the pan-European electronic infrastructure efforts. It consists of official government delegates from all the EU countries. More details on the e-IRG is available at http://www.e-irg.eu/about-e-irg/mission.html
8 http://ncris.innovation.gov.au/Pages/default.aspx
9 The Australian National Data Service (ANDS) aims to influence national policy in the area of data management in the Australian research community as well as to transform the disparate collections of research data around Australia into a cohesive collection of research resources, http://ands.org.au/

- Adoption of e-Infrastructure into mainstream use by a majority of researchers by support from the research funding agencies and JISC;
- Capacity building of the skills base for creating and using e-infrastructure;
- e-Infrastructure for data to serve data intensive research

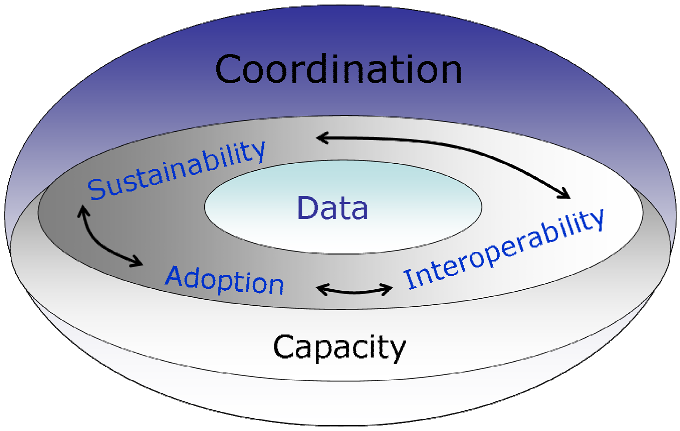

Figure 1: Areas for Action underpinned by co-ordination and leadership

20. Each of these areas is discussed in detail. It is important to emphasise that the areas are closely interlinked with a degree of overlap. Data in particular cuts across all these areas for action as there is real need to improve coordination, sustainability, interoperability, adoption and capacity to fully harness and exploit the exponential growth in data in order to support high quality scientific research.

2. KEY RECOMMENDATIONS

Recommendation 1 The UK's Research and Innovation e-infrastructure should be treated as a strategic National Facility. RCUK should ensure that major Research and Innovation infrastructure projects are included in the Large Facilities Roadmap, and considered for submission to the Large Facilities Capital Fund, where they meet the criteria for the Fund.

21. The recently published *2010 Large Facilities Road Map* starts to recognise the underlying importance of e-infrastructure in delivering leading edge science. In defining large facilities the Road map states:

"*facilities include the traditional large physical installations but increasingly they also take the form of distributed, networked resources that exploit advances in information and communications technology to underpin new collaborative modes of communications. Technologies are leading to the development of means to generate, store and exchange an increasing volume and diversity of data. These provide the wherewithal to address large and complex problems through the integration of many scientific approaches and individual inputs. This means that the nature of large facilities is changing. In particular, in many fields distributed communities will increasingly rely on more centralised facilities to provide data and experimental platforms that provide the basis for integrated and large scale research efforts. These may be focused around particular technologies, or upon the consideration of particular problems or user needs. This evolution of large facility needs is likely to continue for the foreseeable future as technology develops, and therefore the roadmap represents a snapshot of a rapidly changing landscape, which will need to be regularly reviewed.*"

22. This definition now needs to be taken one step further and to conceptualise facilities in the broadest sense, so they fully embrace the key elements of the UK's e-infrastructure (as defined in this report). It needs to move e-infrastructure out of the shadows and raise its visibility as a critical national asset which naturally forms part of the Large Facilities Roadmap. This will help to build confidence and raise the profile of e-infrastructure which in turn will promote greater long term stability and sustainability both through LFCF funding and investment from other agencies and HEIs.

One example of the critical way in which the LFCF could be deployed to drive forward the development of e-infrastructure is in the creation of shared computing services. This could take the form of a core capacity and capability service for UK researchers allowing institutions and projects to access computing resources on demand. A federated infrastructure will be essential to exploit existing and future investments effectively. Initial investment would provide the impetus for institutions and projects to overcome the barriers to deployment of shared infrastructure. This will then be sustained by normal project/institutional investments, but will leverage economies of scale, expanded base expertise, appropriate international engagement and an ability to rapidly exploit new technologies. Such an infrastructure could be in large part based on cloud computing provision, but this would be driven by user requirements not technology.

Recommendation 2 The UK's Research and Innovation e-infrastructure needs to be led and driven to deliver a UK wide vision for research e-infrastructure, to provide a multi-year perspective, identify best practice, coordinate stakeholder investment and champion relevant and fit for purpose cross-disciplinary standards to facilitate coordination.

23. The recent international e-Science Review provided a critical assessment of the UK's e-science initiative. The review concluded that the initiative had created a 'world leading' infrastructure in many areas of UK science. However, the review also warned that to fully capitalise on this infrastructure and to drive forward its further development, dedicated leadership and more systematic, ongoing coordination of investments was now critical:

"*further success will not be achieved purely through technological determinism nor through the uncoordinated activities within the various disciplinary communities. Moving forward requires coordination, clever design, effective leadership, and long-term commitment to a system of linked and balanced interaction between the various communities and sponsors*".

24. The e-Science Review recommended that the UK should "*establish organisation and management structures to continue to treat e-Science as a designated strategic initiative spanning all Research Councils".* Major conclusions and recommendations of the review are attached at Annex 3.

25. The RIN report similarly highlighted a pressing need to improve co-ordination which was viewed as a major obstacle to the efficient and effective exploitation of the UK's e-infrastructure. The report emphasised the need for a more co-ordinated approach in the development of e-infrastructure at many levels: between government departments, between research and development agencies, including the Higher Education Funding Councils, the Research Councils and JISC, and between key national resource and service providers. As one of the interviewees pointed out: 'there are a lot of stakeholders in the field, but nobody occupying the centre ground'

26. The need for this more strategic and co-ordinated approach is particularly important. In the midst of a period of economic uncertainty, cost-efficiency, minimising wasteful duplication and fully capitalising on prior investments are key requirements. This report does not call for the invention of major new infrastructure, through a substantial new e-science initiative or broader e-infrastructure programme. Instead it stresses the need to add value to what is already a well founded e-infrastructure through more modest targeted funding which adds value to existing investment and through coordination of ongoing investment.

27. The ability to effectively 'invest to save' will clearly rest on the early identification of a coherent and widely shared strategic vision for the future development and exploitation of national e-infrastructure. This will need to drive forward the development of common e-infrastructure requirements and standards, promote the more effective sharing and strategic development of technologies software and services, agree common standards across

research disciplines and exchange best practices. It should be stressed that the aim is not to create a single e-infrastructure for all, to interfere with e-infrastructure that is already successfully deployed or tamper with domain specific e-infrastructure. It is rather to exploit the huge potential to share infrastructure where there is obvious benefit in research productivity.

28. The area of data, in particular, is offering a large scope for collaboration across agencies. UK Research Data Service, funded by HEFCE, has recently conducted a feasibility study arguing for standardised approach to data storage (similar to what the Australians are doing, as mentioned above). It may potentially become one of the areas where coordinated joint effort will bring efficiencies through working across dual support system.

Recommendation 3 The leadership of the UK's Research and Innovation e-infrastructure should be delivered by a Director of e-infrastructure, overseen by a Strategic Board with members drawn from areas closely related to e-infrastructure, such as software specialists, research technologists, and specialists in data management. The Directorship would run for five years (subject to approval of a business plan prepared by the Director during the first year) and comprise of a director, a deputy and support staff.

