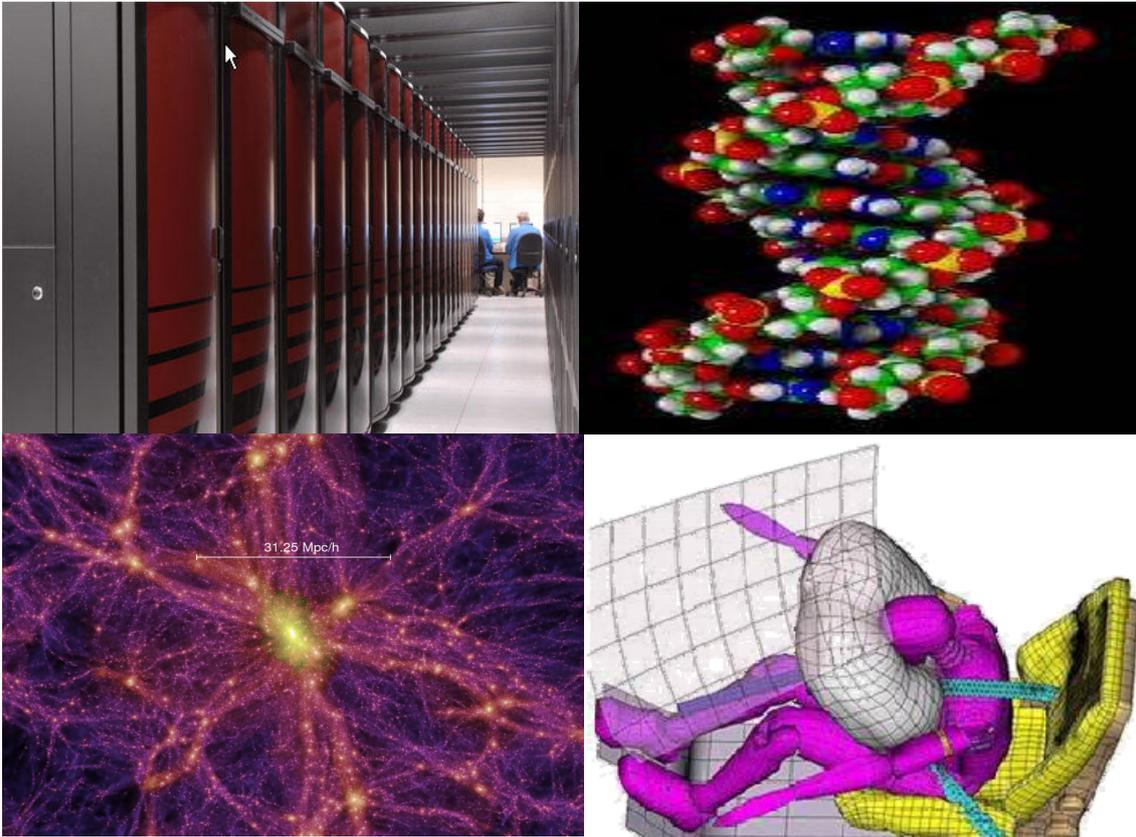# A Strategic Vision for UK e-Infrastructure

A roadmap for the development and use of advanced computing, data and networks

Chairman's foreword

On the 13th July 2011, David Willetts, BIS Minister for Universities and Science, invited academics, industrialists, hardware and software suppliers and experts from the Research Councils to discuss the establishment of an e-infrastructure for the UK. That meeting confirmed that we are experiencing a paradigm shift in which the scientific process and innovation are beginning in the virtual world of modelling and simulation before moving to the real world of the laboratory. This is a shift in which sophisticated analysis and visualisation software are being used to mine massive amounts of experimental data from the life and environmental sciences to uncover new hypotheses and trends. The meeting also concluded that to exploit this revolution we would require a fresh, collaborative approach to software development to bring scientific, industrial and public sector users and hardware and software developers and vendors closer together. This in turn means that the advanced computational capacity of the UK needs to be addressed as a system – an e-infrastructure.

I was commissioned by the Minister to write a report on how we might create an e-infrastructure that would support a strong public-private partnership. I would like to acknowledge the support of those who have given freely of their time for their hard work over the last two months to help me write this report. We have been able to move quickly by building on some excellent analysis. I think particularly of the work of the academic community summarised in the report by Peter Coveney *Strategy for the UK Research Computing Ecosystem[iv],* and the work of the Research and Funding Councils in preparing the *Report of the e-Infrastructure Advisory Group* [viii].

Drawing on these consultations, we have proposed a series of recommendations around software, hardware, networks, data storage and skills which will help to develop the e-infrastructure. The Government has energised these proposals by offering a significant additional sum above their base-line funding to stimulate the partnership in this space. We are recommending the establishment of an E-infrastructure Leadership Council to advise Government on the development of the e-infrastructure and to develop a wider plan for stakeholder engagement.

Professor Dominic Tildesley
Vice President Discover Platforms
Unilever Research and Development
November 2011

Table of Contents

Executive Summary

The objective of this report is to present a ten-year strategy for the development and management of the UK's e-infrastructure. Such an infrastructure, comprising networks, data repositories, computers, software and skills is an essential tool for research and development both in industry and across a wide range of fundamental science. There is a clear correlation between investment in such infrastructure and long-term growth. Pilot studies suggest that for every £1 of public spending in the industrial exploitation of e-infrastructure, £10 of net GVA will be generated within two years rising to £25 after five.

The UK is host to world-leading industry in the automotive and aerospace, pharmaceutical, fast-moving consumer goods, process industries and digital media sectors. All of these large corporations and the related SMEs could, and want to, make better use of e-infrastructure in driving their growth. In addition, the rate of technical innovation in e-infrastructure enabled science, e-science, is accelerating with the development of petaflops machines, petabyte storage and new mathematical methods and algorithms for simulation and data analysis. At first glance the UK would appear to be well placed to exploit this revolution. However, unless a coherent strategy is put in place, the potential benefits will not be realised. This strategy needs to include a financial plan and governance model for a national e-infrastructure which allows industry, academe, the charitable sector and government to work in a partnership for growth.

The UK Government is currently spending approximately £200M p.a. on aspects of e-infrastructure to support the academic community. Government has recently allocated an additional sum of £145M to enhance this infrastructure so that it can be extended to create the partnership while ensuing that our academic research in this field stays at the leading edge. The development of software will be at the heart of this new partnership. The investment will help to develop codes based on new ideas for the academic and industrial communities, will port legacy codes to high end machines for advanced simulation and data-intensive computing, will help to make academic software more robust, and will support the transfer of important commercial codes to high end machines.

To ensure the development of the strategy, we propose the following:

- Implement an agreed and costed strategy that recognises a balanced need for both capital and pre-determined recurrent resources, using the established base-line and additional Government funding to build the partnership. Increase private sector funding over a ten year period.

- Create specialist centres to develop robust software of value to academy, industry and government. Although this is only one element of the infrastructure, it is critical to its success.

- Embed training in e-science in all postgraduate training, including in successful Doctoral Training Centres. Use these modules to offer training to industry where appropriate. Support the publication of reports, technical workshops and conferences for the community funded by this programme.

- Set up a single coordinating body that owns this strategy and can advise BIS Ministers on its implementation and development. This E-infrastructure Leadership Council should be co-Chaired by the BIS Minister for Universities and Science and an industry representative.

- Support the Council with an independent E-infrastructure Secretariat providing the Council with information and costed proposals based on a wide consultation with academic, industrial and government partners.

The strategy outlined in this document is the first step. We recommend that the E-infrastructure Leadership Council, working through the E-infrastructure Secretariat, should develop a wider stakeholder engagement plan for the e-infrastructure community across academe, industry, the charitable and public sectors to ensure that these recommendations are implemented and developed in a way that reflects the ubiquitous nature of this technology.

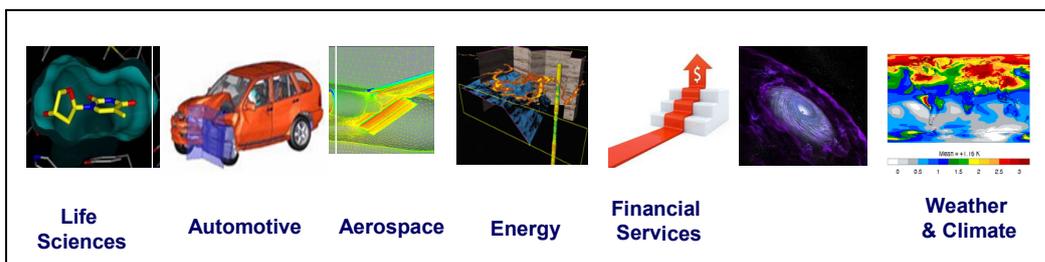# The opportunity of scientific and industrial growth

Today, to Out-Compute is Out-Compete [i]

Introduction

Data generation and analysis using computational methods are at the heart of all modern science and technology. The explosion of new data comes from two directions: cutting-edge experiments, for example nucleic acid sequencing, the Large Hadron Collider and space telescopes; and from *in-silico* simulation and modelling across all length and time scales, for example protein modelling, and the optimisation of manufacturing processes.

This discipline of using digital methods for generating ideas and knowledge from data, often referred to as e-science, is a key enabling technology for the advanced nations in the 21st century. The USA, China and Japan continue to invest huge sums on creating the necessary infrastructure to support these developments because of positive return on investment in terms of growth in the industrial and commercial sectors, and the effect of e-science on healthcare, transportation, renewable and clean energy and climate modelling.
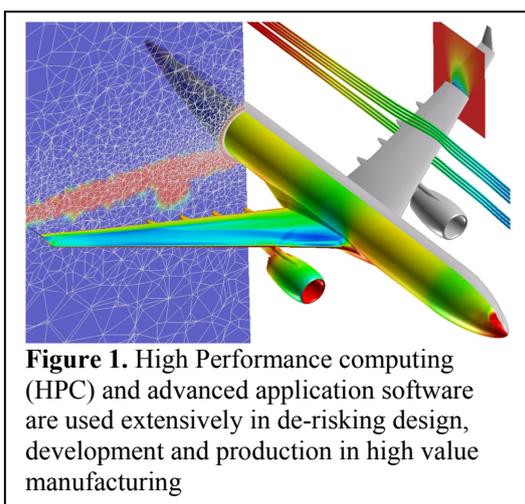
The applications of e-science are ubiquitous[ii]. E-science is used in the pharmaceutical industry for ligand docking for drug design; in the automotive industry for modelling aerodynamic drag reduction; in aerospace towards simulation of a full plane in flight; in the energy arena in the determination of the structure of oil bearing rocks by analysis of reflected seismic waves; in financial services in the prediction of stock trends; and in weather and climate modelling to give more accurate severe weather warnings and reduce uncertainty in climate change prediction.

| Life Sciences | Automotive | Aerospace | Energy | Financial Services | | Weather & Climate |

In this report we use six application areas to illustrate the importance of e-enabled science and innovation.

1. E-science in the aerospace and automotive industry

High performance computing (HPC) refers to the use of leading-edge computers for simulation and modelling and for advanced data analysis. The application of HPC in support of reducing technology and engineering risks is fundamental to high value manufacturing in the UK, and especially in the aerospace and automotive industries.

