Institution: University College London (UCL)

Unit of Assessment: 8 – Chemistry

Title of case study: Impact on UK Government funding and decisions relating to the improvement of UK e-infrastructure

1. Summary of the impact (indicative maximum 100 words)

In 2011, a leading role was given to Peter Coveney in UCL's Department of Chemistry in defining the future strategy for the UK's e-infrastructure, based on the department's expertise and research in this field. This appointment led to the publication of the *Strategy for the UK Research Computing Ecosystem* document, which has since stimulated debate amongst policy makers and informed government policy. On the basis of its recommendations, the government has set up an advisory E-Infrastructure Leadership Council and allocated £354 million to improving the UK's high-performance computing capabilities and wider e-infrastructure, a move that is having wide-ranging industrial and economic impact in the UK. Most recently, in June 2013 the *Strategy* document stimulated further debate about the UK's e-infrastructure at the House of Lords.

2. Underpinning research (indicative maximum 500 words)

E-infrastructures are connected systems of computational technology and resources that can facilitate distributed collaboration and computation. A single e-infrastructure may involve the interoperation of high-performance computing (HPC), cloud computing, mid-range computer systems, databases, networks, image resources, instruments, software and people, which may all be geographically dispersed across a nation or even across the world.

A research group in UCL's Department of Chemistry makes use of e-infrastructures to access globally distributed computational resources, including some of the world's largest supercomputers. In addition to carrying out their own research using computational science, the group also researches the use and development of computational resources. This includes the development of tools that enable users to make use of e-infrastructures more easily.

The unification of distributed computational resources within and between e-infrastructures to solve challenging problems is known as grid computing. It is conducted through the use of software – referred to as middleware – which ties together the resources from the e-infrastructures that constitute the grid; however, these middleware are usually complex and difficult for end users to deploy, with the result that the full power of computational grids is rarely realised. To tackle this problem and encourage the integrated use of dispersed resources, the UCL group developed the Application Hosting Environment (AHE), a lightweight (i.e. simpler than other middleware tools) but powerful middleware tool that hides much of the complexity of the grid from the user.

The first version of the AHE was developed in 2006 by the UCL group and collaborators at the University of Manchester (McKeown and Pickles) [1]. The Manchester team helped to define the architecture; the implementation and deployment were performed by UCL. Version 2.0, with enhanced features and usability, was developed by the UCL group's Peter Coveney (Professor 2002-present) in 2009 [2]. The AHE consists of an ensemble of programs written in Perl with command-line and Graphical User Interface client tools, and is a form of "non-invasive" middleware, meaning that it is not necessary for grid management teams to usually even be aware that the AHE is interfacing with the lower-level middleware resident on the different e-infrastructures. The AHE provides the user with a single interface to all computing resources (e.g. different supercomputers) on which they have accounts, enabling them to perform all manner of complex workflows on these resources, from the most basic form of single job submission to complex orchestrated sets of applications.

Previous research conducted by the UCL group included international leading work on soft matter and computational chemistry, all using high-performance computing, visualisation and computational steering to make new discoveries in these domains of the chemical sciences. The AHE was developed to facilitate these investigations, and has since been exploited by the UCL group to conduct further research pertaining to the chemical and biomedical sciences, which is leading to broad impact in medical applications. For example, it enabled them to use grid

computing resources in both Europe and the US in 2008 to apply molecular dynamics techniques to study the structure and stability of DNA intercalated in layered double hydroxides (LDHs) [3]. Their simulations indicate that DNA intercalated into LDHs is endowed with enhanced structural stability, a finding which gives support to the origins-of-life theory that LDHs could have acted as a protective environment for the first nucleic acids in extreme environmental conditions. In another example, in 2008-10 the AHE enabled the group to use combined UK and US computational resources to estimate and rank the binding affinities of inhibitors with the wild type and variants of HIV-1 protease, an attractive target for HIV/AIDS therapy [4, 5]. This research showed that molecular dynamics can achieve accurate and rapid relative binding affinity ranking. Indeed, the concerted use of computing resources and automated workflow enabled through the AHE make turnaround of an array of binding affinities feasible within three days; such rapidity is crucial for enhancing patient-specific clinical decision support.

UCL owes its prominence in the field to the strength of its published research, which has led to Coveney and others being appointed to leadership positions relating to the use of computers in research, including Coveney's chairmanship of the Collaborative Computational Projects Steering Panel (since 2006) and membership of the UK High-End Computing Strategy Committee (2006-2010). These pivotal positions as a spokesperson for the community have enabled the impacts described in section 4.