29. Whilst much of the existing evidence collected and reviewed by the Expert Group pointed to the need for stronger leadership and better co-ordination of the UK's e-infrastructure, there were few suggestions as to how that might be achieved. The Group recommends the appointment of a Director of e-infrastructure overseen by a strategic board. The Director will need to report into government via a nominated RCUK Chief Executive or nominated CEO deputy (as demonstrated in the Figure 2).

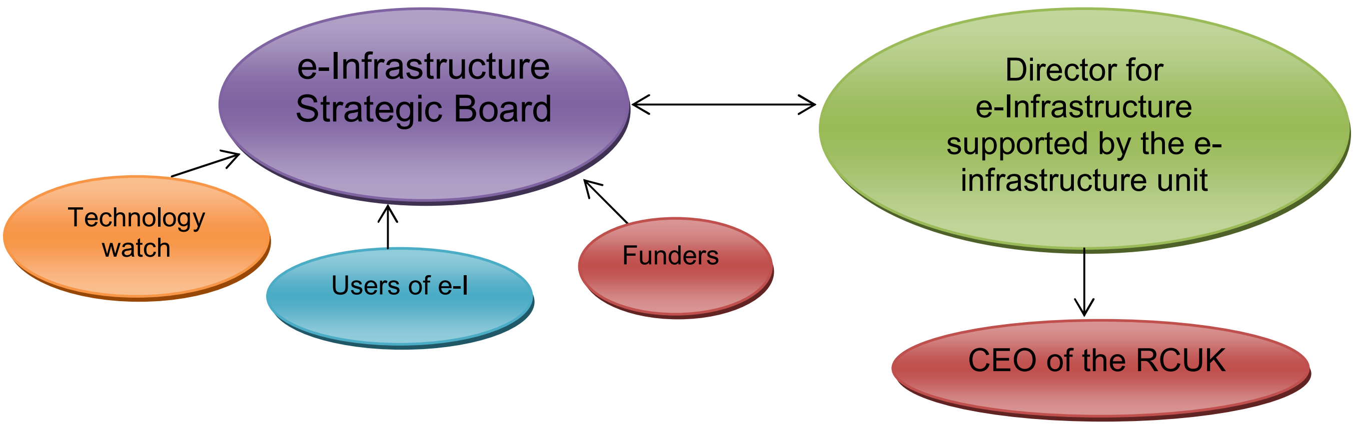

Figure 2: Governance arrangements for UK's e-infrastructure

30. Broadly speaking, the Director will be responsible for:

- strategic leadership for the development of the UK research e-infrastructure;
- development of a coherent and shared strategic vision for the UK's e-infrastructure ensuring its maximum impact;
- conducting an audit of existing e-infrastructure investment to help identifying leverage for coordination and opportunities for efficiency savings;
- identification of strategic investment which would add value to existing e-infrastructure;
- the spreading of best practice and expertise in complementary scientific domains;
- creating strategic synergies with funding agencies (for example, Research Councils, Funding Councils, JISC, British Library), service providers, industry and other strategic partners;
- developing international strategic links to support the development of e-infrastructure.

31. During the first year the Director will be responsible for developing a strategic vision for e-infrastructure through a process of detailed consultation with a range of stakeholders. This will be a critical process in raising the profile and scientific value of e-infrastructure and its importance for technological and economic growth. The international e-Science review rightly points at the importance of taking the *'… opportunities for upper management of universities, government, and industry to learn more about the fundamentals and strategic importance of e-Science and e-infrastructure'.* Based on the consultation, the Director will also be responsible for preparing a detailed action plan which flows from this strategy identifying a potential programme of strategic activities that will add value and further develop the UK's e-infrastructure. The Director will develop a detailed business plan with an accompanying budget for this programme of activities.

32. The Director is expected build a forum for cooperation and partnership with strategic partners other than the Research Councils and JISC, notably: other major research funding agencies such as the HEI-funding bodies, charities such as the Wellcome Trust; TSB and industry; government departments; representatives of the scientific community and individual HEIs and Institutes.

33. The Director will not just be expected to shape and develop this strategy by looking to developments within the UK. The e-Science Review stressed the key international lead taken by the UK in the development of e-infrastructure and this is expected to continue through the Director. The Director will play an important role in building up global links promoting cross-European and international co-ordination, through for example the e-IRG and related ESFRI activities.

The UK's National Grid Service, our contribution to Europe's Grid Infrastructure, is an example of coordinated effort of contributions by Research Councils, JISC and individual HEIs. Integrated with partner infrastructures in Europe, the USA, and elsewhere in the world, the NGS ensures that UK researchers can effectively and efficiently exploit facilities and collaborations across the world. Over the past four years the NGS has established a leadership position in e-infrastructure provision for the UK internationally.

www.ngs.ac.uk

34. One of the major recommendations of the e-Science Review was to *'establish more systematic and better supported mechanisms, including targeted funding, to nurture collaboration and bi-directional knowledge transfer between academia and industry in the creation, provisioning, and application of e-Science'*. There is a strong evidence base of the benefits of close co-operation between academia and industry. The Director will, therefore, be responsible for strengthening the existing links, taking this recommendation further, identifying better ways of connecting academia and industry, engaging relevant industry in the creation and potential provisioning of the e-infrastructure platform for research and innovation, as well as helping industry, where and when possible, adopt and tailor best practices and services from the academia to enhance their own productivity.

Good example demonstrating the benefits of effective cooperation and knowledge transfer between academia and industry is the Distributed Aircraft Maintenance Environment Project (DAME). As part of the e-Science programme, DAME partnered with Rolls-Royce, Data Systems and Solutions and Cybula Ltd to use e-Science to reduce engine maintenance times and to improve the interoperation of the maintenance team. The technologies developed are now used on Rolls-Royce Trent engines and the result was a spin-off company: Oxford Bio-Signals (OBS).

www.cs.york.ac.uk/dame/

35. The Director's activities will be steered by a Strategic Board for e-infrastructure. To give this Board credibility across a range of communities, its membership will draw upon both national and international expertise from areas closely related to e-infrastructure including software specialists, research technologists and experts in data management. It will include relevant funders, users and technology experts, and will be chaired by a senior representative, possibly from industry.

36. The work of the Director and Board will be informed by ongoing dialogue with users and funders and supported by key technology watch activities. In addition there will be a major annual forum that will bring together representatives from the research community, service providers, funding agencies, industry, government and other stakeholders and strategic partners. As part of the annual Forum there may be scope for thematic workshops, for example to address issues around technological developments.

Recommendation 4. The Director would develop a business plan to deliver the strategic vision. This will include a detailed implementation strategy and a fully costed budget statement. The business plan would be considered and approved by the Strategic Board

37. The initial task of the Director will be to develop a coherent Strategic Vision for e-infrastructure and then To develop and deliver the vision it is anticipated that funding will need to be made available to cover a full time appointment of the Director and part time Deputy Director although the exact time commitment and allocation of responsibilities between these two posts may be flexible based upon pre-existing commitments. Whatever the eventual time commitment and breakdown of responsibilities, the Director will need to be a senior figure who will command respect and authority across the full range of stakeholder groups.

38. Funding for the Directorship would also need to cover some basic administrative costs, national and international travel costs and support for the Strategic Board's meetings and the proposed annual forum

39. It is important to stress the need for both a full-time Director and a part-time Deputy Director. This is to ensure that the ambitious task of trying to provide effective leadership across such a wide and diverse waterfront as that covered by e-infrastructure is in any way feasible. This will involve ensuring that the Director and Deputy Director bring together the necessary complementary expertise to develop a strategic vision and action plan that has credibility across the research base and beyond.