Figure 1. High Performance computing (HPC) and advanced application software are used extensively in de-risking design, development and production in high value manufacturing

Both industries face significant yet different challenges in their need to optimise, both product development and time to market, by taking manufacturing from the physical world into the virtual world. To optimise the design and development cycle the UK must increasingly be able to prove products in the virtual environment to enable truly concurrent design, development and verification. UK manufacturing depends upon the integration of its global supply chain elements on which their global advantage depends. This requires

6

instantaneous and seamless access to the latest HPC capability. This technology and supporting e-infrastructure is a national resource that enables competitive UK manufacturing.

In aerospace, worldwide competition demands faster time to market at lowest cost for the most sophisticated and efficient airframes and propulsion. In order to meet these challenges the industry, and the UK is a global leader here, is very rapidly and increasingly employing "virtual prototyping" to reduce product development risk and to evolve new manufacturing processes. The trajectory is for physical testing to be all but eliminated until the final verification phases. This approach places enormous demands on e-infrastructure within the enterprise, but it is essential if the UK is to remain at the top table in global aerospace.

The UK's automotive industry is enjoying a resurgence, largely due to its capabilities in the advanced technology domains, in premium vehicles and in the rapidly developing arena of motor sports, one of the UK's prime seedbeds for automotive technologies. In both of these facets of the automotive industry, the UK has been and is growing as a significant global force, commanding substantial inward investment to the UK. The UK's position currently draws heavily on HPC, supporting infrastructure, and the expertise to use these systems. To remain a leading player, and to capitalise on the position the UK currently has, the industry needs to draw on an e-infrastructure provision that will out-strip its global competitors.

2. E-science in health and the pharmaceutical industry

E-science is central in enabling many of the core processes in pharmaceutical R&D, from sequence analytics and target selection, high throughput chemical screening, ligand docking and protein structure characterisation, predictive safety model construction, drug metabolism modelling through to the molecular characterisation of disease states. To date, the scale of these analyses and the throughput of the assay technology has meant internal infrastructures have sufficed. However due to significant changes in the pharmaceutical R&D model and rapid advances in key assay technologies, to be competitive, the industry needs to consider alternative models.

The drivers influencing change in pharmaceutical requirements are:

- The shift to a personalised medicine strategy in drug R&D: improvements in drug safety and efficacy achieved through the provision of individually tailored therapies using advances in knowledge about genetic factors and biological mechanisms of disease coupled with individual's patient care and medical history. Many drug R&D programmes now have associated biomarker discovery and patient stratification studies, exploring patient and animal specific drug response, metabolism and safety. Typically such studies involve the molecular profiling (e.g. next generation sequencing, transcriptomics) of individual clinical and *in vivo* samples with projects capable of producing terabytes of primary data for relatively little cost (e.g. the $10k/human genome, with $1k soon to be achieved for the sequencing). With such data volumes only set to grow exponentially (or more) and an increasing need to manage, integrate and analyse these data as part of the core R&D process, there is a significant growing demand for advanced HPC solutions.

- Modelling and simulation: the use of predictive technologies to guide bench research and shorten cycle times, at all stages of R&D. These include virtual high throughput screening, predictive safety pharmacology models, predictive pharmacokinteic/pharmacodynamic modelling and systems biology approaches to modelling disease/drug interventions. As the scale and diversity of the data grow this type of modelling will be increasingly important for representing complex systems and predicting behaviour. Such large scale modelling will have significant scale challenges for existing infrastructures requiring computational science research and possibly specialised heterogeneous hardware solutions.

- The move from a tightly integrated vertical R&D operating model to highly networked R&D organisations, partnering with numerous contract research organisations (CROs), academia, charities and biotechnology companies. Such partnerships enable drug R&D projects to harness external innovation and specialist skills and knowledge. However this networked model requires secure e-collaboration environments, allowing iterative, collaborative analyses over shared, often large scale, heterogeneous data sets.

- The explosion of publically available, pharma R&D relevant, massive, heterogeneous and complex biomedical and chemical data sets, such as the 1000 Genomes project or the Cancer Genome Project, or the many transcriptomic, proteomic and metabolomic datasets. The existing policy of internalising, integrating and analysing such public data within pharmaceutical firewalls is no longer feasible requiring access to public/private HPC facilities proximal to these emerging, R&D relevant data sets.

The next generation of pharmaceutical e-infrastructure therefore needs not only to facilitate large scale clinical and molecular analytics and modelling but also allow secure collaborative working over both public and proprietary data sets. To this end many pharmaceutical companies are exploring external HPC services (and research), including commercial cloud solutions, allowing flexible access to significant infrastructure as and when needed. However, such options raise other issues, for example data security (for commercial reasons but also importantly for ethical reasons deriving from the need to protect the confidentiality of the patient data), indemnity insurance, practicalities of moving extremely large data sets over remote networks, access to contextual knowledge (e.g. EnsEMBL, literature and proprietary knowledge) on remote clouds, etc. Despite these issues that are yet to be resolved, investment in UK scientific e-infrastructure research, and especially services enabling public-private research and collaboration (including enabling software platforms), will be of increasing benefit to UK pharmaceutical research and the wider UK biomedical research community over the next few years. Of particular interest will be the role of UK e-infrastructure in enabling a differentiated service, beyond that available from large scale commercial cloud suppliers, reflecting the specific needs of biomedical research.

3. E-science in the digital media

The entertainment and media industry is a $1.7 trillion market, and increasingly at its heart is computing power. Dreamworks, producers of *Shrek* and *Monsters Versus Aliens*, has revenues of $680M and considers HPC to be one its most vital strategic assets[iii]. The production engines of TV, films and gaming are making increasing demands on computer processing, particularly as they increasingly converge and share media assets. (48 frames need to be rendered for every second of stereo animation. As more that 200,000 frames are required for a typical film this would equate to tens of millions of CPU hours.)

The creative industries have always pushed the boundaries of hardware and software to ensure maximum 'realism' in representing the world. Currently, we are witnessing the development of ultra high definition television (with sixteen times the pixel resolution of current HD systems), with concomitant high resolution image capture, Stereoscopic 3D (S3D), and high quality sound capture. The BBC and NHK have already experimented with 8k video format resolutions. Images captured at 8k have 17 times the pixel count of the current HD 1080p format – currently marketed as Full HD. The 8k systems include 22.2 audio, which have three layers of speakers to produce a 3D Soundscape. As the higher resolution and frame rates become a necessity for production, HPC will be a vital commercial requirement in processing the images, audio and metadata if the UK is to maintain a leading role in high quality production.

The UK Visual Effects (VFX) and post-production industry is one of the top three in world, and to remain so it will have to be at the forefront of technological development, including HPC. Currently, VFX work is a very slow process for the creation of a high quality image. Off-line rendering takes an average of ten minutes per frame and can be longer if complex images are

created. Real time rendering, driven by HPC, will be a major commercial driver when film studios and production companies are choosing their post-production houses.

Game developers already use real-time rendering - even at TV HD quality - but to improve image quality and match those of film and future TV resolutions, these developers will need HPC - especially as content is transferred from one medium to another. The introduction of stereoscopic and auto-stereoscopic (glasses-free) 3D, particularly for live events, is currently hindered by the lack of good and reliable real time processing for the alignment of the left and right eye image channels, a task, again, that can only be resolved by HPC.

High end computing will reduce the time to market for producers, increase the quality of experience of viewers and gamers, and expand the market for VFX, post-production, content and entertainment services.

4. E-science in the bio-economy

The revolution in genomics will allow us, for instance, to breed plants from fully sequenced populations. Doing this requires significant computational analysis but will accelerate massively the genetic gain from breeding for traits such as yield, disease resistance. This is paralleled by the need to improve bioprocesses more generally. Here there is a combinatorial problem, in that changing 4 enzymes out of 1000 in a network (the size of a typical problem) can be done in 41 billion ways. This is far too many to experiment on, so we need computer models of the biological systems we wish to manipulate for biotechnology (as well as for biomedicine, see above) – so-called digital organisms. At the individual enzyme level, we shall also need to screen millions of variants for the 'directed evolution' (improvement) of enzyme properties. To exploit properly the methods of synthetic biology and metabolic engineering requires computational analyses that can learn and predict the sequence-activity relationships of such enzymes and the productivity of the organisms that contain them.

Professor Michael Bushell University of Surrey is undertaking an *in silico* study of lignocellulosic bio-fuel processes. This project, funded by BBSRC, which began in 2009, is intended to aid in the design of bio-refining procedures involved in the production of bio-fuels from lignin. The project is to provide a computer package that contains simulations of the metabolic networks of four potential lignocellulose-degrading micro-organisms and four potential producers of bio-ethanol.   Using this package, it will be possible to simulate a process using each possible combination of degrader and producer. Variables can be altered, including the composition of the raw materials involved, using linear programming to optimise metabolic activity for bio alcohol production. This will allow users to screen all possible combinations of lignocellulose degrading micro-organisms and producers of bio-ethanol using a particular raw material, optimise the bio-ethanol production and screen for the production of potentially valuable by-products. Current bio-refining process design relies on trial and error experimentation supported by limited, *ad hoc* process models. New modelling of the entire genome of micro-organisms offers the opportunity to exploit gene sequencing programmes to examine the entire theoretical repertoire of potential microbial species. This could lead to a new source of bio-fuels giving the UK a global market advantage in the increasing 'green economy'.

5. E-science in weather and climate modelling

In the 2008 case for the current Met Office supercomputer, a socio-economic investment appraisal showed that the investment would provide enhanced weather forecasts and climate services which would in turn contribute to delivering a net extra £0.5Bn of socio-economic benefit to UK society for a five year whole-life cost of about £50M.

The current supercomputer has been operational since 2009. For weather, it is now enabling small-scale high-impact events to be better forecast, with warnings given to the public and civil response authorities with more-accurate quantification of uncertainty at lead times adequate for actions to be taken to reduce impacts. A secondary benefit is that continued incremental year-on-year improvements are being made in the quality of routine automated UK and global forecasts at all forecast ranges. These improved capabilities are delivering socio-economic benefit across a range of sectors including volcanic ash incidents, aviation safety and efficiency, wind-storm damage, flood damage, road and rail transport, maritime safety and utility company operations. For example, for one important sector, civil aviation, global GHG savings were indicatively quantified at 30 MtCO$_2$e (Megatonnes CO2 equivalent) over the five-year life of the supercomputer, of which 18 MtCO$_2$e were attributable to the enhanced supercomputing capability.

For climate research and climate services, the new supercomputer is allowing climate models to be run at higher resolutions. These models are better-representing key atmospheric, ocean and ice processes and better-capturing weather features. Biological and chemical processes are being included in models (especially the carbon cycle) to generate so-called 'earth system' models and seasonal and decadal climate forecasts are being developed. Socio-economic benefits arise because the new supercomputer brings bring forward the time when more-certain climate change predictions become available. This in turn is leading to reduced Green House Gas (GHG) emissions through better-targeted climate mitigation measures. In addition, adaptation actions can be more effective, so the impacts from climate change and natural variability, particularly in the next 2 to 3 decades, are reduced. Benefit is delivered, for example, through improved planning for costly, long-life infrastructure projects.

