Prometheus, Vol. 25, No. 1, March 2007

e-Research Infrastructures and Open Science: Towards a New System of Knowledge Production?[1]

RALPH SCHROEDER

ABSTRACT *Major efforts are currently under way to develop e-infrastructures for research. At the same time, there have been a number of calls for research in the digital realm to foster 'openness', for example in 'open access' policies. This paper explores the relation between e-infrastructures and open science, and argues that there are a number of factors apart from research policy that will shape e-research. These include not only the legal and economic environment, but also the ethos of science, 'open' initiatives outside of research, the momentum of large technological systems, and the activism of experts and wider social movements. The paper assesses the strengths and weaknesses of these factors, as well as the tensions and confluences between openness and e-infrastructures.*

Keywords: e-research; infrastructure; open science; social implications of technology; Internet and research; research policy.

Introduction

This paper presents an overview of initiatives to develop e-infrastructures and to promote open science. There are several factors that shape these emerging technological infrastructures for research and their openness. These include the norms or ethos governing scientific research, the relation to 'open' initiatives outside research, the dynamic of large technological systems, and the groups promoting openness as part of a wider agenda. Gauging the prospects of open e-infrastructures for research entails going beyond the economics of innovation and the legal framework of intellectual property which have already been widely discussed. A broader view will also consider infrastructures from the perspective of the sociology of science and technology and will focus on the convergence and divergence between the initiatives of different social groups and movements. The paper argues that this broader approach is essential for assessing the strengths and diffuseness of different components of the agenda where 'open science' and e-research meet.

Prometheus ISSN 0810-9028 print/ISSN 1470-1030 online
http://www.tandf.co.uk/journals
DOI: 10.1080/08109020601172860

Three further arguments are made: one is that the 'openness' of e-research infrastructures already has considerable momentum, but there are also major limits to their extension which are mainly a product of the limited attention space in research. A second is that although research policy can contribute to steering e-infrastructures towards openness, policymakers must also recognize that other factors shaping these e-infrastructures have dynamics of their own, including those that may limit openness. Finally, because of the relation between technological and scientific advance, the technological infrastructure and its open organization are both needed if they are to succeed in enhancing knowledge production. That is, openness as a norm critically underpins the widest possible extension of e-infrastructures, just as a more accessible technical infrastructure will produce more effective knowledge dissemination; and the opposite, a fragmented infrastructure with limited access will yield narrower social benefits and slow the advance of knowledge.

Definitions and Aims

In recent years, there have been a number of initiatives to support collaborative research via electronic networks. As we shall see shortly, these have gone under various labels, including e-Science, cyberinfrastructure, e-infrastructure and the Grid. Since this is a rapidly developing area in which terms are still used in a fluid way, it is important to provide clear definitions at the outset which delimit the phenomena under discussion. In this essay, 'e-research' will be used as a generic label to encompass research under all of the labels just mentioned, but also defined specifically as the use of advanced computing tools and high-powered networks for collaboration and sharing resources in scientific or academic research. This collaboration and sharing can take place between researchers, between projects, between institutions, between disciplines and between different locations. e-Research, however, will mean here not just putting research online, but rather using leading edge computing tools—including shared databases and instruments, tools for distributed work, and shared computing resources—to foster collaboration.

Recently, there have also been many 'open' initiatives. Open source software is perhaps the best-known, but there are various others, including in fields such as business, government, and cultural production. As this paper is limited to research, and to give the term some analytical purchase, 'open science' is used here to designate research efforts that aim to contribute to a resource that is shared among scientists or researchers. 'Open' will be used specifically to designate unrestricted access and use, being free of charge to users, and using non-exclusionary (open) standards.[2]

The term 'e-infrastructure' will also be used in a narrow sense to designate the technical system, akin to a large technological system or infrastructure (to be discussed further below), that supports it. Again, to define the term more precisely, e-infrastructure will be taken in the first instance to designate the physical or material components of this technological system, the advanced electronic networks that make use of the Internet and the Web, as well as, secondarily, the organizational networks that are supported by this system.