40. The Strategic vision and action plan will be accompanied by a detailed budget covering a programme of strategic activities. The budget will need to be carefully scrutinised by the Strategic Board prior to being placed before BIS for further consideration. It is not currently clear what the size of the budget might be. That will be determined by scale and depth of agreed programme of activities.

A few examples of possible Director's investments:

1. Specific targeted projects, using the ENGAGE programme as a model for enabling adoption and interoperation of existing infrastructure. For example, NERC's investments in Geo-Spatial services (maps) could be adopted by ESRC services, and even MRC and BBSRC – software for zooming over maps could be used to zoom over tissue images.
2. Targeted investments in services where no one Research Council has ownership yet common services would yield benefit. A clear example is the citing and quality assurance of data. This needs a data citation service, interoperation with journals, library and digital repository systems, data quality methods and training. And to enable adoption by international efforts such as DataCite. A citation service removes a barrier to data sharing. More data sharing means less need to continually reinvest in the same data capture over and over again.
3. Targeted investments in preparing services for better sustainability, particular the services, data and software that is inter-disciplinary or trans-disciplinary or foundational and thus suffers from the tragedy of the commons in the sense that everyone wants to use it but nobody wants to look after it. The SSI (Software Sustainability Institute, funded by the EPSRC) could be an instrument for this.
4. Coordinating multiple investigations into the same infrastructure. For example duplication implied in multiple Institutional repositories. Another example is cloud computing. EPSRC, BBSRC, UKTI, JISC and individual HEIs are all investigating cloud computing for researchers but the efforts are not joined up. We could obtain real economies by engaging with cloud providers for a UK Research Cloud in a coordinated manner.

3. STRATEGIC CONTEXT - AREAS FOR ACTION

41. There are five broad, interconnected areas where there is a real opportunity to add value to the UK's e-infrastructure through dedicated leadership, stronger co-ordination and further strategic investment. These should form the basis of any future action plan (indicative action plan suggested by the Expert Group is attached at Annex 5). The areas are:

- **Sustainability** of new and existing mission critical e-Infrastructure supported by the research funding agencies;
- **Interoperability** between new and existing e-Infrastructure supported by the different research funding agencies;
- **Adoption** of e-Infrastructure into mainstream use by a majority of researchers by support from the research funding agencies;
- **Capacity** building of the skills base for creating and using e-infrastructure;
- e-Infrastructure for **data** to serve data intensive research

3.1. LONG-TERM SUSTAINABILITY OF E-INFRASTRUCTURE

"Sustainability … is a recurring concern[14]"

42. It is widely acknowledged that sustainable e-infrastructure is a prerequisite for scientific advancement. It allows scientists to build confidence in the availability of core technologies, software, data and services encouraging adoption and widespread use. However the RIN report stresses that there are fundamental concerns with the lack of any apparent systematic strategy to develop a stable e-infrastructure across the UK research base which is inhibiting uptake and exploitation in many areas of science and research. The international review of e-science also points to the pressing need to "*sustain the operational e-infrastructure for e-science"*, and that "*sole reliance on business as usual grant funding through the individual research councils is not likely to take full advantage of the gains to date or the potential for the future*.".

43. A first critical step towards achieving sustainability is to agree what constitutes national mission critical e-infrastructure. This will require all key stakeholders and strategic partners to establish what represents 'core' e-infrastructure in their respective different scientific domains. Such consensus, whilst difficult to achieve, is essential for setting up a base line for the funding and future development of e-infrastructure. Without credible commitment to a sustained infrastructure, individual researchers, research groups, and perhaps most importantly research institutions will be unwilling to commit to it themselves, trusting rather the ad-hoc development of effort funded through the short term project grants, which brings with it all the attendant shortcomings of nugatory investment at best and wasteful duplication at worst.

44. Clearly different sustainability strategies apply to different kinds of e-Infrastructure (e.g. services, software, data sets and hardware) as well as different disciplines (current spending on e-infrastructures by individual Research Councils and JISC can be seen at Annex 4). In particular the importance of maintaining access to research outputs such as scientific papers and reference datasets cannot be overstated. To achieve this goal national 'memory institutions' such as the British Library, and possibly international ones such as the European Bioinformatics Institute, must form an integral part of these strategies.

> The ESRC has been funding the Economic and Social Data Service since the mid 1960s and through a process of sustained investment has built a world leading data facility. The service now supports nearly 42,000 active users from academia, government and commercial sector promoting high quality research, informing policy analysis and shaping business practice. The success of the service rests on its early identification as a vital national resource for the storage, management and dissemination of social and economic data and its continued status as a mission critical part of the UK's e-infrastructure
>
> www.esds.ac.uk

3.2. E-INFRASTUCTURE INTEROPERABILITY

"A basic prerequisite for the success of a UK national e-infrastructure … is the integration and interoperation of the infrastructure's component parts."

45. New e-Infrastructure investments will only achieve adoption by users if they are interoperable with the software used by the researchers. e-Infrastructure investments in one discipline can only achieve adoption in another if interoperability is duly addressed. Interoperability between existing and new e-infrastructure is a key area that is closely related to coordination and sustainability. This is particularly important to achieve the economies promised through deploying and exploiting shared infrastructure and identifying the 'core' e-infrastructure that is interoperable. Spending on interoperability is about targeted benefits. For example, common standards and interoperability of data depositories are important steps towards a long-term sustainable and interoperable national strategic network.

The importance of interoperable e-infrastructure can be demonstrated through the example of e-health. The seamless integration of the Taverna workflow system originally developed to serve the Life Science community, into a Shared Genomics workbench makes tools written for bioinformaticians easier for clinical researchers to use. Effectively Taverna allows a scientist with limited computing background and limited technical resources and support to construct highly complex analyses over public and private data and computational resources, all from a standard PC, UNIX box or Apple computer. Taverna is now used in, for example, astronomy, digital document preservation, instrumentation simulations, environmental genomics and social census data analysis.

www.taverna.org.uk

Similarly, the standards-based interoperability of OpenCDMS (Open Source Clinical Data Management system) in clinical centres is to enable clinical researchers to manage the full life cycle of their clinical research project, from design through to archiving, without any specialist knowledge of databases or IT systems

www.opencdms.org

3.3 'CROSSING THE CHASM': ADOPTION OF E-INFRASTRUCTURE

46. "Crossing the chasm" (as shown in the Figure 3) refers to the challenges faced when e-infrastructure has proven itself with a small group of early adopters but has not been adopted by the mainstream majority. Barriers to adoption are practical (uncertain sustainability, unsupported, irrelevant), technical (too complicated to use, not interoperable with other systems, insufficiently robust, take-on costs too high), cultural (no reward in adopting, rewards for reinventing), legal/policy (licenses and data sharing policies can discourage adoption, local HEI IT policies forbid access to services and software) and capacity (insufficient skills or insufficiently skilled research workforce). In short, to cross the chasm e-infrastructure must be planned to

be usable from the onset, actively supported once developed and positively welcomed by researchers.