Currently, the Met Office has scientific developments available which would deliver further significant improvements in forecast accuracy. However, the current supercomputer has insufficient capacity to enable the Met Office to implement these developments in its operational forecast model and exploit the resulting improvements in forecast accuracy. With increased supercomputing capacity the Met Office would be able to improve short-range weather forecasts, long-range predictions, and climate change projections - all of which would deliver considerable further socio-economic benefit to the UK. For example, in May 2011, the House of Commons Transport Select Committee published its report *Keeping the UK moving: The impact on transport of the winter weather in December 2010*[iv]. The Committee recommended that further investment in Met Office supercomputing be made: "£10 million would be a small price to pay for improving the Met Office's long-range forecasting capability, given the cost to the UK economy of transport disruption due to severe winter weather". Such investment would provide improved evidence and advice on the future probabilities of extreme weather enabling more informed decisions on resilience investments.

6. E-science in basic research

The academic community in the UK increasingly uses e-science to underpin much of its work. Current programmes of funded research include [iv]

- Research into climate change, dispersion of pollutants, next-generation power sources, energy distribution, nanoscience and new materials;
- Responses to emergencies and other time-critical incidents. These currently include simulation and prediction of fires, earthquakes and extreme weather events, but in the future will extend to real-time clinical decisions;
- The retrieval and mining of complex structured information from large corpora of historical and other texts, and the integration of this information with relevant databases;
- The scanning and documentation of ultra high-resolution 3D scans of museum objects and associated machine-readable metadata, and making the resultant data widely available;
- Patient-specific computer models for personalised and predictive healthcare developed through the European Virtual Physiological Human and other programmes;
- Medicine where complex surgery, tumour imaging and cancer treatments depend on advanced computing;
- Basic biology and drug design which depend both on advanced computing and access to large, diverse databases;
- Novel public transport planning services devised using accessible public data;
- Simulation of population dynamics and aging to predict future care requirements in support of planning by administrations.

In astrophysics there is an intimate connection between theoretical research, which today is overwhelmingly reliant on modelling and simulations using HPC, and the exploitation of data from expensive observational facilities. The UK continues to be a recognised world leader in theoretical astrophysics. For example, the most cited paper in astronomy published in Nature during the 2000s presents the "Millennium simulation", in which UK-based researchers played a key role. The Millennium simulation data have subsequently been used in at least 400 other papers by astronomers across the world.

A Revolution in E-enabled Science and Innovation

Introduction

The importance of e-science and innovation for academy and industry rests on rapid advances in the following areas: the development of new software algorithms for modelling and simulation; data intensive computing and storage for the mining of knowledge; high performance computing hardware and architectures; open and accessible networks to link researchers to computational hardware and data; and the development of skills in e-science. These are the five components of a national e-infrastructure that will need to supported by a governance structure to coordinate the whole effectively.

1. Software and applications

Software for computer simulation and for data-driven computing is a critical and valuable resource. Access to high quality software is a major competitive differentiator, both for academia and industry, whether running on specialist or commodity hardware. The typical high-level software lifetime of 10-20 years greatly exceeds the 2-3 year cycle of hardware replacement, from which we can see that software represents major investments in skilled scientists and engineers amounting to 10s to 100s of person-years equivalent to many millions of pounds.
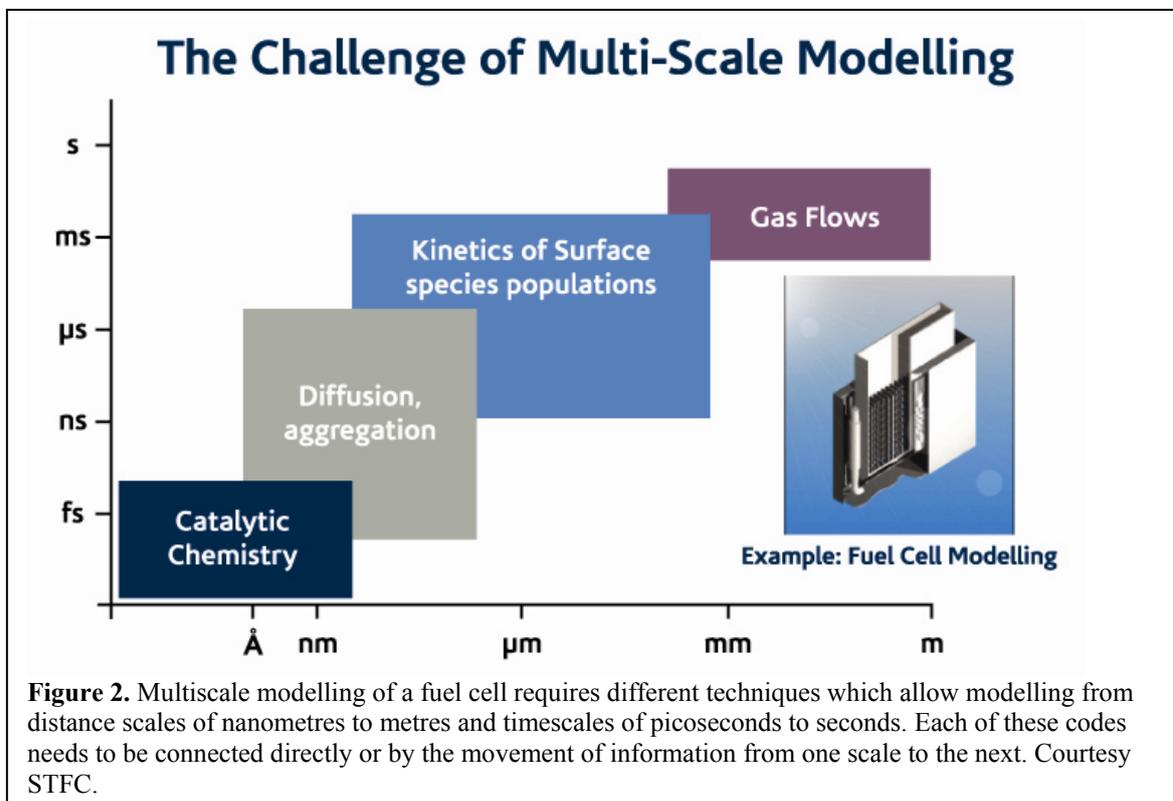

The current situation in software

Currently, academic software development for the e-infrastructure in the UK is mainly funded by the Research Councils. Through the Computational Science and Engineering Department at Daresbury, the STFC is host to a number of Collaborative Computational Projects (CCPs) which assist universities in developing, maintaining and distributing computer programs and promoting the best computational methods. These CCPs are funded by the STFC, the Engineering and Physical Sciences Research Council (EPSRC) and the Biotechnology and Biological Sciences Research Council (BBSRC) and are focussed on major open source, computational science codes. The Edinburgh Parallel Computer Centre (EPCC) has a strong team of software engineers with an expertise including: grid computing; data integration; and computer simulation. EPCC has an in depth knowledge of database programming, network programming and parallel programming technologies (for example MPI and OpenMP). There are many other examples of research grants to Universities that are awarded specifically for software development.

In the private sector, an excellent example of software and applications is afforded by the CFMS Advanced Simulation Research Centre (ASRC) opened in Bristol in 2010 (http://www.cfms.org.uk/). CFMS is an independent, not-for-profit organisation formed by UK industrial concerns (Rolls-Royce, Airbus UK, Williams Grand Prix, BAE Systems, MBDA and Frazer-Nash Consultancy) and its focus is the delivery of more intuitive and powerful simulation-based design processes. Alongside supporting its members, the CFMS-ASRC is committed to making itself available to high end technology focussed on SMEs in the automotive, aerospace and wider industrial sectors. A similar model would be the Virtual Engineering Centre of the University of Liverpool located on the Daresbury Campus (http://virtualengineeringcentre.com/).

There is an exciting opportunity for a significant step change in simulation through the development of models across many different length and timescales. In the past the automobile industry would have been content to model the collapse of the rear bumper of a car on impact. Now

the industry has the ability to model the effect of the impact on the structure of the whole car, the deployment of the air bag and even the effect of the impact on the internal organs of the passenger. Simulations can involve the deployment of many different modelling techniques to the same problem, so that in the simulation of a fuel cell, it may be necessary to apply quantum mechanical modelling of the chemistry, molecular dynamics simulation of the diffusion of species, mesoscale models of surface population and computational fluid dynamics of gas flows in one calculation (Figure 2).

Figure 2. Multiscale modelling of a fuel cell requires different techniques which allow modelling from distance scales of nanometres to metres and timescales of picoseconds to seconds. Each of these codes needs to be connected directly or by the movement of information from one scale to the next. Courtesy STFC.

In order to ensure leadership for the UK, we need to step beyond the current situation and bring together software development teams comprising experts from industry, the mathematical and physical sciences, informatics and computational science, together with the appropriate domain experts. This expertise will need to be focused on the development of new codes based on new ideas for academic and industrial communities, on the porting of legacy codes to high end machines for simulation and data-intensive computing, and for the transfer of important commercial codes to high end machines.

2. Data-intensive computing and storage

The present revolution in e-science is driven by the massive increases in the size and complexity of digital datasets. Industry analysts[ix] predict that digital data volumes will grow by at least 44 times between 2009 and 2020 with every industry segment facing the challenge of managing huge volumes of data.

As an example, the explosion of genomic data arising from continuing developments in sequencing technologies has led to a supra-exponential increase in the size of databases such as those maintained by the European Bioinformatics Institute (EBI). EBI-EMBL's Nucleotide Sequence Database grew from ca. 10 Gigabases in 2000 to ca. 200 Gigabases in 2009. This is the now the output of a single experimental sequencing machine in just one week. Current projections suggest that the amount of data storage capacity for gene sequences alone will rise from 6PB in 2011 to 66PB in 2019. Similar explosions of data and opportunities exist in the environmental sciences

(ocean and earth sciences), astronomy, and in health and well-being (understanding consumer and patient needs, semantic modelling, neurobiology etc.)[v]

The question becomes: how can such large volumes of data be managed successfully and be usefully interrogated to deliver meaningful insight and business value? New methods are required to analyse different types of data. Techniques that apply to large volumes of static data (such as medical image archives) are different from those required to analyse streams of data (such as traffic management systems and CCTV images). This is leading to a convergence of HPC and data analytics technologies as the two requirements merge and new parallel algorithms are developed. The data analytics market alone is predicted to be worth around $18 billion by 2015 and has applications across all industry sectors from healthcare to financial services.

Computer architecture has already changed to integrate computing power and data storage. Sequential analysis of large datasets is giving way to parallel computation, performed by groups of identical, 'commodity' off-the-shelf (COTS) computers connected together with high speed networks to form a local area network. Such 'Beowulf' clusters can be programmed to allow them to distribute processing tasks but are technically unsuitable for working on problems that are not readily parallelisable. Computing Clouds, large pools of accessible, virtual resources (hardware, storage, services) that are dynamically reconfigurable, offer an alternative approach to data analysis. In addition to compute cycles for analysis, such remote, networked clouds can now offer back-up and archival services through standard protocols[vi].

As storage densities increase, a sustained annual investment can be expected to result in increased data volumes for the same capital expenditure year-on-year helping the infrastructure to keep pace with data growth. As data volumes increase, technologies for indexing, annotating and retrieving data become increasingly important and must be addressed in parallel with the capital expenditure.