This essay therefore centrally deals with a new set of research initiatives, how ideas and practices about openness are shaping this research, and the infrastructure that will emerge from the interaction between them. Since the early shaping of e-infrastructures is likely to have a profound impact on research in the decades to come, as with other large technological systems, it is important to identify the constraints and possibilities in this process early on. To do this, the paper will first give an overview of the various e-research and e-infrastructure initiatives and their promotion of openness. Then it will provide an analytical discussion of the factors influencing these initiatives, beginning with the norms of openness in science and in the next section some related initiatives towards openness. This is followed by a discussion of how 'open' and 'closed' forces are ranged against each other. The paper then moves on to infrastructures; how this concept applies to e-research and its significance for contemporary science. This significance is then put into a larger context which identifies the potential obstacles and limitations to the extension of infrastructures for research. The essay concludes by gauging the prospects of open e-research infrastructures.

e-Research and Research Policy

e-Research has been promoted under different labels in different contexts. In the United States (US), e-research mainly goes under the name of 'cyberinfrastructure', in the United Kingdom (UK), the term 'e-Science' is used, and the European effort tends to be labelled 'e-infrastructure'.[3] These efforts are relatively recent, emerging in the last 5–10 years, but they are assuming an ever greater part of the agenda of research funding agencies and policymakers. Regardless of the different labels, the policy documents and guidelines for research are in agreement about promoting 'openness', and this promotion of openness is clearly shaping research.

In the US, for example, there is now a separate 'Office of the Cyberinfrastructure' within the National Science Foundation. The recently appointed director of this organization, Dan Atkins, was also part of a group that authored a report, known as the 'Atkins report', which has become an influential policy document in e-research.[4] The Atkins report clearly promotes open science. For example, it advocates 'open platforms'[5] and talks about the Grid as an 'infrastructure for open scientific research'.[6] Further, the report argues in favour of 'mutual coordination among directorates and … common standards and … a common infrastructure'[7] and mentions 'open source'[8] software as an example.

The Atkins report, which dealt with science and engineering, was followed by the Berman and Brady report which sought to extend the Atkins report to the social and behavioural sciences.[9] And although this report does not talk about open source or standards as such, it endorses the Atkins report and stresses a 'shared Cyberinfrastructure'[10] and provides many examples of sharing resources such as common databases. Finally, a similar report for the arts and humanities by the American Council for Learned Societies includes a section on open standards and spells out various ways to promote openness, which include open source methods for software development.[11]

In the European Union (EU), the e-research initiative of the European Community consists of several parts, but it is also generally pushing for open science. In relation specifically to e-research, for example, the e-Infrastructure Reflection Group (e-IERG) consists of national representatives of EU governments whose aim is to inform policy on the developments of e-science infrastructures across the European Research Area (ERA). The latest 'e-Infrastructures roadmap' of this group advocates open source software and argues that 'current Intellectual Property Rights are not in the interests of science'.[12] Two scenarios, it goes on to argue, of a laissez-faire approach to software development or of a commercial approach to software, are inferior to a third, whereby public resources ('structural funds') would be put into developing and sustaining Grid software.[13] It needs to be emphasized that the group is not hostile to the private sector; instead, it argues that 'science and industry should work closely together in order to make sure that a set of open standards and a broad community of services that is of use to all emerges'.[14] Above all, the group argues that an infrastructure that is shared across Europe will bring benefits that are difficult or impossible on a smaller scale.

The UK arguably has an e-research effort which is second only to the one in the US since the UK also began to develop its national programme early on (in 1999). In this case, similar open policies are being supported by the National e-Science Centre, the National Centre for e-Social Science and the Arts and Humanities e-Science Support Centre. The implementation of open source software and open standards in the UK e-science community, for example, is included in the mission statement of the Open Middleware Infrastructure Institute (http://www.omii.ac.uk/). And the Joint Information Systems Committee, an umbrella body which provides information and communication technologies for the UK's education and research sector (which includes the e-science centres), funds the 'OSS Watch', the 'Open Source Advisory Service' (http://www.oss-watch.ac.uk/) which promotes open source software in UK higher education.