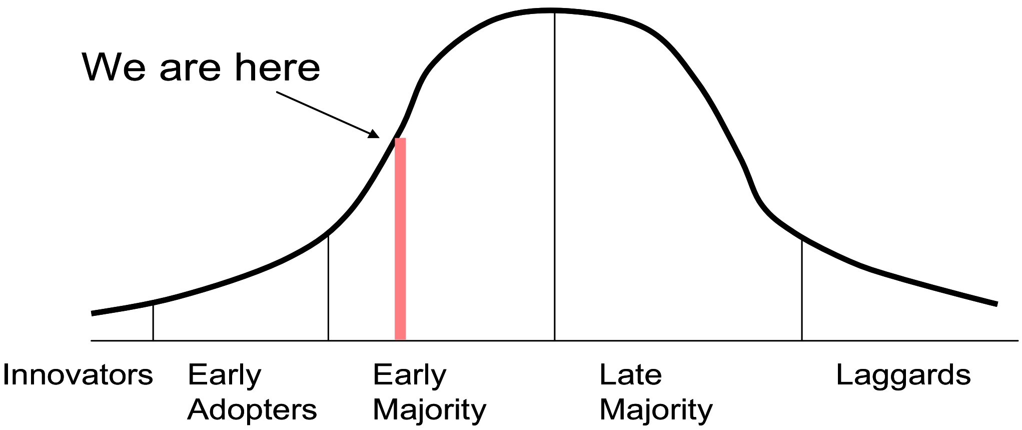

Figure 3 *"Crossing the chasm" diagram using the prototype of Geoffrey Moore's 1991 business classic, Crossing the Chasm, applied to Technology Adoption Life-Cycle (*Moore*, Crossing the Chasm,* 2nd edition, pp 11-12*)*

47. One way of achieving this objective is to encourage the use of mixed teams in e-infrastructure related projects. This would bring together researchers, research technologists and IT specialists who collectively can share their experience of best practice, reuse and re-purpose existing e-infrastructure and deliver products that meet the needs of the research community. Such teams would help to dissolve many of the cultural barriers to uptake and use and help to address one of the recommendations of the international e-science review called for *'training, making services available, and tailoring current services to more specific needs of disciplinary and interdisciplinary communities'.*

The JISC-funded Engage Academic Groups with e-Infrastructure project (Engage) is a powerful example of the advantages of purposeful adoption of e-infrastructure. Engage undertook a series of short development projects to make software originally developed for specialised research purposes usable by inexperienced users and available on national and local e-infrastructures. It empowered researchers to exploit the benefits of e-infrastructure by developing and deploying new software solutions on available UK e-infrastructure. Significant achievements of the project can be demonstrated with the following examples:
Climate modeling: making the Genie climate system model easy to use over the National Grid Service for post-graduate teaching and even public use;

Protein sequencing: making workflows to protein sequences publicly available, thus enabling students to perform analysis on a database of 15,000 protein sequences;

Ancient documents: adapting image-processing tools developed to study ancient manuscripts and making them available to different communities of researchers.

www.jisc.ac.uk

3.4. CAPACITY BUILDING OF THE SKILLS BASE

"e-infrastructures are often seen as complex and challenging"

48. Capacity building is an indivisible component of e-Infrastructure efficiency and is a prerequisite for successful adoption process. Only well-trained data producers and users are properly equipped to take advantage of existing and future e-infrastructure both in the UK and beyond, helping to close the chasm between the small community of early adopters and the rest. It is also important to develop the skills required for effective data production and consumption from the early years of research by embedding the necessary training in training courses. This training should target young and mid-career researchers, providing them with the set of skills necessary for early adoption. This set of skills should be embedded within research teams and become distributed widely across the whole research sector.

49. There is also a need to increase capacity to develop and not just exploit e-infrastructure. For this purpose it is vital to create and maintain a professional career structure for research technologists within the research organisations. Currently these critical, but often under-rated, staff survive on short-term research contracts. This fundamentally undermines the capacity to build a stable cadre of skilled research technologists who can share their expertise and experience, promote good practice and avoid duplication of effort through the use or re-purposing of existing e-infrastructure.

50. HEI Human Resource practices currently undermine efforts to promote sustainable team science and the pooling of skilled research technologists across multiple awards. Just when we need flexible research practices, HR policies are forcing inflexible hiring models on investigators.

3.4. DATA

"A new, fourth paradigm for science [is] based on data intensive computing. In such scientific research, we are at a stage of development that is analogous to when the printing press was invented."

51. Data are being collected on an unprecedented scale. Scientific data, in particular, are doubling every year.[10] The exchange of data is the most immediate and cost efficient means of inter-discipline, inter-researcher and inter-agency collaboration, leading to unparalleled opportunities to extend high quality scientific productivity across and within disciplines. It has been argued that some scientific research is as a consequence entering a 'fourth paradigm' based on data intensive computing.

52. New practices enabled by digital data pervade all disciplines. In the social sciences, "born digital" data enables new studies which stand to influence policy and practice. Arts and humanities researchers too are working with an increasing volume of digital content and benefiting from e-Infrastructure to work digitally and on the international stage.

[10] For example, there are more than 120,000 data report downloads annually supported by the Economic and Social Data Service. An article is published in PubMed every 30 seconds.

53. Under this paradigm data are not a mere by-product but an asset in their own right; arguably the primary long term asset. Data intensive research recognises the central role of data as a capital asset for innovative research. Our success as a nation in our research and innovation will depend on our ability to manage and process data and turn it into valued, reusable information and sharable knowledge. Every recent review of e-infrastructure has highlighted the importance and scale of data. For example the RIN report highlighted the need to store, manage, manipulate and share data more effectively as a key issue, whilst the international review of e-science identified data as a 'grand challenge' for research over the next decade.

> UK Biobank is an illustrative example of data resource that links, at unprecedented scale, many different sources of information on individuals for inter-disciplinary research. It aims to enable studies to be carried out on the relationships between genes, lifestyle and health through the collection of DNA samples and information from half a million people across the UK.
>
> www.ukbiobank.ac.uk

54. Traditionally data, computation and software have been treated as independent and non-interacting entities. This is clearly inappropriate; data fuels and validates computer intensive simulations that predict more data; data mining, search and visualisation are computer intensive tasks; simulations, visualisations, search engines and databases are software.

55. This report thus places data at the core of its proposed action plan, stressing now they are interlinked with and underpinned by other areas, For example, a dataset that is sustained will be adopted as it can be relied upon by its user base. A widely adopted dataset will improve its case for being sustained. A dataset that is able to be exchanged and exploited by multiple stakeholders and in multiple domains (that is, it is interoperable) is more likely to be adopted and more likely to be sustained through broadening its user base. Capacity building is essential to develop a skills base capable of contributing to and exploiting the dataset. Coordination and leadership would: share best practice for stewardship of data; help identify key datasets and ensure their stability; target standardisation practices needed to support interoperability and user access; and support cross-council software, services and policies needed to widen adoption, reward data sharing and encourage data reuse rather than reinvention.

56. The expanding volumes of research data require not only the development of new technologies, but thoughtful and purposeful data management and sharing mechanisms to enable researchers to exploit this key resource. This involves data collection and preservation for reuse and repurposing where e-Infrastructure plays a central role. This also involves the development and adoption of metadata standards, finding aids, tools for data manipulation and persistent identifiers. These facilities can enable data reuse that in turn leads back to the need for co-ordination and also training.

57. The issue of enabling reuse also leads to an obvious, but complicated issue of data sharing and openness of research process in general. The current culture (held in some environments) of protecting the researchers' 'intellectual capital' does not facilitate a sharing culture and open access to research outputs. This needs to be addressed in a coordinated manner across the funding agencies, but also involving government departments. Without doubts data are key to an informed public policy. Opening up access to non-personal public data has been strongly advocated by the government with its project 'Making Public Data Public' that is seen as a platform for developing new technologies and new services.[11]

58. Along with encouraging data sharing, there is a clear need to have an infrastructure in place to enable access and, interoperability, and one that supports data citation that will credit providers, assure quality and provenance and maintain confidence amongst data users. Such data citation approaches are currently being deployed internationally, for example, through the international DataCite and Orcid initiatives, in which the British Library is a partner. They are also being implemented by programmes funded by the UK's Research Councils, such as the IPCC data management infrastructure currently under development by the British Atmospheric Data Centre (a NERC data centre).. A coordinated, cross-council effort is needed. The international e-science review identifies the need to:

'*At every opportunity establish and support policies for openness: open-source code, open data, and open courseware. To the extent possible, these should be freely available with terms of use that encourage reuse. Work with international standards activities especially for interoperable data'.*

59. The e-Science review then points to the need to

'…consider a highly centralised, large data centre model for storing the information and preserving the bits, together with a distributed model for curation by disciplinary specialists... [and] plan for continued exponential growth in scientific data'.