The issues in data-intensive computing that will need to be considered in building an infrastructure include: the massive size requirements for storage and its geographical positioning with respect to the experiment or simulation that created the data; effective data retrieval and archival of the data across the network; recognition and development of parallel analysis algorithms for inherently parallel problems and their deployment across high performance machines; the recognition of inherently non-parallelisable problems and their solution on local environments; and finally an understanding of when it is better not to store data but to recreate them as required for analysis.
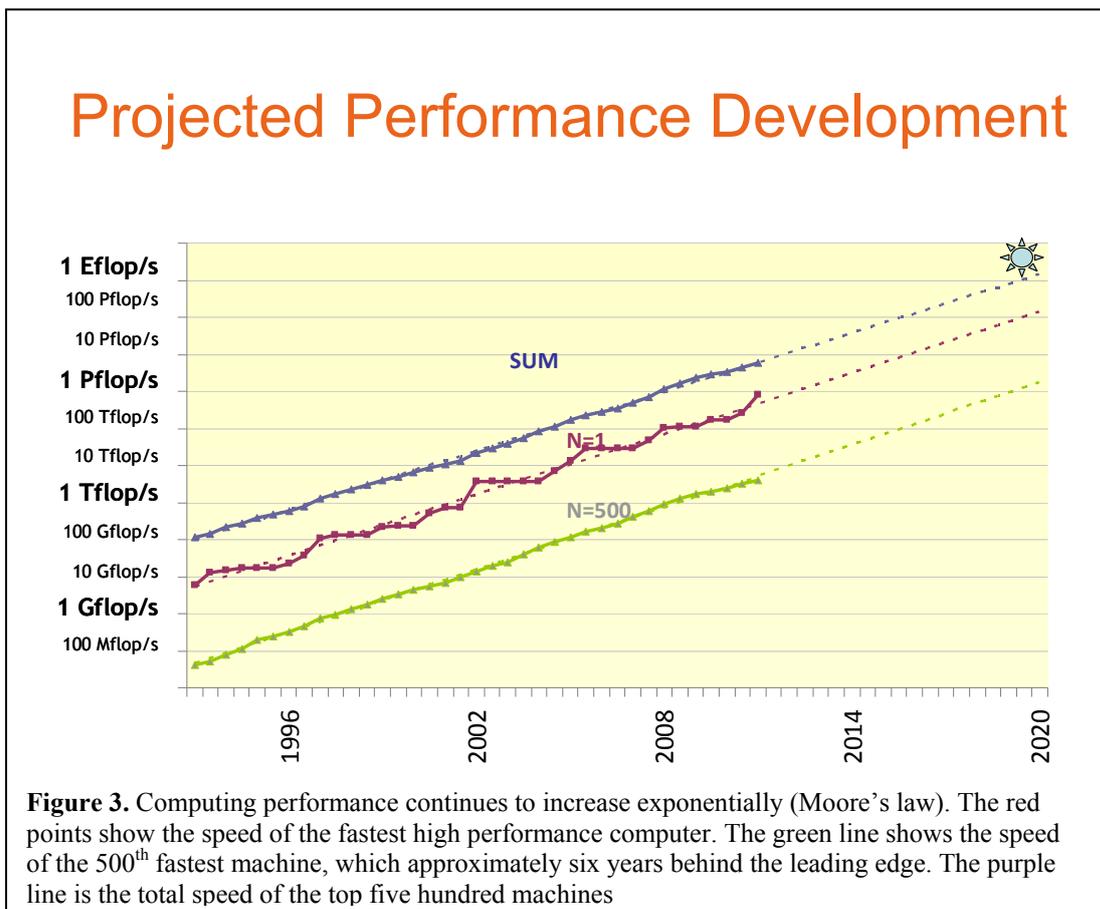

Unilever wished to identify active compounds for anti-perspirant benefit through the inhibition of a key protein target. Existing actives for the target were identified from 5½ M data points within the ChEMBL database. These were used to build a model relating the structure of a compound to its inhibition activity. The model was then applied to an internal database of the structures of 13.8M compounds that are available from suppliers to identify 166 novel, putative, actives. These were then acquired and tested experimentally for enzyme inhibition activity and this led to 24 compounds with an IC50 in the nanomolar activity range. The best was 112x more potent than the previously identified Unilever lead. From identification of target to final results took 4 months.

---

3. Performance growth and the change in architectures

The advance in the power of computers has been phenomenal, with decade upon decade growth in performance of around one thousand-fold every ten years coupled with static or falling prices (Figure 3). A well known but anonymous quote says *"If cars had developed at the same rate as computers you would be able to buy a Rolls-Royce for a penny and travel to the moon and back on a thimbleful of fuel"*. In earlier years this growth was driven mainly by faster clock frequencies and

instruction level parallelism, programs ran faster on the latest computers with no need to change the software. However the industry has recently (around 2004) reached a limit whereby clock speeds cannot increase further without the chips consuming too much power and getting too hot. This has halted the growth for individual processor cores but not for systems as a whole. Performance increases have been driven in recent years by a massive increase in parallelism.

The leading computer systems now have of the order of 100,000 processor cores and 1000-core systems are commonplace in R&D departments and universities. In 2008 these 100,000-core machines reached petascale performance (10^{15} operations per second) and well-founded predictions indicate that exascale performance (10^{18} operations per second) will be achieved by around 2018. At this time, we will have around 1000 times more computational performance than is available today. These systems will contain hundreds of millions of cores and will require significant technology breakthroughs in power consumption and management, interconnect bandwidth, memory, packaging and systems software all of which have cascading implications for industry.

Figure 3. Computing performance continues to increase exponentially (Moore's law). The red points show the speed of the fastest high performance computer. The green line shows the speed of the 500th fastest machine, which approximately six years behind the leading edge. The purple line is the total speed of the top five hundred machines

As performance and parallelism grow at the top-end, and while the pursuit of exascale performance is in itself an exciting challenge for the IT industry, it will be revolutionary largely because the new technologies required will have to allow today's elite petascale systems to become available in a single rack. Thus, the real value behind the drive towards the exascale is the trickle-down effect which will allow a multi-petascale system to be affordable for a local industry or small business. Even PCs and workstations, the bread-and-butter resource of the scientist and engineer, will not be immune as they too are increasingly equipped with multi-core processors. This continuing growth in price/performance and power-efficiency comes at a cost: the new systems will be based on complex multi-core and accelerator-based architectures which are much more challenging to program than are today's systems, requiring a revolution in software design and development

The issues in high performance computing that will need to be considered in building an infrastructure include: the planned purchase and upgrading of major national super-computer resources; the balance between this provision and specialist high-end machines for particular disciplines (e.g. quantum chromodynamics, astronomy, genomics) and/or regions; and the balance of this resource with campus HPC provided through individual universities and clusters of universities[vii].

4. Networks

Modern data communication networks based on optical fibres have a very high inherent transmission capacity, and access to networks of this kind will be essential to realise the full potential of the e-infrastructure vision by enabling the wide-scale, high-volume, low-latency data movements that are foreseen.

These networks are generally available nationally through telecommunications companies providing commercial services on the open market, and also through organisations with a specific remit to provide network services to national academic, research and education communities. In the UK this organisation is JANET (UK) which provides the JANET network as a service on behalf of JISC and its funding partners.

JANET operates its core network built on 40 and 100 Gbit/s link components, and many instances of these can be operated in parallel so there remains great potential for augmenting capacity where needed. Typical access capacities for universities and research organisations are currently in the range 1-10Gbit/s, generally delivered as a single communications channel over a transmit/receive pair of optical fibres. Where closely co-located, this would be a direct connection between the JANET and the site equipment, but in other cases the link would be provided through a third party service from a telecommunications company. These link capacities can also be increased where required to provide either additional IP service capacity or dedicated point-to-point circuits/lightpaths. Various technical approaches might be used for implementation, and for some key strategic locations with a high potential for future demand there may be a need to extend direct fibre access from the JANET core to the sites in question. The next standard for high-capacity commercial networking products is likely to be 400Gbit/s around 2015, followed some time later by 1Tbit/s products. JANET will be engineered to accommodate these technologies, and all components of the e-infrastructure environment, from end-systems through both local and wide-area networks must be prepared to evolve through technical upgrades at an appropriate level to realise the high-throughput environment required for an effective e-Infrastructure.

Any organisation that wants to participate in the e-infrastructure environment will need to establish a connection of appropriate capacity via a suitable network service provider. The global ensemble of national research and education networks which includes JANET and its peers in Europe connected together by GEANT (the pan-european research and education backbone network) provides an environment with the advanced service features and community coherence to support large multi-national projects. There are already project examples in the areas of physics, astronomy and High Performance Computing, with other disciplines actively planning similar activities, and building on this environment would be a natural path towards a more comprehensive and coherent e-infrastructure.

The issues in networking development that will need to be considered in building an infrastructure include: finding the right regulatory and funding environment for industrial users to be able to gain more comprehensive access to networks such as JANET and GEANT; the research effort to sustain high-throughput end-to-end data-transfer rates as the network capacity increases; and the additional effort to provide the skills and training needed for advice and guidance on matching end-systems to high-capacity networks.

5. Skills

All components of the national e-infrastructure require highly skilled staff to operate effectively. This will only be achieved by re-training qualified scientists and engineers in new e-science skills and providing explicit encouragement and training opportunities for the next generation of young researchers in this field.

In the area of software development the need has been identified for the development of new codes as well as the evolution, rewriting, and hardening of existing codes. This will be achieved by the formation of multi-disciplinary teams including domain scientists, industrial R&D engineers, computational scientists, computer scientists, numerical analysts and hardware technologists.

The UK needs to develop a young, and growing, community of researchers who are able to exploit current and future leading-edge high-performance and data-centric computing to the fullest extent. Where appropriate this should be linked to existing computer science and domain-specific training programmes e.g. by working in partnership with universities to provide technology and computationally-focused training modules as part of existing Doctoral Training Centres (DTCs) and similar graduate training constructs. Where academic-industrial partnerships are already established there is a strong need to train industry's scientists and engineers in the use of the latest techniques in high-performance parallel applications and in high-throughput data management and visualisation. These modules could be based on those developed in the DTCs.

It is important to place a higher value upon the position of "scientific programmer" and also "data curator" in the academic environment and to offer more career opportunities to these staff. Scientific programmers combine the knowledge of the underlying scientific discipline with implementation, optimisation and parallelisation for high-end systems: they are important in obtaining highly-efficient application implementations.

Training programmes should also develop the skills needed to adapt existing codes and write new codes that will be efficient for modern programming languages and emerging computer architectures. Training should also help to encourage the transfer of software development between disciplines. This will ensure the emergence in the UK of a new generation of researchers who have the ability to realise the potential of high performance and other computing in a research environment that is increasingly interdisciplinary and multi-scale.

Scientists outside the domains of engineering and physical sciences are particularly unlikely to have been exposed to the necessary skills, and will need special mentoring if they are to make the most of the opportunities offered by e-science.

The problem and a solution

In the previous sections we have outlined the revolution in e-enabled science and innovation and the corresponding advances in technology that lie at the heart of much of modern science. The UK, as judged by international panels, is world leading in many of these disciplines. At the same time we have highlighted the opportunities of applying these techniques in industry and the potential of this area to grow the UK economy and provide tens of thousands of jobs over the next ten years. Increasingly the needs of industry and the public sector are the same as those of academe.