In addition to research funding bodies, a number of non-governmental organizations (NGOs) are pushing for open science on the level of research policy. One example will suffice: Science Commons (http://sciencecommons.org/) is an NGO whose aim is advocacy on behalf of open e-science. Science Commons comes under the umbrella of the Creative Commons organization, which promotes copyright licensing that is 'open' in the sense that it offers 'commons' alternatives for those who might otherwise use conventional copyright and instead uses licensing that preserves open access. In other words, it promotes a form of licensing that is an alternative to for-profit copyright and patenting, thus making it easier to share scientific knowledge without forcing it into more exclusive copyrights or patents. In fact, there are now several different types of creative commons licenses which differ in the strictness with which this 'openness' or 'freedom' is interpreted. Science Commons is one among several Creative Commons initiatives and it is itself subdivided into several more specific areas of concern (publishing, data, and licensing). What unites these specific areas is the aim of promoting open access to and use of scientific knowledge.

Science Commons, in turn, has links with several other organizations promoting similar aims. A number of institutions such as publishing repositories and NGOs support and have adopted Science Commons aims. A related effort, for example, is the recent 'Global Information Commons for Science Initiative' by the International Council for Science's Committee on Data for Science and Technology (CODATA) which aims to promote better understanding and coordination towards an '"information commons" for global e-science' (http://www.codata.org/resources/newsletters/newsltr91A4.pdf).

Many other examples of openness policies among research funding agencies and NGOs could be given, both within the US/EU/UK and beyond. It can be seen that these policies are almost identical in their thrust and aim to promote openness in all the above-mentioned four senses. To what extent these policies are implemented in individual research projects remains to be seen,[15] but this will also depend on a number of factors outside the realm of funding bodies and research policymakers.

Science and the Norm of Openness

Although 'open science' is a relatively recent term, it is possible to argue that modern science has always been 'open'.[16] Merton proposed that the 'ethos of modern science' consists of 'four sets of institutional imperatives—universalism, communism, disinterestedness, organized skepticism'.[17] In the context of open science and e-research, the most relevant of these is communism, by which Merton did not mean the political ideology, but rather more generally the 'common ownership of goods': 'The scientist's claim to "his" intellectual "property"', he argued, 'is limited to that of recognition and esteem which … is roughly commensurate with the significance of the increments brought to the common fund of knowledge'.[18]

Within the sociology of science and technology, Merton's account of the norms of science has been extensively criticized. Hess has reviewed these criticisms and concludes that 'it is possible to salvage Merton's delineation of the norms of science, but only as a prescription of how scientists should behave ideally';[19] in other words, these norms do not reflect actual scientific practice. Others have argued, however, that, with some refinement, they sometimes *do*.[20] In the case of e-research infrastructures, these norms of scientific enquiry are perhaps less important than the fact that they underpin a system of scholarly communication (in a broad sense which includes not just publication, but also access to the tools for research). And in this respect, Merton's norms of 'universalism' and 'communism' are part of the broader openness of the research process as it has been defined here.

Apart from Merton, some other sociologists of science have argued that openness is an essential institutional characteristic of scholarly communication and thus a necessary condition for the advance of scientific knowledge. That is, scientific knowledge must remain open to being improved upon via new communication because closing off communication would prevent the ongoing refinement of knowledge.[21] The question whether e-infrastructures are 'open'—in the sense of the 'communism' of shared resources and the 'universalism' of open communication—and live up to the institutional norms of science is thus a key question.[22]

The Varieties of Openness

Apart from research policy and the norms of science there is also, however, a much broader impetus towards openness. In fact, there are so many 'open' initiatives now that it is difficult to keep track. Thus 'openness' movements can be found in the realms of culture and media, politics and political activism, and business and economics. Here it will suffice to give just a few examples which do not belong to the realm of academic research but are closely related to it. All of these projects or initiatives meet one or more of the four aspects of the definition of openness given earlier, and some meet all four, including being free of charge to certain institutions and sometimes to anyone.