3. References to the research (indicative maximum of six references)

[1] The application hosting environment: Lightweight middleware for grid-based computational science, P. V. Coveney, R. S. Saksena, S. J. Zasada, M. McKeown and S. Pickles, *Comp. Phys. Commun.*, 176, 406-418 (2007) doi:10.1016/j.cpc.2006.11.011

[2] Virtualizing access to scientific applications with the Application Hosting Environment, S. J. Zasada and P. V. Coveney, *Comp. Phys. Commun.*, 180, 2513-2525 (2009) doi:10.1016/j.cpc.2009.06.008

[3] Computer simulation study of the structural stability and materials properties of DNA-intercalated layered double hydroxides, M. A. Thyveetil, P. V. Coveney, H. C. Greenwell and J. L. Suter, *J. Am. Chem. Soc.*, 130, 14, 4742-4756 (2008) doi:10.1021/ja077679s

[4] Rapid and accurate prediction of binding free energies for saquinavir-bound HIV-1 proteases, I. Stoica, S. K. Sadiq and P. V. Coveney, *J. Am. Chem. Soc.*, 130, 8, 2639-2648 (2008) doi:10.1021/ja0779250

[5] Accurate ensemble molecular dynamics binding free energy ranking of multidrug-resistant HIV-1 proteases, S. K. Sadiq, D. W. Wright, O. A. Kenway and P. V. Coveney, *J. Chem. Inf. Mod.*, 50, 890-905 (2010) doi:10.1021/ci100007w

References [3], [4] and [5] best indicate the quality of the underpinning research.

4. Details of the impact (indicative maximum 750 words)

Prior to the impacts described in this section the UK's e-infrastructure suffered from a fragmented funding regime, with resources being funded by a variety of bodies representing diverse user groups. There was also a distinct lack of a coordinated policy and strategy for UK e-infrastructure, and no single coordinating body that represented all relevant stakeholders. This fragmented organisation of the national infrastructure, together with no guarantee of future provision, discouraged business from becoming involved, inhibited collaboration between organisations and resulted in the UK losing its ability to compete internationally in the field.

In 2011, as a result of the strength of UCL research in the field, Coveney was asked by the UK e-Science Institute to lead on the publication of a report into the UK's e-infrastructure, with the objective to develop a strategy to overcome the problems of fragmented funding and coordination. The *Strategy for the UK Research Computing Ecosystem* report [A], published in October 2011, was the result of a number of meetings (with the involvement of UK Science and Universities Minister Rt Hon David Willetts MP in July 2011). It set out a holistic approach for the UK's e-infrastructure involving the different Research Councils, funding bodies and user communities in a

coherent collaboration. Non-university users of the e-infrastructure include automotive, pharmaceutical, software and hardware industry sectors; charity-funded research laboratories (for example, the London Research Institute); and many publicly funded research institutes, including the Met Office, Culham and AWE (Atomic Weapons Establishment). Specific recommendations made in the *Strategy* report were to (i) set up an independent coordinating body to govern the e-infrastructure that reports directly to the Government, (ii) drive the development of software, hardware and networking for all stakeholders (including national supercomputing service providers and Research Councils) in the e-infrastructure, and (iii) offer e-science training to develop the next generation of computationally aware researchers who are able to use the improved e-infrastructure. Coveney's role in directing and shaping these recommendations was directly consequent on the UCL team's research in computational science, including computational chemistry.

This document in turn fed into and informed a key policy document entitled *A Strategic Vision for UK e-Infrastructure*, commissioned by Willetts and published by the Department for Business, Innovation and Skills (BIS) in January 2012 [B]. This document addressed the advanced computational capacity of the UK as a system (i.e. as an e-infrastructure), and proposed a series of recommendations on software, hardware, networks, data storage and skills to help develop this e-infrastructure. In particular, the *Vision* report took forward the specific recommendations made in the *Strategy* report, demonstrating that the *Strategy* report informed Government understanding of the issues. As a direct result of the first of these recommendations the Government created the UK E-Infrastructure Leadership Council (ELC) in March 2012 [C]. The ELC has the responsibility to advise the Government on e-infrastructure development and to develop a wider plan for stakeholder engagement. It is currently working on mechanisms for engaging industry and healthcare with the newly funded e-infrastructure, including an "On Ramp" programme that seeks to identify and remove barriers to the uptake of e-infrastructure by UK PLCs. The strong influence of the *Strategy* report on the *Vision* report is clearly evidenced in the Chairman's foreword and the references of the latter report [B]. For example, the Chairman wrote: "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. […] 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*" [B].