At present, however academe, industry and the public sector are unable to reap these advantages because of the lack of a guaranteed e-infrastructure. This section elaborates the problem and then suggest a solution building on the E-infrastructure Advisory Group Report[viii] and the Coveney report[iv].

The problem

- HPC in the UK is losing touch with the most advanced high-end supercomputers. The world's fastest super-computer in Japan is approximately thirty times more powerful than the UK's HECToR. In the last four years the number of UK machines in the world's top 500 has dropped by 50% while the number in France and Germany has shown a steady improvement. According to the IDC[ix] the US is spending $2.5Bn p.a. on high-end computing provision and they predict a rise to $5Bn pa. The leadership that the UK has shown in developing HPC technology and expertise over the last twenty years is under threat.

- The software application cycle is 10 years, or more, as compared to 2-3 years for hardware. Industrial software development and vendor software lags behind. There is a substantial volume of legacy software that needs upgrading. The lack of high quality open-source and commercial software prevents large and small enterprises from exploiting e-science. Open source software permits scrutiny and reproducibility, core tenets of the scientific process.

- Individual firms cannot keep up with the rapid developments in hardware and are not able to use e-science most effectively to increase growth. This is true for both larger corporations and SMEs. Significant economic growth will only be achieved through a shared infrastructure and a sharing of the costs of establishing and running it.

- Operating costs including cooling, electricity and staffing are growing especially for high end facilities. The next generation of petaflops machine and the petabyte data bases will be even more demanding in terms of running costs. Future provision will require sharing. It is vital that all capital planning includes all necessary recurrent cost provisions.

- Firms cannot interact easily with a fragmented national infrastructure or without a clear forward commitment to a strategy. Any move from a local, in-house provision of computing to the use of a shared infrastructure will only be achieved if a clear plan is developed and adhered to. (This problem applies equally to academe.)

- There is no single coordinating body for e-infrastructure in the UK that can represent all relevant stakeholders, both providers and users, in the public and private sectors. This has resulted in the lack of a coherent strategic plan for e-infrastructure. Unless action is taken to coordinate the UK provision the socio-economic benefits from e-enabled science and innovation will stall.

- There is fragmentation of planning and funding across e-infrastructure in the UK. For example, in the academic sector, different Research Councils have established different models of computing provision. A more efficient and effective provision might be achieved if the balance across all elements of the infrastructure was agreed and implemented by the Research and Funding councils.

- At present there is no agreed plan for coordinating provision of campus, regional and national computing facilities. There are clear advantages to the universities to share hardware provision and software development where possible (as exemplified by the High Performance Computing Wales cluster) but the lack of a clear forward plan for the UK hampers these developments.

- There is no coordinated plan for skill development. A steady stream of software engineers and developers is needed. In particular we need experts in mathematical and

discipline-based skills to be trained additionally in software engineering for e-science. These types of developers are in short supply.

A solution

A solution to this problem which would enable the UK to take full advantage of the rapid developments and applications in e-science could be to:

- Create a ten-year roadmap to define the components of the infrastructure: networks; data and storage; compute; software and algorithms; security and authentication; people and skills. This roadmap would serve the academic, government and industrial communities and be benchmarked against other national developments such as those in the US, Japan, China, France and Germany and provide for UK contributions to international developments (e.g. PRACE[x]).

- Create a substantive and internationally-leading software base which is robust and reliable, which encapsulates new as well as well-established algorithms, and which will help to harden academic software. This base will include industrially relevant vendor software, which is scalable and well optimised for current as well as new and emerging architectures and which is accessible across the academic, government and industrial communities.

- Create secure data and information stores in strategic locations with data-analysis provided through cloud environments, working with open source software.

- Ensure that important public databases are available to all UK researchers (see the 8 UK data centres outlined in the recent JISC report[xi]).

- Provide broad access to the infrastructure for industrial partners, suppliers and Independent Software Vendors (ISVs), as well as the academic community. Access to the infrastructure would be for mutual benefit. For example an ISV might provide versions of their code for use by limited parties in exchange for its migration on to a high end machine. Normal access for non-academic partners would be through a rental or pay-as-you-go model where all aspects of the e-infrastructure could be purchased.

- Assist the development of a portfolio of training modules in computational science, numerical algorithms grid-computing, parallel programming, cloud computing, data-centric computing, e-science, computer animation and computer graphics. Include elements of this training in a number of established Doctoral Training Centres where e-science would sit along side disciplines such as physics, biochemistry, archaeology, social sciences.

- Develop a single coordinating body to drive closer cooperation and enable effective industrial access, while insuring that UK academe has access to leading edge capability. This coordinating body should develop a vision for UK e-infrastructure that will increase productivity and efficiency for all of the partners.
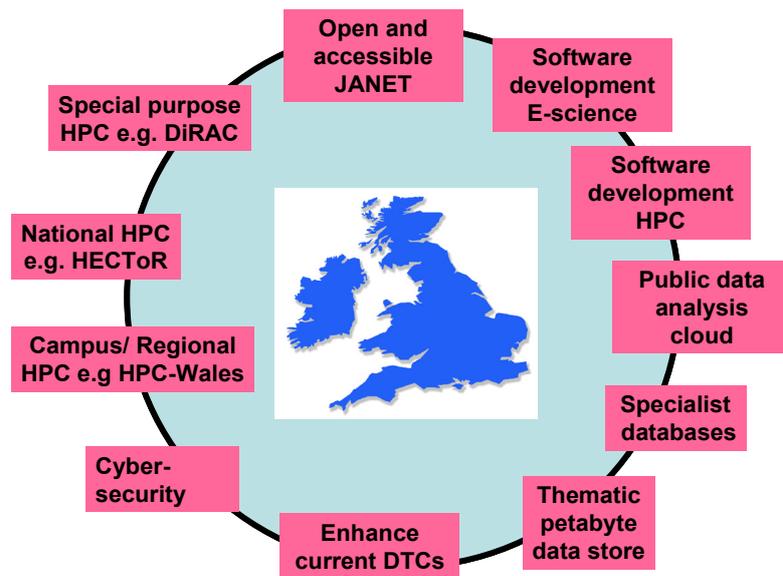
The e-infrastructure ecosystem

Figure 4 The proposed e-infrastructure for the UK including high performance computing, data driven computing and storage, software development, training and skills, networks and security and authentication.

The current baseline

In this section, we will outline the current baseline spending which could reasonably be associated with the e-science infra-structure as it is currently established for the public sector (Research Councils and HEFCE). This includes recurrent and capital spending on: networks: data driven computing and storage; high performance computing; software development; people and skills; security and authentication. In Figure 5 we show this spending over the period from 2009 to 2013.
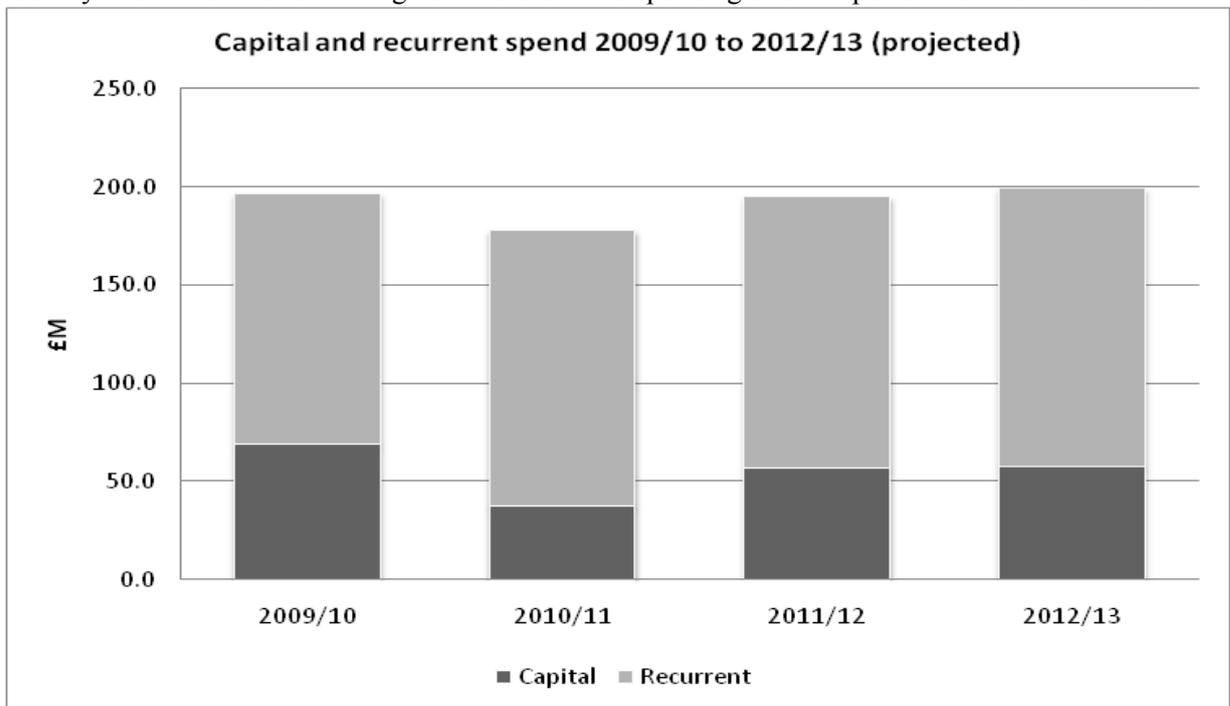

Figure 5. The capital and recurrent spends on e-infrastructure over the period 2009 to 2013.
It is important to understand what is included and what is not included in these numbers and we have chosen to illustrate this with a narrative in Appendix A where we outline the capital and recurrent expenditures separately. Note that the recurrent is considerably greater than the capital.

We propose a similar approach to that adopted by the Research Councils for the funding of the large scientific facilities in the UK; ISIS, the Diamond Light Source and Central Laser Facility. In the Large Facilities Funding Model, the Research Councils agree on the appropriate operating costs for these facilities from the Science Budget. These funds are ring fenced and administered (in this case by STFC) on behalf of all Research Councils.

For the e-infrastructure we recommend that the proposed E-infrastructure Leadership Council, supported by the proposed E-infrastructure Secretariat (see E-infrastructure Leadership Council), should advise the Government on this baseline funding in future. This ring fenced budget should be held by BIS and the relevant Research and Funding Councils, and administered under the oversight of the E-infrastructure Leadership Council using the governance structures of the Research and Funding Councils and the Large Facilities Capital Fund. Funding from the private sector (large corporates, SMEs and trading funds such as the Met. Office) will be additional to this baseline.

Additional funding

Recently, the Government has agreed an additional capital spend of £145M to enhance the e-infrastructure in 2011-2012. We recommend that this money should be used to establish a partnership between academe, industry and government that will develop over the next ten years. The E-infrastructure Leadership Council should monitor progress with this partnership as part of the implementation and development of this Strategy.

An emerging business model

The e-infrastructure model that we are proposing has clear advantages for academe, charities, suppliers, industry, Government and the UK economy more widely.