In the commercial realm, for example, there is Open Business (http://www.openbusiness.cc/), an organization which promotes business models and ideas related to creative commons, open source software, and free access and use. A major aim is to act as a network for disseminating knowledge about alternatives to business practices that rely on restrictive intellectual property rights. Similarly along the lines of a knowledge commons, there is the Open Knowledge Network (http://www.openknowledge.net/) which promotes tools to support local knowledge across the global South and has a number of projects that use open source and open standards to engage in this work. In terms of access to knowledge, the Open Archives Initiative (OAIster http://oaister.umdl.umich.edu/o/oaister/) is a digital libraries project to provide better availability for difficult-to-access digital resources and to encourage means of finding these through common open metadata standards.

Apart from access, a key question is openness to contributions. For example, although primarily a network of researchers, the International Virtual Observatory has begun to open up participation to amateur astronomers, by allowing them to contribute to open data repositories (http://www.ivoa.net/). The most widely known example of open contributions, of course, is Wikipedia (http://en.wikipedia.org/wiki/Main_Page), which can also be regarded as an open access resource. This kind of openness as participation is different from openness for dissemination and widening benefits, for example, in licensing for pharmaceutical development to ensure that affordable medicines for infectious diseases can be developed for the developing world. This is what the Institute for One World Health (http://www.oneworldhealth.org/) is trying to do.

The most well-known example of an 'open' initiative is open source software. Open source is critical to e-research and e-infrastructure since open source software tools are important shared resources in many of these efforts. Open source has also, according to Weber, gone furthest in addressing some of the legal and business-related issues of e-research.[23] One question is therefore to what extent open source collaboration and software practices can provide a model and viable tools for an open e-infrastructure.[24]

Open source is the attempt to create software that can be freely used and developed within terms of a license that safeguards continuing free use and access. A fundamental distinction is between software that is 'purely' non-commercial (it contains no proprietary code and stipulates that all code must remain non-proprietary) and software that allows proprietary—or commercial—appropriation and development. The advantages of different models, whether it is best to develop so that it can be commercially exploited, or to stay 'pure', have been subject to extensive debate.[25]

Openness and the open source model can also be seen as a social innovation (or set of innovations), diffusing into different arenas. So, for example, sourceforge (http://sourceforge.net/), a repository for open source software code, has spread and been applied, for example, to bioforge (http://www.bioforge.net/forge/index.jspa), which aims to apply a similarly open means of providing access to biotechnology research.[26]

Other initiatives promoting open access to knowledge, especially in publishing and related areas, are too numerous to chart.[27] These include libraries, texts, data and other digital objects. Willinsky also points to the similarities between open science, open access and open source software.[28] Much like open source software, he argues, which is open to contributions and continual refinement from a larger community, research, too, can be seen as a product of continually building on the openly published achievements of other researchers and improving on them. Willinsky concludes that 'the current convergence among various *open* approaches to intellectual property represents a common commitment to larger public sphere. These approaches extend well beyond the university and yet it remains the primary institutional force in sustaining this open economy'.[29] At the same time, collecting these efforts under one umbrella is not intended to give the impression here that they are part of a concerted or centrally organized effort. In fact, a key feature of open science is that it consists of disparate initiatives, mainly by researchers and academic institutions and NGOs.

Open versus Closed

Against the spread of open research policies and related initiatives, there is also a contrary trend, towards securing intellectual property rights (IPR) in such a way that ownership is protected; in other words, that access to data and publications, the use of software and tools, and also standards should conform to a proprietary or exclusive regime.[30] This is sometimes referred to as the new 'enclosure movement'[31] on the analogy with what is often regarded as the origin of private property in the first enclosure of public (commons) agricultural land.