As a result of these two policy documents (mainly the *Strategy* report, which was already published at the time), in October 2011 the Government earmarked £145 million (increased to £165 million in early 2012) to improve the UK's HPC capabilities and wider e-infrastructure [D, E]. This investment has resulted in improvements to existing e-infrastructure such as the UK academic network JANET and high-capacity data storage across the Research Councils, and has funded a number of HPC activities in the UK including Emerald, one of the largest Graphics Processing Unit systems in Europe, now located in the Rutherford Appleton Laboratory in Daresbury, and IRIDIS, a conventional 100-teraFLOPS supercomputer, located at the University of Southampton. These activities have vastly increased the computing capacity and capability in the UK. This has wide-ranging industrial and economic impact, including allowing the NHS and UK PLCs to use these resources, and attracting hi-tech industry from overseas. For example, companies including IBM, Intel and Cisco have already made investments into the UK following this Government investment in e-infrastructure [F].

The *Vision* report (and therefore the *Strategy* report) has further influenced Government spending, with the Government's on-going commitment to improving UK e-infrastructure demonstrated by the allocation of an additional £189 million in capital funding to this area in December 2012. This investment was part of the £600 million capital funding for science announced in the Chancellor's 2012 Autumn Statement, and the amount allocated to e-infrastructure was communicated by Willetts at Policy Exchange in January 2013, where he described e-infrastructure – or "Big Data" – as one of the "eight great technologies" which will propel the UK to future growth [F]. The £189 million, to be invested over the next two years, has been allocated to areas including "e-Infrastructure for Biosciences" and "Administrative Data Research Centres" [G]; it was announced in February 2013 that £30 million has already been awarded to the Hartree Centre in Daresbury and its academic and industrial partners, in part to develop new technology to make

supercomputers more efficient [H].

The work of the UCL team has also informed policy debate. In June 2013, Coveney was invited to speak at a House of Lords meeting about scientific infrastructure [I], as part of an inquiry into that topic launched by the Science and Technology Committee in May 2013. He reminded the audience of the recommendations for UK e-infrastructure made in the *Strategy* report, and explained that whilst the initial Government investments and formation of the ELC are excellent outcomes of this report, there are a number of critical recommendations that have yet to be implemented. These include holistic management of investments in e-infrastructure, a focus on software development in addition to investments in hardware, and the provision of appropriate training across a wide range of disciplines. His presentation sparked considerable discussion and debate amongst meeting attendees about these and other issues relating to UK e-infrastructure.

5. Sources to corroborate the impact (indicative maximum of 10 references)

[A] *Strategy for the UK Research Computing Ecosystem* (2011), available online: http://wiki.esi.ac.uk/w/files/f/f5/ResearchComputing-glossy.pdf – corroborates the recommendations of the report and Coveney's role in producing it.

[B] *A Strategic Vision for UK e-Infrastructure* (2012), available online: http://www.bis.gov.uk/assets/BISCore/science/docs/S/12-517-strategic-vision-for-uk-e-infrastructure.pdf – corroborates the influence of the *Strategy* report on the *Vision* report.

[C] ELC website: https://www.gov.uk/government/policy-advisory-groups/122 – corroborates the creation of the ELC in March 2012 and what it does.

[D] *BIS: Innovation and Research Strategy for Growth* (2011), available online: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/32450/11-1387-innovation-and-research-strategy-for-growth.pdf – corroborates that the Government allocated £145 million to e-infrastructure in October 2011.

[E] Transcript of David Willetts' oral statement "Science and the City" (2012), available online: https://www.gov.uk/government/speeches/science-and-the-city--2 – corroborates the increase in Government funding for e-infrastructure to £165 million in early 2012.

[F] Transcript of David Willetts' speech at Policy Exchange (2013), available online: https://www.gov.uk/government/speeches/eight-great-technologies – corroborates that the Government allocated £189 million "Big Data" in January 2013, and corroborates that investments into the UK have been made by companies including IBM, Intel and Cisco.

[G] Research Councils UK press release (2013), available online: http://www.rcuk.ac.uk/media/news/2013news/Pages/130124.aspx – corroborates how the £189 million has been allocated.

[H] George Osborne press release (2013), available online: https://www.gov.uk/government/news/chancellor-announces-boost-for-state-of-the-art-computing – corroborates the award to the Hartree Centre of £30 million from the £189 million.

[I] Agenda for Scientific Infrastructure seminar at House of Lords (11 June 2013) – corroborates the attendance of Coveney at the meeting and the discussion questions. Available on request.