Advantages to academe

It has been said that "3 months in the lab can save a whole afternoon on the computer." In other words, the productivity of scientists and engineers who have the skills and abilities to harness the tools of e-science will be massively greater than those who do not, and it is vital that the UK maintains our current lead. This extends to the 'reading' (i.e. text mining) of the scientific literature, where even PubMed alone (the peer-reviewed papers relevant to biomedicine) is increasing at 2 papers per minute. Improved connectivity will also make far more effective the constant video interaction necessary for optimal inter-laboratory collaborations. The use of data standards for interoperability, and the existing requirement to make data available openly, will allow the UK to lead in the use and exploitation of publicly available datasets, in all scientific disciplines, including social science and the humanities.

The infrastructure will not only provide better science but will also spawn new kinds of science. In addition to increased computational resources and more effective networking and data handling, the infrastructure will encourage interdisciplinary projects extending over the boundaries of the Research Councils allowing, in particular, for the transfer of software from one domain to another.

The provision of a clear e-infrastructure roadmap, with indicative funding, out to 2020 will provide a solid foundation for UK academics to tackle leading edge problems in all disciplines with confidence. The deep involvement of industrial partners in the infrastructure will bring the academic community close to a set of new of interesting and applications of high commercial value and facilitate the important process of knowledge transfer which is at the heart of Government strategy.

Advantages to industry

Industry analysts report that the major inhibitors to further industrial adoption of high-end computing include the cost of applications, the lack of parallel application development tools, the lack of skilled staff and the limited scalability of home-grown applications. The proposed infrastructure and business model will allow industry to realise the full potential of e-science. By sharing a common infrastructure with academy and government, industry will have full access to expertise and skills in programming and parallelisation without the need to build large in-house teams, which can be isolated and where it is difficult to maintain skills at the highest level. By renting resources, or paying for computing cycles, industry can avoid massive depreciation costs before the value of a particular modelling or analysis approach has been full validated. The danger that large and expense hardware resource may sit idle in a company environment due to the peaks and troughs of a project lifecycle is avoided by sharing with other industries and the public sector.

Advantages to medical research charities

Biomedical research in the UK is supported by a significant charitable research sector. For example just two, the Wellcome Trust and Cancer Research UK, together spent over £1 billion on their support for research and health information projects in the year to March 2010 (source - Charities Commission Register). As medical and biological research becomes increasingly computational and data driven, IT costs associated with research projects are rising to 5-10% (in some cases, such as genomics studies, even higher) of overall research spend. As bioinformatics becomes a key part of the business of medical research, access to a pool of HPC expertise, consultancy, software skills and training opportunities would allow UK biomedical researchers to maintain their world leading rankings in this area.

Advantages for suppliers

According to analysts such as IDC[ix] and Interset360[xii] the overall market for HPC is growing with a compound annual growth rate of 6-7%. This is against a backdrop of declining revenues in the IT market as a whole. All sectors, from financial services to advanced manufacturing are growing as a consequence of increased demand for higher fidelity models and the requirement to run at ever larger scale. In addition to the systems themselves there is a pull through effect in other sectors of the IT market where, for example, additional storage systems are required to manage the vast quantities of new data being generated. This in turn creates an increased demand for the data analysis required to extract value from large-scale data. For every pound spent on HPC systems there is at least another pound spent on storage, applications and services. The infrastructure that we propose will allow the effective deployment of massively parallel systems and data storage. This will provide hardware vendors with a clear road map for the UK, allowing them to invest in staff and resources, and significantly advance opportunities for serious partnerships for developing and testing at the leading edge comparable with those being established in other European countries such as France and Germany.

In addition to core scientific applications, a need exists for the new development tools such as profilers and debuggers, libraries of mathematical and scientific subroutines, system management applications and other tools designed to ease the burden on the programmer. The range of partners and hardware architectures associated with the e-infrastructure will encourage ISVs to adopt innovative licensing models that are not tied to processor count.

Advantages for the UK

Economic infrastructure drives competitiveness and supports economic growth by increasing private and public sector productivity, reducing business costs, diversifying means of production and creating jobs.[xiii] There is a clear correlation between investment in infrastructure and long-term growth. The Government has recognised this in the National Infrastructure Plan 2010.[xiv]

That Plan highlights the importance of digital communications and the significant positive impact both on gross value added in the economy and on employment in the information and communications technology sector and the wider economy.[xv]

With the right e-infrastructure, Government can make the UK a better place to live and to do business by:

- enabling improvements in scientific and business productivity and growth through more efficient ways of working, and more efficient communication and exchange of information with peers, customers and suppliers;
- accelerating growth and job creation through new business formation and growth in the technology sector;
- increasing the supply of skilled people available to harness the opportunities offered by leading edge e-infrastructure;
- and supporting better and more efficient ways of delivering public services.

By investing in e-infrastructure the Government is also further increasing the attractiveness of the UK for inward investment, as set out in the UKTI Strategy "Britain Open for Business."[xvi]

It is clear therefore that by joining with the private sector to develop an appropriate e-infrastructure and advanced computing capability the Government is positioning the UK at the forefront of the exploitation of digital technologies.

Building the partnership for growth

Successful harnessing and exploitation of current and future computational platforms in European economies could lead to an increase in Europe's GDP of up to 3% within 10 years. The economic returns can be viewed as coming from two major areas:
- The HPC supply chain (potential to add 0.5% to 1% to Europe's GDP by 2020)
- Industries that leverage HPC to improve their products and services (potential to add 2% to Europe's GDP by 2020)

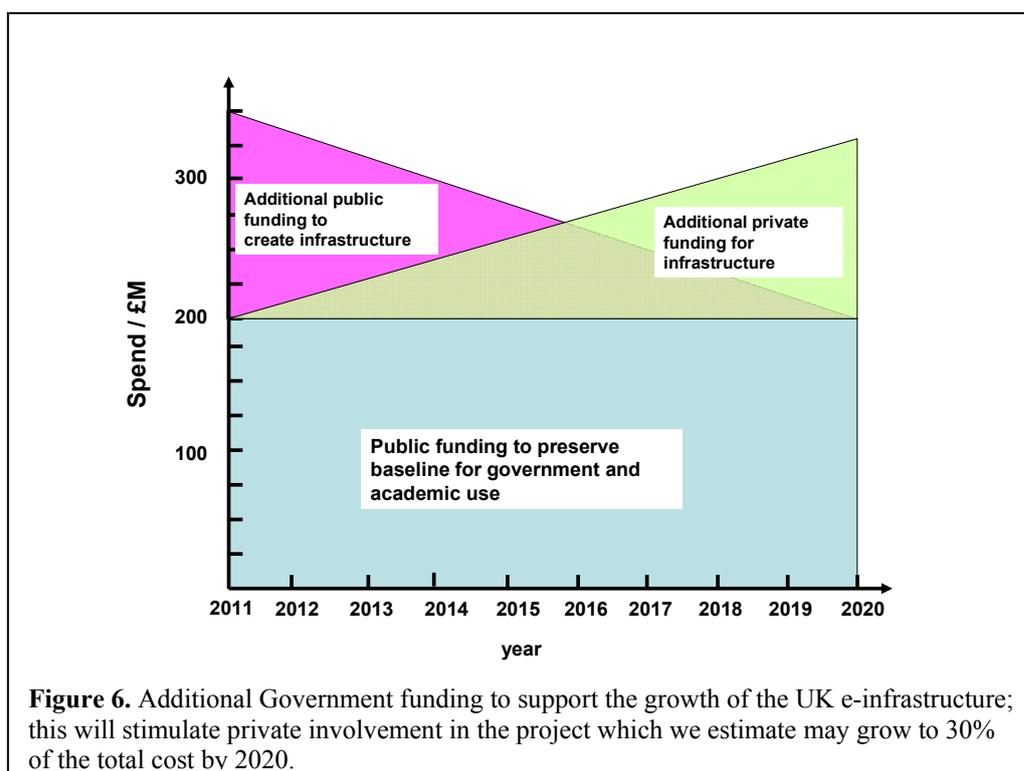These predictions come from an independent report commissioned by and delivered to the EU by IDC in 2010[ix].

HPC is itself a global industry. Current IDC market tracker data suggests that the world wide market is around $10.5Bn in 2011 growing to around $14Bn in 2015. This figure rises to $18.5Bn when software, middleware, storage and services are included[xvii]. The UK is recognised as a world leader in the writing of software capable of efficiently exploiting modern, multi-core hardware, evidenced in particular by the numerous ongoing collaborations between UK researchers and large US HPC laboratories.

The right e-infrastructure in the UK will attract highly skilled staff, visitors and collaborators to the UK. It will lead to permanent staff, long-term collaborators and a large compute resource being located in the UK. It will allow the training of developers with the necessary skills to exploit all new computer technologies and so equip the UK with a highly skilled workforce in this area.

As outlined previously, the UK Government expects to support the Research Councils, University e-infrastructure and JANET(UK) with a baseline spend of just under £200M p.a. from within the science budget. In a recent announcement[xviii] Government has provided an additional £145M in 2012. This additional funding is to enhance the infrastructure to make it an attractive proposition for the private sector to invest in over the next ten years.

There are elements of the e-infrastructure that no private sector company will invest in as they do not require sole use of these assets. Some companies invest heavily in HPC for competitive advantage where a business case based on productivity gains can be made. Where individual companies cannot justify such investment, they need access to a national e-infrastructure for key skills and services. They will pay for this service. The use that industry will make of a national e-infrastructure is visible tip of iceberg in terms of their computational investment. However, if the UK national infrastructure is not adequate, industry will move its own investments overseas – the whole iceberg.

We estimate that over ten years the industrial contribution to the e-infrastructure could rise to £100M p.a. As large UK-based industrial R&D gains confidence in the infrastructure, we would expect their own supply chains comprising a large number of SMEs to increase the investment to a level of approximately 30% of the total by 2020.

Figure 6. Additional Government funding to support the growth of the UK e-infrastructure; this will stimulate private involvement in the project which we estimate may grow to 30% of the total cost by 2020.

One model would involve the private sector buying in to the e-infrastructure on a needs basis. This payment could be for computational cycles on a high end HPC machine to support their internal hardware provision. It could also pay for access to secure data-storage, access to public sector databases, software development, upgrading of legacy software, or general advice and support in the e-science arena

Several expressions of interest in such a model have already been received by the Government from large UK-based industries but we expect that the bulk of the revenue from 2015 onwards could come from small and medium-sized enterprises, who need help and support to take full advantage of the e-science revolution.

An interesting example of the way smaller enterprises can be stimulated to use an e-science infrastructure comes from Scottish Enterprise, which has increased the number of companies using HPC in Scotland. They are funding a small set of "HPC Adopter Projects" to get companies over the initial hurdle of using the infrastructure. Last year they funded a series of pilots and undertook an exercise to measure the Gross Value Added (GVA) to the economy properly.

Three projects, with Deep Casing Tools Ltd, FIOS Genomics Ltd and Prospect FS Ltd, cost £256K with £110K coming from Scottish Enterprise. The companies matched the SE funding with in-kind contributions. The resulting GVA study showed that by 2013, £10 net GVA will be generated for every pound of Scottish Enterprise spend with a five-year projection of £25.

This model, with real industrial engagement, could be used as a template for a much greater increase in the use of e-enabled science and innovation across the UK. Our analysis shows that there around 1,400 companies who could make use of HPC with around 300 who are quite likely to take up the opportunity.