There are a number of reasons why protecting IPR is currently moving to centre stage. One is that whereas before, patents and copyright were subject to national jurisdictions with international agreements for enforcement beyond borders, it has in recent years become necessary to ensure that this enforcement extends globally, and especially to the developing world and for those forms of intellectual property such as patents for medicines or copyright for entertainment (music and video material) which may not have enjoyed this protection or where it is being challenged. Another reason is the increasing pressure to commercialize research, including research funding pressures for academic research to produce 'added value'.

Some have argued that a narrow constituency of private interests has been winning the battle to enshrine global enforcement of commercial IPR.[32] At the same time, as we have seen, moves towards open research and openness in other domains have also been gaining strength. To be sure, the two do not necessarily conflict or confront each other directly, so that proprietary commercial research, for example, could co-exist with openly accessible research results on the same topic. At the same time, a battle is currently shaping up between

> two distinctive regimes or environments for the conduct of research: the actors in the realm of 'open science research' expect reciprocal sharing of discoveries among themselves and the rest of the world, while those in the world of private profit-oriented and proprietary R&D expect to receive payment for the right to use their inventions (and to pay others for the use of theirs).[33]

A key question, therefore, is whether e-research infrastructure will implement openness in the tools that are being developed.

The movement towards openness in research can thus be seen as a legal battle, but it also resonates with a larger ideological or cultural movement that is shaping the agenda about globalization, social justice and common resources. This larger environment is bound to affect policy and practice in the research community and the technological infrastructure that supports it. One reason why a wider social context is crucial is that although science has been an autonomous institution in society since the late nineteenth century with the emergence of the modern research university, nowadays science needs allies.[34] This is partly because the equipment for carrying out scientific research in the age of big science and large technological systems has become very expensive, and partly because of the growing controversies around science.

There are a number of actors within this larger social movement. Most narrowly, they consist of researchers, and Haas has coined the term 'epistemic communities' to characterize expert activism. He defines epistemic communities as a 'network of professionals with recognized expertise and competence in a particular domain and an authoritative claim to policy-relevant knowledge within that domain or issue-area'.[35] More broadly, they consist of the 'activists beyond borders' who are networked within NGOs to promote social justice.[36] And more broadly still, they also include the mid-level government officials described by Slaughter and who, she argues, constitute a new force in international relations building global networks to make internationally binding agreements—thus exercising 'soft power'.[37]

The movement towards openness aims to maximize the benefit of research, and in this 'advocacy' mode, researchers and others will stress the social and economic benefits of science for society. In doing this, they are supported by wider movements in society and by a public receptive to science that is seen as a 'public good'. Whether the engagement between science, the public and the wider society in this case should be regarded in terms of a conflict or controversy between these groups (mainly over resources), as a conflict sociologist might see it,[38] or in terms of a functionalist framework whereby interest groups and the research community need to be aligned in a non-zero-sum game, remains to be seen.[39] In assessing the significance of the agenda of this global social movement it is important to recognize, however, that even when NGOs and the scientific community and social movements are worldwide, the target of these movements is nevertheless the state;[40] particularly as the state is the main source of funding of academic research and of research regulation.

Large Technological Systems, the Leading Edge of Knowledge, and e-Research Infrastructure

e-Research infrastructures are networked systems in which technologies and social institutions are intertwined. To use Hughes' term, they are 'large technological systems' which combine extensive networks of physical artefacts with the organizational capacity to implement and sustain them.[41] Two features of these systems are particularly salient: one is that the early phases of these systems are decisive for their longer-term shape and direction; and the second is that these systems take on a momentum of their own.[42]

At this point we can return to definitional questions since the term 'infrastructure' implies something broader than 'large technological systems'.[43] This is because e-infrastructures for research can be seen as a (relatively) small part of a larger communications infrastructure (the Internet and the Web) and also as a small part of a larger social infrastructure of education and research. e-Research infrastructures add to or extend these larger social (communication and education) infrastructures by building or grafting onto them the tools for collaborative research and shared resources, including data stores, steering of remote instruments, and the like. e-Research tools thus have an 'infrastructural' character in that they underpin a number of different scientific and technological (or engineering) fields, and represent an infrastructure in this sense, but they are only a new and small part of a larger whole.

e-Research infrastructures are therefore both: a large technological system insofar as they consist of a number of interdependent social and technical systemic parts (and large because the system covers the globe); and an infrastructure insofar as it supports research. Both terms can therefore be used, depending on whether the emphasis is on the technical system or the support to the research community—but the key point is that e-research infrastructures are *sub*systems within large technological systems and within social infrastructures—namely, those dedicated to e-research.