> The vital importance of data-sharing can be illustrated by the following example: by analysing sequence data appearing on public databases in real time and posting resultant analyses on a wiki site, 4 research teams working across three time zones revealed the origins and evolution of the H1N1 virus just eight weeks after the first reported case of swine flu in Mexico.
>
> *BBSRC Strategic Plan 2010-2015*

60. Data sharing is not just a matter of data citation to give credit to data providers , It is also a matter of data quality and provenance for data consumers. Data curation practices and support services in different disciplines are a prime area for exchange across disciplines. However, such exchange requires knowledge exchange effort, specialisations of software and standards.

[11] From the keynote speech given by the Minister for Digital Britain, Stephen Timms, for the RSA/Intellect "Technology in a Cold Climate' symposium, 27 October 2009

61. While the UK is very well positioned with regard to its data depositories, particularly in the environmental, social and life sciences, there is still a need for infrastructure to enable data sharing and interoperability across all research disciplines. This will, in addition, help establish the 'fundability' of vast research data[12] preventing existing data from being 're-invented' and strengthening the provision and capacity for secondary data analysis and exploitation. This implies a need for a national network of data depositories with common sets of standards and services. These would provide appropriate, coordinated support and promotion for data sharing and interoperability, identification of key community datasets and coordinated models for their sustainability and long-term storage, the creation and adoption of standards, capacity building in data management and many other relevant issues at both a strategic and operational level. The UK has key skills in the area, for example, NERC has recently embarked on an ambitious 3-year programme to move towards having the data, systems and services to support this vision of interoperable data - for the environmental sciences. This strong UK skills base means that we are well positioned to move forwards with e-infrastructure investments in this area.

4. TOWARDS A CONCLUSION

62. A well integrated e-infrastructure has the potential to dramatically increase collaboration, sharing and reuse across the research community, driving gains in research productivity and innovation thus fostering economic development and competitiveness of the UK. The UK's current, advanced position compared to other nations, should not be taken for granted. There is a need for more integrated e-infrastructure across disciplines and sectors within national boundaries but also across national boundaries. Research and innovation is a globally undertaken enterprise and to be competitive there is a need to think strategically about the development of UK's e-infrastructure as a national and global asset that gives our country both privileged position, and responsibility as a global leader.

63. The success of the UK e-science programme was largely built on the joint coordinated effort of various funding agencies. This secured UK's leadership position, as recognised in the international e-Science review. The programme demonstrated why a coordinated effort is important and how it can be achieved. A strong recommendation that followed was to *'continue funding policies that strongly encourage or require the creation and adoption of shared e-infrastructure.'* It suggested not only cost-effective and energy-efficient use of e-infrastructure, but also facilitation of intellectual interoperability between disciplines to jointly meet various research challenges. This demanding task is achievable provided an agreed mechanism for coordination can be put in place to ensure targeted and co-ordinated investment that reduces ongoing costs through efficiency savings and implementation of 'invest to save' approach.

[12] One of the challenges we face is the unprecedented growth in volumes of data. This is a common challenge across the globe. For example, in the USA volumes of data being generated are predicted to be reaching 1000 Petabytes in 10 years by the US Department of Energy. This brings not only the challenge of storing data, but to a large extent the challenge of finding data that could be reused for secondary analysis.

64. In this report we have taken the recommendations of the e-science review and findings from a range of other studies a step further by defining a co-ordination structure for taking forward the areas of action. We believe that without taking the right steps now, there is a risk of not just losing our leadership position, but also jeopardizing our capability to collaborate in addressing global challenges.

Annex 1

MEMBERSHIP OF THE EXPERT GROUP

1. Professor Carole Goble, University of Manchester, Chair
2. Shearer West, AHRC
3. Amanda Collis and Paul Gemmill, BBSRC
4. Richard Boulderstone, British Library
5. Lesley Thompson, EPSRC
6. Jeremy Neathey, ESRC
7. Malcolm Read, JISC
8. Tony Peatfield, MRC
9. Mark Thorley, NERC
10. Michael Jubb, Research Information Network
11. Neil Geddes, STFC
12. Kristine Doronenkova, ESRC, secretary to the Group

Annex 2

BACKGROUND TO THE REVIEW

The *Science and Innovation Investment Framework 2004-2014* highlighted the Government's commitment in working with interested funders and stakeholders to consider the necessary UK e-infrastructure. The OSI report *Developing the UK's e-infrastructure for science and innovation* (hereafter OSI report), published in 2007, documented a substantial assessment of the *status quo* and developments required for taking forward the UK's e-infrastructure for science and innovation.

The OSI report outlined a broad vision for the UK's e-infrastructure for research to provide a vital foundation for the UK's research base. It acknowledged the increasing possibilities for knowledge support and the creation of wealth but also the need to cope with rapidly advancing technology to provide a sustained yet responsive foundation to contribute to the solutions of multiple problems of the modern interdependent world. There has since been significant progress across all areas identified in the OSI report[13]. However, it was acknowledged that there is a need for a progress review that would take account not only the developments within the recommended areas, but will have a broader view on the developments related to e-infrastructure that are not mentioned in OSI report.

E-infrastructure must adapt to rapid developments in technologies, expansion of the knowledge base, continuous demand for data and new practices of scholarship and innovation. A number of major technical developments are not mentioned in the OSI report. For example, "cloud computing", "green computing", "Web 2.0", "carbon footprint", "software as a service", "open access repositories", "crowd-sourcing", and "semantic linked open data" are missing. Similarly, a number of strategic UK e-infrastructure programmes have gained visibility, including "data.gov.uk" and "Digital Britain". These issues are, however, widely acknowledged and taken into account in this report.

The evolution of cloud computing, for example, has the potential to radically change the delivery of IT services across a wide range of applications. It brings a paradigm shift in the approach to technology infrastructure, but also in information sharing making it ever more accessible. However, a number of questions remain around security and data protection in the provision and use of these services. JISC, and EPSRC all have cloud computing initiatives that are insufficiently coordinated.

Environmentally friendly and sustainable computing, often referred to as "green computing" or "green IT", is another important development that underpins the progress of the national e-infrastructure and is shifting the balance between centralised and distributed computing and data storage provision. It also has economic implications as it strives to achieve improved systems performance and use, including the economics of energy efficiency. While progress has been made

[13] This, for example, includes the development and use of virtual research environments (VRE) by researchers in a number of disciplines; the improved access to research outputs by the further development of institutional and subject-based repositories; the rising awareness and support of the open access (including deposit mandates) and open source movements through-out the research lifecycle.

by individual data centres by the use of virtualisation and other energy reduction measures these remain ad hoc, local and uncoordinated achievements.

To start the review process a light-touch evaluation of progress completed since the OSI report was undertaken by the Research Information Network (RIN) and completed December 2009. The evaluation included a series of specialist interviews and consultations with experts and stakeholders and identified several key areas, largely interlinked, to focus future efforts and scarce resources. The findings of the RIN evaluation were further discussed by the Expert Group comprising representatives of the Research Councils, JISC and British Library. This report is a joint outcome. While not exclusive these areas are identified as areas of key importance for advancing national e-infrastructure. The reported general evaluation of the UK e-infrastructure is that it provides a solid base for further developments.