The E-infrastructure Leadership Council

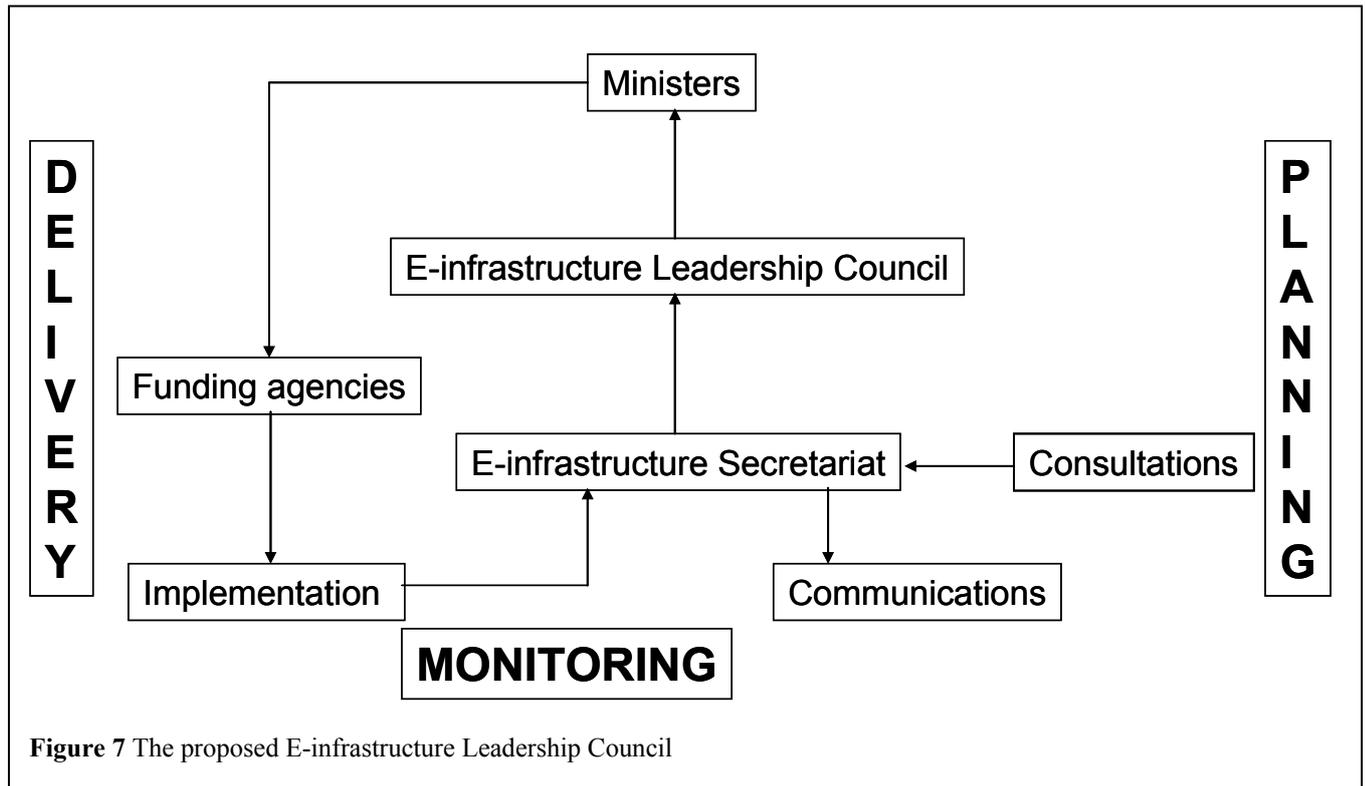**Figure 7** The proposed E-infrastructure Leadership Council

Figure 7 shows the proposed governance of the e-infrastructure. At the heart of this we are proposing that Government establishes an E-infrastructure Leadership Council (ELC). The purpose of the ELC is that it should make recommendations to Government on all aspects of the e-infrastructure as this infrastructure supports academe, industry and aspects of Government computing. The ELC should be a forum where academe, industry the charitable and public sectors can come together to exchange views and discuss all aspects of the development of the e-infrastructure. This should include not only domestic aspects but also developments overseas. The ELC should make recommendations on capital spend (which will be critical for decisions pertaining to equipment associated with networks, HPC and data-driven computing) and recurrent spend (which will be critical in supporting our vision of software development, training, networks and authentication and security). It should be chaired jointly by the Minister for Universities and Science and an industry representative. It should have a composition of experts from academe, industry, the Research and Funding Councils, other Government Departments and the charitable sector. Membership by invitation from Government will normally be for two years and the ELC should normally meet four times a year.

The ELC should be supported by an E-infrastructure Secretariat (ES). This will be a small independent team staffed from BIS and the Research and Funding Councils. Secondees from other

stakeholder groups should be encouraged. It will provide position papers and draft recommendations to the ELC. It will consult widely with users in academe, industry and Government to reflect the needs of the community accurately in the information that it provides to the ELC. The ES will also be responsible for communication to the e-science and innovation community.

Recommendations from the ELC will go to the Minister. His decisions will be reflected in the ring fenced baseline budget held by BIS and the relevant Research and Funding Councils. The Secretariat will keep industrial and Government partners informed so that developments in the e-infrastructure can be included in their R&D planning processes. There will be strong interaction between all stakeholders and the ES during the implementation phase to ensure that the ELC is kept informed of progress.

Recommendations

We are making the following recommendations:

- The ELC to agree the costed, ten year roadmap to achieve an infrastructure with the 5 elements of the e-infrastructure and the appropriate security and authentication. The baseline funding should be held by BIS, and administered using the governance structures of the Research and Funding Councils and the Large Facilities Capital Fund. This will provide the strong foundation required to extend the infra-structure beyond the public sector. Use the recently announced capital spend of £145M (2011-2012) to extend the e-infrastructure to a partnership between academe, industry, the public and charitable sectors. This partnership should be strongly developed over the next 10 years so that approximately one third of the cost can be sustained by the private sector through relationships with large corporations and SMEs.

- Drive the development of software for all of the partners in the e-infrastructure. This is fundamental in establishing the value of e-infrastructure for the UK and has needs to be given an equal footing with the purchase of hardware and networking. To achieve this, we should establish a small number of internationally competitive software centres in the UK. These centres will support the development of new algorithms for use on leading-edge machines (parallel and multithreaded architectures). They will port and update legacy software from academic groups and make this software widely available; the centres will try to assure the robustness and quality of this type of open source code. The centres will also establish strong relationships with ISVs to port commercial software to high-end machines. We expect these new software centres to work directly with industry through thematic centres such as the Advanced Simulation and Research Centre (ASRC) in Bristol and the Virtual Engineering Centre (VEC) in Daresbury. These Centres must be strongly driven by the user communities in the public and private sectors.

- Embed e-science training in successful Doctoral Training Centres. Use these modules to offer training to industry. An explicit commitment to e-infrastructure will play an important role in the maintenance of an HPC skills base post initial qualification, providing roles for career path development and a pool of expertise for consultancy, training and skills transfer between industry and academe. This should be supported with an agreed programme of reports, technical workshops and conferences for the community funded by this programme.

- Set up a single coordinating body that owns this Strategy and can advise BIS Ministers on the implementation and development of the Strategy. This E-infrastructure Leadership Council (ELC) should be co-chaired by the BIS Minister for Universities and Science and an industry representative. The composition should reflect the stakeholder groups:

- Support the Council with an E-infrastructure Secretariat (ES), charged with providing working papers and proposals to ELC and for consulting with stakeholders. The ES should engage with the existing and potential e-infrastructure community across academe, industry, the charitable and public sectors to ensure that this Strategy is implemented and developed in a way that reflects the needs and requirements of all the partners. The ES should be a small, independent group staffed by BIS and the Research and Funding Councils. The ES should co-opt technical experts on an ad hoc basis as required.

Next Steps

- Establish the E-infrastructure Leadership Council and the E-infrastructure Secretariat.

- The E-infrastructure Leadership Council to make recommendations on the implementation of this Strategy to determine the future shape and size of the e-infrastructure.

- The E-infrastructure Leadership Council to develop a detailed plan for private sector engagement in the e-infrastructure, working with the TSB. This should include detailed discussions with large corporations, small to medium enterprises and trading partners.

- Develop a wider stakeholder engagement plan for the whole existing and potential e-infrastructure community across academe, industry, the charitable and public sectors to ensure that this Strategy is implemented and developed in a way that reflects the ubiquitous nature of the technology.

Appendix A. The current baseline for e-infrastructure

The following is a short description of elements of the six components of the current baseline spending on e-infrastructure separated into capital and recurrent costs. This covers the time period 2009 to 2013.

Capital Expenditure Figures

Networks

Capital investment under this cost heading relate solely to the cost of provision for the JANET backbone infrastructure.

Data and Storage

Capital investments associated with this cost sub-heading are dominated by the investments made by ESRC and BBSRC. Examples of investment activity include BBSRC/MRC/NERC's contribution to the ESFRI ELIXIR project and ESRC's support for archives and databases such as the Economic and Social Data Service archive (ESDS), Census, and Scottish and NI Longitudinal Studies. This heading does not include investments associated with large embedded/specialised storage infrastructures e.g. parallel file systems or network attached storage associated with HPC

systems. In addition HEI/campus investments in storage infrastructure have been excluded on the basis that the data could not be obtained.

Computational Capability

The Research Councils have provided significant support to activities such as hardware refreshes for the HECToR national high performance computing service (EPSRC), the distributed DiRAC HPC facility (STFC) and ongoing support for the NCeSS hub and nodes (ESRC). In addition, the figures also recognise the significant year on year investments made by the HEI sector in the provision of local mid-range HPC resources. HEI's are not currently able to bid directly or indirectly to central Government for capital funds in aid of e-infrastructure and as such the investments recognised in these figure have been facilitated by individual HEIs as part of their own institution's research strategy. There has also been considerable co-investment from industry (for example the public funding of DiRAC was supplemented by a $0.5M donation of equipment and software from IBM University Relations).

Software and algorithms

STFC recently made a significant investment in supporting the mission of the International Centre of Excellence for Computational Science and Engineering at Daresbury Labs which will be dedicated to the task of developing applications codes for the next generation of computing systems, architectures and application paradigms.

People and Skills

These figures are for the Northern Ireland Longitudinal Study- Research Support Unit (ESRC)

Security and Authentication

No investments specifically recognised over the period although this is likely to be provided by JISC on behalf of the UK through JISC Innovations type activities such as Shibboleth etc and this has been recognised at a notional level of circa £1m pa.

Recurrent Investment

Networks

The costs included relate to the running costs of JANET (UK), network operation costs and user support services.

Data and storage

Ongoing costs across the period are predominantly associated with NERC and BBSRC's support for their data centres (e.g. electricity and staff etc), data support services (staff) and database access (licences). Similar activities were noted for other councils (STFC and ESRC) although the scale was an order of magnitude lower (i.e. less than £1M).

Computational Capability

As with Data and Storage the expenditure under this heading is mainly concerned with the staff resource necessary to run large computational facilities, electricity for power and cooling of systems, maintenance contracts with suppliers for hardware and user support functions such as the Computational Science and Engineering function provided by NAG for the HECToR service. In

addition, HEI expenditure has also been recognised under this host heading for electricity and staff costs associated with the provision of local cluster systems.