The potential benefits of this infrastructure or large technological system should therefore be seen in terms of the creation or extension of other social infrastructures or large technological systems—but *within the limits of* e-research. If we think about some other large technological systems and social infrastructures that have been developed on a national scale and beyond since the late nineteenth century—energy, transport, and communication—these have had a number of society-wide benefits. (It is worth mentioning in this context that the term 'Grid' has been used on the basis of the analogy with the electricity infrastructure or grid.) e-Infrastructures benefit research, and through research benefit society.

Bearing this limitation of e-infrastructures in mind, it is now possible to specify how they nevertheless add to and extend current research capabilities. Most importantly, e-infrastructures are currently at the leading edge of scientific and technological advance. This is because research technologies have always been critical to scientific advance, rather than the other way around. Science and technology only became closely coupled in the middle of the nineteenth century, but since then, instruments (technology) have been critical to modern 'high-consensus rapid-discovery' science because, as Collins has argued, the reproducible results created by instruments that can be cloned and refined are essential to scientific advance.[44] In the twentieth century, big science[45] and large technological systems have dominated the leading edge of 'high-consensus rapid-discovery science' because of how large-scale scientific initiatives coupled with expensive (and again large-scale) technologies have been able to transform the physical environment. In this transformation, science and technology have thus been intertwined in an 'adventure of the interlocking of representing and intervening'[46] in which the research instruments for 'big science' have become increasingly large, expensive and complex, while the development of large technological systems has gone hand in hand with the organization of large-scale scientific research.

More recently, the Internet/Web as a large technological system has become an essential research instrument in driving research. If we focus on the most prominent current example, biotechnology, which has arguably been at the leading edge of scientific advance in recent decades, the developments in this field are unthinkable without the enhanced data processing capacities of computing, and more recently a network in which new data can be stored and accessed.[47] Put differently, machines have been driving knowledge, and biotechnology is only the current example of how research instruments, including high-performance computing and powerful networks of information and storage, are essential to scientific advance. Moreover, biotechnology is clearly 'big science' and increasingly dependent on e-infrastructures.

The question about the intellectual and organizational shaping[48] of this new part (or subsystem) of the large technological system and infrastructure is therefore fundamental to scientific advance. The degree of 'openness' of this infrastructure partly depends on its social shaping, as evidenced in the research policies and in the various open initiatives described earlier. But this shaping is also intellectual because these powerful research instruments have a dynamic—or momentum—of their own, and follow the extension of other big science initiatives, large technological systems and infrastructures as models. And yet the openness or closedness of these systems or infrastructures at this stage is clearly one of the bottlenecks—or 'reverse salients' as Hughes[49] prefers to call them—the resolution of which will be a key switchtrack which determines their future direction.

This is an important point, because those who argue that social institutions (legal, policy, economic drivers) shape open e-research[50] overlook the scientific and technological momentum inherent in this system. Research will increasingly move online and rely on large-scale computing because this is a more powerful means of automating the research process. And while the system will need to be adapted to different social institutions, the reverse is also true.

One way of conceptualizing e-infrastructures then is not just in terms of enhancing economic growth and innovation, but also as supporting knowledge production. There are extensive debates over the extent to which developed societies are characterized by knowledge production, and whether the organization of knowledge production has changed in recent decades.[51] For our purposes it is sufficient to note that considerable resources continue to be devoted to developing e-research and the large technological system that supports it; which is one indication that this system is deemed to be essential for creating what is potentially a new system of knowledge production.