In 2009 the RCUK commissioned an International Review of e-Science which reported in February 2010 (RCUK e-Science review)[14]. This review was a thorough, comprehensive and substantial assessment of e-Science, e-Research and e-Infrastructure in the UK, calibrated against the international state of the art. Consequently, we draw on the findings and recommendations of the review panel's comments and recommendations.

There are other reports and evidence which has also been drawn upon during the preparation of this report. – This includes, for example, Century of Information Research[15]; *Towards 2020 Science*[16]; UK Strategy for Data Resources for Social and Economic Research 2009-2012 (The National Data Strategy)[17]; Strategic Plans of the Research Councils[18] and JISC Strategy[19]; US Report of Interagency Working Group on Digital Data to the National Science and Technology Council[20]; US National Science Foundation reports on Virtual Organisations[21] and

[14] The RCUK e-Science review was conducted by an international panel of 15 experts in the latter part of 2009. It reviewed a substantial evidence base of documentation and conducted a week-long programme of presentations, interviews and visits covering 60 projects in December 2009.

[15] Century-of-Information Research (CIR). A Strategy for Research and Innovation in the Century of Information, Paper was prepared by the e-Science Directors' Forum Strategy Working Group and abridged by Professor Malcolm Atkinson and Professor Paul Jeffreys. Paper was first presented at the Conference 'Oxford eResearch 2008, 11-13 September 2009, University of Oxford; *Promethus, Vol 27, No.1, March 2009*

[16] Microsoft report *Towards 2020 Science* sets out the challenges and opportunities arising from the increasing synthesis of computing and the sciences. It also seeks to identify the requirements necessary to accelerate scientific advances. Available at http://research.microsoft.com/en-us/um/cambridge/projects/towards2020science/

[17] The Strategy sets out priorities for the development of research data resources both within and at the boundaries between the social sciences and other areas of scientific enquiry. More information is available at http://www.esrcsocietytoday.ac.uk/ESRCInfoCentre/nds/

[18] Each Research Council has a vision or strategic plan which sets out the Council's overarching research aspirations and priorities over a 5 year plus period. Strategic Plans of the Research Councils can be accessed at http://www.rcuk.ac.uk/aboutrcs/operation/strategies.

[19] JISC Strategy seeks to emphasise activities designed to bring benefits to the education sector in the short-term, while at the same time maintaining investment in those projects with mid- and long-term benefits is available at http://www.jisc.ac.uk/aboutus/strategy/strategy1012.aspx

[20] 'Harnessing the Power of Digital Data for Science and Society', September 2008. The report advocates the creation of a comprehensive framework of transparent, evolvable, and extensible policies and management and organizational structures that provide reliable, effective access to the full spectrum of public digital scientific data

[21] Beyond Being There: A Blueprint for Advancing the Design, Development, and Evaluation of Virtual Organisations; Final Report from Workshops on Building Effective Virtual Organisations, May 2008.

Cyberinfrastructure[22]; Understanding Infrastructure report[23]; EU e-IRG reports[24]. This list is not exhaustive, but provides an overview of the sources that informed the work on this report.

This report sets out to review the OSI report and the recommendations of the RCUK e-Science review and other evidence in order to propose a course of action for the development of the UK's e-infrastructure. The agencies specifically addressed are: the Research Councils and JISC. The report also seeks to build a forum for cooperation and partnership with other stakeholders, notably: other major research funding agencies such as the Wellcome Trust and TSB; industry; government departments; the HEI-funding bodies; representatives of the scientific community and individual HEIs and institutes. These are referred to as "strategic partners" throughout the report and action plan.

The OSI report and many others introduced a number of recommendations to ensure the development of the UK's e-infrastructure, but failed to propose a prioritised, concrete plan for driving those recommendations forward. This report proposes new structures and key areas for action to guide further developments, agency policies and cross- and within agency strategic planning.

---

[22] Revolutionizing Science and Engineering through Cyberinfrastructure; Report of the National Science Fou8ndation Advisory Panel on Cyberinfrastructure, February 2003, available at www.nsf.nsf.gov/oci; Cyberinfrastructure Vision for 21st Century Discovery, National Science Foundation, Cyberinfrastructure Council, March 2007, available at http://www.nsf.gov/pubs/2007/nsf0728/nsf0728_1.pdf

[23] Understanding Infrastructure: Dynamics, Tensions, and Design; Report of a Workshop on "History and Theory of Infrastructure: Lessons for New Scientific Cyberinfrastructures', January 2007; more information is available at http://www.si.umich.edu/InfrastructureWorkshop/

[24] For example e-IRG White Paper 2009 – a live document summarising on-going discussions around key e-Infrastructure areas and topics that require immediate policy actions; e-IRG roadmap 2009 (consultation paper).

Annex 3

Major Conclusions and Major Recommendations of the RCUK International Review of e-Science

Major Conclusions

The e-science movement has emerged from a combination of push and pull. It is propelled by the push of the exponential growth of ICT capabilities, coupled with the pull of the demand for transformative tools and methods now needed to support the complexity, diversity and integrative needs of modern and future scientific research. The fundamental goal of the global e-Science movement is to determine how to use ICT as a foundation (as e-infrastructure) for transformative enhancement of the *doing* of research, in ways that create more transformative benefit from the *results* of research. E-Science is about transforming knowledge discovery in science; it is about innovation to support innovation.

The path linking research, knowledge production, and innovation is complex and nonlinear, but investments in e-Science are critical for expanding the knowledge creation that lies at the heart of innovation. Innovation is fundamental to advancing economic and social well-being. Furthermore, investments in e-Social Science offer the potential to better understand those pathways and to fully turn the UK's investments in basic research into advances in innovation and economic prosperity, including high wage jobs. The e-Science Programme can also be a pilot to show the way towards ICT-enabled environments for more effective and inclusive, anytime and anywhere, life-long learning.

The technologies and practices of e-Science, together with the e-infrastructure on which it rests, must be both a topic as well as an enabler of research and development; and this duality needs to be made synergistic. E-Science as a topic of research includes both technological and social (behavioural, economic, legal, ethical) dimensions. As an enabler of research it requires establishing and nurturing mutually-beneficial relationships between those skilled in design and evaluation of e-science environments, those pushing the edge using these environments, and those providing operational services and training. It is intrinsically an interdisciplinary, multi-role team effort.

The Panel has concluded that the UK e-Science Programme is in a world-leading position along the path described in the title of this report: *Building a UK Foundation for the Transformative Enhancement of Research and Innovation*. The investments to the present are already empowering significant contributions to well-being in the UK and the world beyond. The UK must now decide whether to create the necessary combination of financial, organisational, and policy commitments to capitalise on their prior investments, and to move to the next phase of building capability, growing adoption and achieving competitive advantage.

The successful creation and adoption of e-Science is an organic, emergent process that requires ongoing, coordinated investment from multiple funders together with coordinated action from multiple research and infrastructure communities. It requires nurturing robust infrastructure and a continuous cycle that couples research, application development, and training processes. It is the balance between these processes that drives success in e-Science. None of this is easy, but the rewards for success are enormous.

Major Recommendations for Action

The previous sections of this report contain interwoven findings and recommendations, as well as direct responses to the questions in the evidence framework in Section 4. We have given both a retrospective assessment in Section 2 as well as our vision of opportunities and challenges for the future in Section 3. The UK e-Science community has also given the RCUK the benefit of their assessments and advice for the future of the e-Science Programme in Section 4 of the Evidence Document. All of this, we hope, will be studied carefully by the e-Science community and the relevant funders and be helpful for plotting the future of e-Science in the UK.

We conclude this report with a list of a dozen major recommendations for action at a more general level than in earlier sections. The emphasis here is on *what to do,* rather than being very prescriptive about *how to do* it. These are not in a priority order, but most are quite interdependent.