Software and Algorithms

Expenditure within this cost heading is mainly dominated by activities sponsored by BBSRC and EPSRC. Examples of recurrent investment include support of the SLA and CCP activity at Daresbury Labs, the NAIS Science and Innovation award, applications software development awards, support for the Software Sustainability Institute, the IDEAS factory awards and co funding for EPSRC-NSF software development projects (EPSRC) and significant directed support into software development for bio-informatics and genomics (BBSRC).

People and Skills

Expenditure over the period has been predominantly associated with ESRC's support of activities such as the National Centre for Research Methods, Welsh Institute of Social and Economic Research Data and Methods and Survey Resource Network. In addition NERC and BBSRC have invested significantly in studentships and skills base development in line with their research priorities over the period.

Security and Authentication

There are no investments specifically recognised over the period although this is likely to be provided by JISC on behalf of the UK through JISC Innovations type activities such as Shibboleth etc and this has been recognised at a notional level of circa £1m pa. Shibboleth is a web based authentication mechanism, and there is now significant effort being invested in standardising a more generic mechanism for authentication that could be used between users' applications and e-Infrastructure resources. The standardisation is proceeding through a working group in the Internet Engineering Task Force[xix], and JANET is leading a pilot activity (project Moonshot[xx]) to provide supporting infrastructure and example applications within the UK community. This has generated a great deal of interest in national and international grid and HPC communities where its potential has been recognised.

Glossary of terms

ASRC: the Advanced Simulation and Research Centre, Bristol. Alongside supporting its members, the CFMS-ASRC is committed to making itself available to high end technology focussed on SMEs in the automotive, aerospace and wider industrial sectors.

BBC: the British Broadcasting Corporation.

BBSRC: the Biotechnology and Biological Sciences Research Council is a UK Research Council and is the largest UK public funder of non-medical bioscience. It predominantly funds scientific research institutes and university research departments in the UK.
BIS: the UK Government Department for Business, Innovation and Skills (BIS).

CCP: the Collaborative Computational projects are funded by the STFC, the Engineering and Physical Sciences Research Council (EPSRC) and the Biotechnology and Biological Sciences Research Council (BBSRC) and are focussed on the development of major open source, computational science codes.

CFMS: CFMS is an independent, not-for-profit organisation formed by UK industrial concerns (Rolls-Royce, Airbus UK, Williams Grand Prix, BAE Systems, MBDA and Frazer-Nash Consultancy) and its focus is the delivery of more intuitive and powerful simulation-based design processes.

Cloud Computing: cloud computing is the delivery of computing as a service rather than a product, whereby shared resources, software, and information are provided to computers and other devices as a utility (like the electricity grid) over a network (typically the Internet).

CRO: a contract Research Organisation

DTC: Doctoral Training centres (sometimes called Centres for Doctoral Training) are centres for managing the Engineering and Physical Sciences Research Council (and the Economic and Social Research Council) PhD-funded degrees in the United Kingdom. Each DTC involves a UK university (or a small number of universities) in delivering a four-year doctoral training programme to a significant number of PhD students organised into cohorts. Each Centre targets a specific area of research, and also emphasises transferable skills training

E-infrastructure: an e-infrastructure, comprises networks, data repositories, computers, software and skills and is an essential tool for research and development both in industry and across a wide range of fundamental science. It enables e-science and technology

ELC: The E-science Leadership Council will offer recommendations to Government on the capital and recurrent provision necessary to build and maintain a first-class e-infrastructure for the UK.

EMBL: the European Molecular Biology Laboratory specialises in innovation in life sciences research, technology development and transfer, and provides outstanding training and services to the scientific community in its member states. This publicly-funded non-profit institute is housed at five sites in Europe whose expertise covers the whole spectrum of molecular biology.EPCC: the Edinburgh Parallel Computer Centre is a team of software engineers with an expertise including: grid computing; data integration; and computer simulation. EPCC has an in depth knowledge of database programming, network programming and parallel programming technologies

EPSRC: the Engineering and Physical Sciences Research Council is the main UK government agency for funding research and training in engineering and the physical sciences, investing more than £850 million a year in a broad range of subjects – from mathematics to materials science, and from information technology to structural engineering. `

ES: the E-science Secretariat will be a small independent team staffed from BIS and the Research and Funding Councils. Secondees from other stakeholder groups should be encouraged. It will provide position papers and draft recommendations to the ELC.

E-science: e-science is the discipline of using digital methods for generating ideas and knowledge from data. This data can come from experiment or from simulation and modelling.

Exa: prefix meaning 10^{18} as in exabyte (a unit of storage) or exaflop (10^{18} floating point operations per second, a unit of computing speed)

GDP: Gross domestic refers to the market value of all final goods and services produced within a country in a given period. GDP per capita is often considered an indicator of a country's standard of living.

GVA: Gross Value Added is a measure in economics of the value of goods and services produced in an area, industry or sector of an economy. In national accounts GVA is output minus intermediate consumption[1]; it is a balancing item of the national accounts' production account

GEANT: GÉANT is the pan-European data network dedicated to the research and education community. Together with Europe's national research networks, GÉANT connects 40 million users in over 8,000 institutions across 40 countries.

HECToR: HECToR is the UK's high-end computing resource, funded by the UK Research Councils. It is available for use by academia and industry in the UK and Europe.

HEFCE: The Higher Education Funding Council for England distributes public money for teaching and research to universities and colleges. In doing so, it aims to promote high quality education and research, within a financially healthy sector.

HPC: High performance computing is the use of leading-edge computers for simulation and modelling and for advanced data analysis.

HPC-SIG: The HPC Special Interest Group was formed in 2005 in response to the significant funding for University-level computing funded primarily by the SRIF-3 funding round. Members are drawn primarily from Computing Services in the Higher Education sector with representation from related organisations such as the National Grid Service and funding bodies.

IDC: the International Data Corporation is a global provider of market intelligence, advisory services, and events for the information technology, telecommunications and consumer technology markets.

JANET: JANET is the network dedicated to the needs of education and research in the UK. It connects the UK's education and research organisations to each other, as well as to the rest of the world through links to

the global Internet. JISC: The Joint Information Systems Committee is the UK's expert on information and digital technologies for education and research

MPI: the Message Passing Interface, is standardized and portable message-passing system designed by a group of researchers from academia and industry to function on a wide variety of parallel computers.

NHK: *Nippon Hōsō Kyōkai*; official English name of the Japan Broadcasting Corporation

Peta: prefix meaning 10^{15} as in petabytes (a unit of storage) or petaflop (10^{15} floating point operations per second, a unit of computing speed)

SME: a small to medium sized enterprise.

Software: the computer codes required to perform simulation, modelling, data analysis, visualisation on high performance computers through the implementation of algorithms

STFC: the Science and Technology Facilities Council is an independent, non-departmental public body of the Department for Business, Innovation and Skills (DBIS). The Council has a broad science portfolio and works with the academic and industrial communities to share its expertise in materials science, space and ground-based astronomy technologies, laser science, microelectronics, wafer scale manufacturing, particle and nuclear physics, alternative energy production, radio communications and radar.

VFX: visual effects (commonly shortened to Visual F/X or VFX) are the various processes by which imagery is created and/or manipulated outside the context of a live action shoot. Visual effects using computer generated imagery (CGI) have become increasingly common in big-budget films.

References

[i] M. Faraci, *American Business's Secret Competitive Weapon: HPC,* at
http://www.forbes.com/2009/05/22/high-performance-computing-leadership-managing-hpc.html

[ii] M. Sawyer and M. Parsons, A *Strategy for Research and Innovation through High Performance Computing*, Planet HPC Setting the Roadmap for High performance Computing in Europe.

[iii] Dreamworks at
http://www.compete.org/images/uploads/File/PDF%20Files/HPC_Dreamworks_072809_A.pdf

[iv] http://www.parliament.uk/business/committees/committees-a-z/commons-select/transport-committee/news/winter-weather-response-news/

[iv] P. Coveney, *Strategic plan for the UK Research Computing Ecosystsm*, 2011

[v] T. Hey, S. Tansley, and K. Tolle eds., *The fourth paradigm: data-intensive scientific discovery*. Redmond, WA: Microsoft Research; http://research.microsoft.com/en-us/collaboration/fourthparadigm/2009. 2009

[vi] P. Szegedi, *NREN' strategic perspective on storage and Cloud "Build or Buy", (Terena)*. 2011,

[vii] P. Cajella, C. Gardiner, C. Gryce, M. Guest, J. Lockley, O. Parchment, and I. Stewart, *HPC-SIG Report 2010,* 2010.

[viii] P. Williams, *Report of the e-Infrastructure Advisory Group*, report to RCUKEG and Professor Adrian Smith BIS, 2011

[ix] E. Joseph, S. Conway, C. Ingle, G. Cattaneo, C. Meunier, and N. Martinez, *Strategic Agenda for European leadership in Supercomputing IDC Final report of the HPC study for DG Information Society of the European Union*. 2010

[x] PRACE. Partnership for advanced computing in Europe, http://www.prace-ri.eu/ , 2011

[xi] JISC, *Data centres: their use, value and impact*, http://www.rin.ac.uk/data-centres, 2011

[xii] C. Willard, A. Snell, S. Gouws Korn, and L Sergevall, *High Performance Computing Forecast for 2011 through 2015. Economic Sectors and Vertical Markets*, published by Intersect360 Research, 2011.

[xiii] *Going for Growth*, OECD, 2009, highlights that investment in physical infrastructure increases long-term economic output more than other kinds of physical investment.

[xiv] http://www.hm-treasury.gov.uk/d/nationalinfrastructureplan251010.pdf

[xv] J. Atkinson, R. Kärrberg, P. Castro, D and Ezell, S, *UK's digital road to recovery*, Liebenau 2009.

[xvi] http://www.ukti.gov.uk/uktihome/aboutukti/aimsobjectives/corporatestrategy.html

[xvii] E. Joseph, S. Conway, L. Cohen, and C. Hayes, *IDC HPC Market Update*, June 2011.

[xviii] Reference the Minister's recent announcement at the Conservative Party Conference

[xix] IETF Abfab, http://tools.ietf.org/wg/abfab/, 2011

[xx] Project Moonshot, http://www.project-moonshot.org/, 2011

Acknowledgements

I would particularly like to acknowledge the support of Mike Ashworth, STFC, Paul Best, CFMS, Peter Coveney, University College London, Ian Dix, AstraZeneca, David Docherty, Digital Media Group, Andy Grant, IBM, Tony Harper JLR, Doug Kell, BBSRC, Martin Ridge, BIS, David Salmon, JANET (UK) and Lesley Thompson, EPSRC.

I would also like to thank Oz Parchment (Southampton University), Richard Blake (STFC Daresbury), John Bancroft (STFC Daresbury), Dai Jenkins (EPSRC Swindon), Susan Morrell (EPSRC Swindon), Mark Parsons (EPCC Edinburgh University) and all those who attended the e-infrastructure seminars at BIS, 1Victoria Street, on the 13th July and 21st September 2011.

The image of Nature on page 11 is reprinted by permission from Macmillan Publishers Ltd: Nature, copyright 2005.