1. Structure and leadership
Establish organisation and management structures that continue to treat e-Science as a designated strategic initiative spanning all Research Councils and having ongoing designated funding. Provide high-quality dedicated leadership for a strategic e-Science Programme, and provide the leader with adequate authority and resources to catalyse real synergy within and between funder, researcher, and service providers. The leader needs resources to co-fund with specific projects funded by the individual Councils. This recommendation includes exploiting more systematic, coordinated investments in the infrastructure, development, and adoption of e-Science between several components of BIS, including the Research Councils, JISC, the Technology Strategy Board, and the funders of higher education and facilities. E-Science Programme leadership should also seek coordination with e-Science-related funding from outside of government including the Wellcome Trust, the EU, and others.

2. Industry-academic collaboration
Establish more systematic and better supported mechanisms, including targeted funding, to nurture collaboration and bi-directional knowledge transfer between academia and industry in the creation, provisioning, and application of e-Science. There are multiple goals here: (1) to identify better ways of connecting academia and industry through e-Science to accelerate the transformation of research outputs to beneficial innovation; (2) to enlist relevant industry in the creation and perhaps provisioning of the e-infrastructure platform for e-Science; and (3) to help industry adopt and tailor best practices and services from e-Science to enhance their own productivity.

3. RCUK e-Science Centre network
Sustain and strengthen the RCUK network of e-Science Centres. Challenge and support these Centres to both serve their regional constituencies and be members of a network for the common good with others in the UK and international partners. Establish the remit of various centres in a way that stresses complementary expertise and sharing. Emphasise and assess expectations that the Centres be proactive in engaging with each other, with UK industry, and with other countries. Establish polices for sustaining this network but in a way that periodically (e.g. every

five years) requires re-competition to enable new entries into the field and to retire less effective activities.

4. Sustaining advanced e-infrastructure
Sustain the operational e-infrastructure for e-Science created to the present. Informed by an ongoing review of future needs, evolve it to: (1) higher capacity; (2) more complete function; and (3) leading-edge, distributed system architecture. E-infrastructure, as we are using the term, includes the hardware, software, organisations and people to provide generic but tailorable services such as networking/communication, computation, visualisation, data repositories, digital libraries, online observatories, sensor networks and instruments, and distributed (virtual) collaboration. Middleware provides the glue to integrate these services in secure ways. In particular the UK needs to invest in e-infrastructure in ways that (1) anticipates the continuing exponential growth in scientific data and the increasing ability to extract knowledge from it; (2) provides the UK research community access to petascale-level computing and anticipates the future needs for access to exascale; (3) continues to build on and strengthen the grid model of distributed computing, but also explores the adoption of emerging models of cloud computing for research. Scaling scientific computer codes to the peta- or exa-level is generally a major challenge that would require R&D support.

5. Supporting complementary roles
Recognise in programme calls and funding policies that there are people in several complementary roles that need to be funded in a balanced way. There are (1) researchers seeking to innovate in the application of e-Science methods; (2) researchers in computer science and some aspects of social science contributing to designing better e-Science methods and services; (3) professional software engineers or informatics specialists who build reliable production-grade systems; and (4) professionals who administer and operate the supporting e-infrastructure. There are people effectively spanning several of these roles. There may be the need to better define and reward new professional identities and job types within academia that span role 2 and 3 above.

6. Sharing for cost and science effectiveness
Continue funding policies that strongly encourage or require the creation and adoption of shared e-infrastructure. This is important not only from a cost-effective, efficient-energy use, and environmental-impact perspective, but also for facilitating intellectual interoperability between disciplines, institutions, facilities, and data resources essential for many grand-challenge research endeavours. Since scientific research is intrinsically global, place great emphasis on creating UK e-infrastructure that harmonises with the e-infrastructure in other countries. Doing so will enhance the sharing of data and unique, expensive instruments and will reduce constraints on collaboration at a global scale.

7. Role for arts and humanities
Encourage and support even more participation of the arts and humanities research communities in the e-Science programme. (We saw some excellent beginnings in our review.) Arts and humanities are poised to achieve large benefit from e-science methods and infrastructure as the human record becomes increasingly digitised and multimedia. For example, a field called "corpus computing" is emerging due to the ability now to compute across enormous collections such as those being created by industrial partnerships with academic libraries. Copyright management for such work

will remain an issue, but for public domain and open-license materials, the field is now wide open.

8. Role for social sciences
Building on a strong start, encourage and support even greater leadership by the social science research community in the adoption of e-Science methods particularly, for example, given capabilities to explore enormous data sets, analyse social networks, and explore very complex systems through simulation and modelling. The social science community should also be encouraged to contribute more to deeper understanding of more principled ways to design effective virtual research environments, collaboratories, and four-quadrant environments (see last section below). Many science communities are creating such distributed knowledge communities, but many are sub-optimal or outright failures, usually for social and behavioural rather than technical reasons. If research is conducted within such technology-mediated environments, we can potentially capture and later mine not only the artefacts of knowledge work but also the processes.

9. Crossing the chasm; refreshing innovation
Develop a dual strategy that both (1) accelerates the adoption of e-Science methods in the "mainstream market" of researchers as discussed in Section 3 ("crossing the chasm"); and (2) refreshes the investments in the "early market" to produce the next wave of innovation in e-Science services and application. Goal (1) involves training, making services available, and tailoring current services to more specific needs of disciplinary and interdisciplinary communities. This process needs an integrated formative assessment activity that includes continuous monitoring and that helps ground and inform the activities of "early market" communities. The assessment should inform a spiral, iterative design process. Also include special opportunities for upper management of universities, government, and industry to learn more about the fundamentals and strategic importance of e-Science and e-infrastructure.

10. Data stewardship at enormous scale
Continue the strong focus on creating practices and services for appraisal, curation, federation, and long-term access to scientific data. Complement or broaden the activities of the Digital Curation Centre with the creation of coordinated and sustained production services for curation and stewardship of scientific data. Consider a highly centralised, large data centre model for storing the information and preserving the bits, together with a distributed model for curation by disciplinary specialists. The academic digital libraries centres might be encouraged to assume some major responsibility for scientific data. Seek and promote international cooperation. In all of this, plan for continued exponential growth in scientific data.

11. Openness as a general policy
At every opportunity establish and support policies for openness: open-source code, open data, and open courseware. To the extent possible, these should be freely available with terms of use that encourage reuse. Work with international standards activities especially for interoperable data.

12. Towards functionally complete, four-quadrant, research environments
Place greater emphasis on the overarching goal of establishing capacity for collaborative, international, interdisciplinary team science to occur routinely in "functionally complete, four-quadrant environments" built upon e-infrastructure. "A four-quadrant environment" refers to a blended virtual-physical environment in which

the activities of a group can flow easily between all four quadrants in a 2-by-2 matrix with same versus different for the two dimensions of both time and place. It subsumes the concept of virtual research environment. “Functionally complete” means that the environment supports access to all the people, the information and data services, the observatories and facilities, the computational services, and the collaboration and communication services necessary for a scientific team (or more generally, a community of practice) to carry out its work. Such environments could become both necessary and sufficient for participation in a research endeavour. They could accelerate and broaden participation in scientific discovery and learning. They offer the potential to support both explicit and tacit knowledge creation, and thus to support a blend of learning *about* science, learning *to do* science, and learning *to be* a scientist.

Annex 4

DIRECTOR E-INFRASTRUCTURE – JOB SCOPE

Director

Key activities:

1. The Director will be responsible for providing the strategic leadership for the development of the UK research e-Infrastructure.
2. The Director will draw advice from the e-Infrastructure Strategic Board, of which they will be a member.
3. The Director will be responsible within ToR for developing and securing funding for the 'value add' actions of e-Infrastructure. The business plan should have detailed plans for the CSR position and indices plans for the following period.
4. The Director will report to the nominated Chief Executive of the RCUK or nominated CEO deputy. Within an operational framework agreed by the Chief Executive the Director will have considerable freedom to operate to best effect.
6. An administrative and technical team will support the activities promoted by the Director. The Director will have overall management responsibility for the team and its operation consistent with the normal accounting and staff arrangements in the host institution:
7. Attributes required:
    - Build global links so the UK is able to pursue quality
    - Ability to rapidly gain respect required to fulfil the role
    - Ability to operate effectively both through influence as well as directive mode
    - Strong technical expertise and the ability to quickly assimilate technical issues outside current knowledge
    - Ability to think laterally and work through issues to reach the desired solution while gaining the support of others
    - Ability to travel extensively
    - Advanced communication skills oral and written
    - Track record of building and maintaining strategic partnerships with many kinds of stakeholders.

e-Infrastructure Unit

To provide the appropriate administration and technical support to allow the Director to fully meet their role and responsibilities.

Key activities:

- Supporting the Strategic Board through generation of papers, taking minutes and arranging meetings
- Putting together authoritative reports that can be used by others to adopt effective cross talking e-Infrastructure
- Running the annual e-Infrastructure meeting
- Gathering current information together so a full picture of UK and e-Infrastructure is maintained so investments are made in the right context.

- Providing information to the community, councils and strategic partners. Responding quickly and comprehensively to enquiries – the unit is a front door to UK e-infrastructure.
- Be responsible for liaison with e-infrastructure related activities in the funding agencies
- Budget management

Annex 5

Spending of the Research Councils and JISC in support of their e-infrastructure:

Agency	Estimate p.a.	Form
AHRC	£4m	Capital funding for "Digital Equipment and Database Enhancement for Impact"
BBSRC	£38.4m	Large facilities and software resources primarily funded though Bioinformatics and Biological Resources Fund (BBR)
MRC	£ 10.8m	Research facilities and resources, e.g. Data Support Service, High-Throughout sequencing hubs, Methodology hubs
ESRC	£24.5m	Research facilities and resources, e.g. data services, large datasets, research methods (capacity building element)
EPSRC	£21m	Including annual costs for HECToR and other e-infrastructure, as well as support for researchers and research students through grants who use e-infrastructure
NERC	£10m	To support its network of environmental data centres
STFC	£5-10M	Computing infrastructure and data centres.
JISC	£60m	Including JANET (£35m) that meets the broader needs of the HEIs, not only research.

Annex 6

<u>Action Areas and indicative actions for *Delivering the UK's e-infrastructure for research and innovation*</u>

There are five broad, interconnected areas where there is a real opportunity to add value to the UK's e-infrastructure through dedicated leadership, stronger co-ordination and further strategic investment.

These form the basis of a future action plan. The areas are:

- **Sustainability** of new and existing mission critical e-Infrastructure supported by the research funding agencies;
- **Interoperability** between new and existing e-Infrastructure supported by the different research funding agencies;
- **Adoption** of e-Infrastructure into mainstream use by a majority of researchers by support from the research funding agencies;
- **Capacity** building of the skills base for creating and using e-infrastructure;
- **Data** - e-Infrastructure for data to serve data intensive research.

The areas are closely interlinked with a degree of overlap:

- Infrastructure that is adopted by a user base has a case for sustainability;
- Infrastructure that is interoperable with pre-existing and used infrastructure is more likely to be adopted;
- Infrastructure that is sustained is more reliable as an adoption choice and worth investing in interoperability measures;
- Infrastructure that has community capacity to create and use it is more likely to be adopted.

Data cuts across all these areas for action as there is real need to improve coordination, sustainability, interoperability, adoption and capacity to fully harness and exploit the exponential growth in data in order to support high quality scientific research.

An indicative action plan for the Research Councils and JISC, as suggested by the Expert Group, is presented here. Similar action plans for other stakeholders, such as funding councils and TSB, would be developed in partnership with those stakeholders by the Director.

The actions are intended to be taken by each Research Council and incorporated into their base operating procedures. They are also intended as an action plan for cross council interactions. They are also viewed intended to be an indicative action framework for the Director.

Area	**Principles**	**Examples of actions**
1. Sustainability of e-infrastructure	Need for appropriate suite of sustainable, compatible and modern e-infrastructure	To develop a strategy for different requirements and models for sustainability of: - services - software - data sets - hardware
2. Interoperability between e-infrastructure	Need for interoperability between existing and new e-infrastructure	To agree appropriate core e-infrastructure in different scientific domains that should interoperate and who is responsible for funding this.
3. Adoption of e-infrastructure	Need to: - maximise the up-take and use of e-infrastructure. - ensure that e-infrastructure is prepared for adoption from its inception and throughout its development. - promote the availability and value of e-infrastructure.	To examine the barriers to adoption and bridging the ‘adoption chasm’ through community-lead targeted developments and software hardening like the JISC ENGAGE programme and the EPSRC’s Software Sustainability Institute. To set up a reward structure for the adoption and adaptation of existing infrastructure to maximise usage of current and future investments. To develop standards used in peer reviewing of new e-infrastructure related projects, ensuring that all of them include clear and concrete pathways for adoption and exploitation from the outset.

Area	Principles	Examples of actions
4. Capacity building of the skills base.	Need to increase the capacity to develop, exploit and adopt sustainable and interoperable e-infrastructure	To agree on the set of skills required within research teams that would include skills related to research technology. To develop a plan for training, and capacity building, that would include focused promotion activities, to ensure the take-up of existing e-infrastructure beyond the early adopters. To create and maintain a professional career structure for highly skilled research technologists within Research institutions that recognises and values them.
5. Data	Need to - recognise the importance of data as a capital asset in all research disciplines. - support data intensive scientific research, recognising the importance of linking data and computation. - promote sharing and reuse of data within disciplines and as a cost efficient means of collaboration across disciplines. - promote an open access policy outside the scientific community and across agencies.	*Leadership and Coordination* To explore potential for a National Data Roadmap (similar to the UK Strategy for Data Resources for Social and Economic Research) to create a national relevant network of data depositories with common set of standards. *Sustainability* To develop a strategy for different requirements and models for sustainability of data sets, data services, data access/mining software and long term data storage. To make the e-infrastructure needed for large scale data archiving, large scale data networks and high throughput data processing part of the Large Facilities roadmap. To explore potential of cloud computing and services to support data management requirements *Interoperability* To set up mechanisms for enabling and sustaining better data integration of existing datasets in different scientific domains that should be exchangeable and adoption of commonly acceptable standards and mechanisms, and who is responsible for

Area	Principles	Examples of actions
	- develop plans for leadership, coordination, sustainability, interoperability, adoption, and capacity for research data.	funding this. To set up a mechanism for the adoption of existing metadata standards in existing data sets and the creation of missing metadata standards where necessary. *Adoption* To set up a mechanism for ensuring that all new data-related projects include specific pathways for data standardisation, adoption, sharing and sustainability at the outset and are rewarded for producing reusable datasets. To introduce recognition and reward structures for data reuse and data sharing in order to promote a culture of data and information sharing within and across research communities; specifically to establish data citation and attribution mechanism for scientific data; and to extend legal deposit legislation to digital data/digital material. *Capacity* To develop a plan for strengthening the research capacity to produce and consume data in the most effective way through establishing a mechanism of training for young and mid-career researchers and research technologists in data management, including the professional training of curators and information specialists. To create and maintain a professional career structure for data curators and information specialists within Research institutions that recognises and values them.