The Limited Impact of Open e-Infrastructures

There are several limitations to the extension and impact of openness and e-infrastructures. The most important—and to my knowledge completely overlooked—is the limitation on the attention space. The concept of attention space on the research frontier has been used to analyse intellectual change as a whole,[52] but it is necessary to adapt this notion to the realm of e-research. This can be done by noticing that whereas 'attention space' applies to knowledge or ideas, similar ideas have been applied to technology under the labels of lock-ins, path-dependence, and different types of monopolies.[53] And although these are not identical, they all go against the thrust of a characteristic of openness that is often taken for granted in the debates discussed so far: namely, that it is possible to create new systems without displacing existing ones and that open research is infinitely expandable and yet imposes no restriction on other systems or parts of systems.

In the world of research, moreover, the status order which governs what counts as knowledge relates not just to the *amount* of attention, but also to *who* pays attention. In academic research in particular, this status order is well-entrenched. The same applies, to come back to e-research infrastructures, to electronic networks: these networks must lock users (or 'customers') in, or they must dominate the 'attention space' to obtain competitive advantage in the market for symbolic goods or tools; in this case, scientific knowledge or instruments in their various forms. Current debates about the 'sustainability' of e-research can be interpreted in the light of this limitation: only those tools or projects that gain enough attention and users, or those tools that are deemed useful for making research advances, can be sustained with resources.

This points to the truth in what sceptics say about e-research and e-infrastructures; that many of these initiatives lack users and will atrophy. In other words, many resources and tools will be eliminated from lack of attention and the inability to sustain themselves with sufficient resources, as is true of much research and new technology generally. This limitation, however, is also useful for allowing us to see the flipside more clearly: the success of openness will depend on generating and sustaining the largest possible user- or attention-base.[54] It is possible to take this point even further: there are certain areas of scientific advance that are tightly coupled to—or highly dependent on—the research instruments they use.[55] In areas that are more loosely coupled, it may be the case that a number of these instruments can exist in parallel, but they may also be harder to sustain if research does not depend on them.

At the same time, the attention space in academic research is not necessarily restricted in a zero-sum way. New technology creates spaces and newly expanding domains (for example, new specialisms or subdisciplines) that will, again, complement and add to existing ones—but again, these spaces fill up. In the case of open source software, for example, open source can complement existing software and expand to provide solutions to problems that are not yet covered by open source. Similarly, in academic research, it is possible to have increasing rates or volumes of knowledge production and increasing specialization of research outputs for e-research infrastructures to expand and extend into (for example, new online databases). The limitation is nevertheless the focus of attention on a few research results, and this effect persists online. Specifically in relation to e-research and e-infrastructures, there will be one or a few dominant databases, and similarly one or a few tools and types of software will dominate. Even open access findings in e-research will be governed by the limited online attention space within different disciplines or knowledge domains.

Limited attention space is thus the main limitation on open e-research infrastructures. There are a number of further and related limits. One is not to overestimate the impact of open e-research infrastructures (which an 'advocacy' perspective tends to do[56]): for example, researchers mainly build on each other's work, and since most research is done in institutions where researchers have access to the resources they need, whether published in open access journals or in expensive subscription-only journals, for example, makes little difference; a shift towards openness will have a limited impact on their output. Note, however, that there may be a network or bandwagon effect: if open access research is cited more, because of how researchers access such material in practice, this will reinforce the drive to publish in open access forums. Still, the much-discussed benefits, for example, of reaching a public outside of research institutions or reaching researchers in the developing world[57] will only have limited impacts.

Another limitation (that has already been mentioned in passing) is that this system or infrastructure is not entirely new. The emerging e-research infrastructure is being grafted onto existing ones—of electronic publishing, databases, instruments, networks, and the like. Like other new technologies, it complements and adds to, rather than replacing and superseding existing technologies.[58] Still, the e-research infrastructure is new insofar as it consists of more powerful networks and computing tools, and therefore the openness or otherwise of these networks and tools, or of this new part of the system, will make a difference in use: researchers and others will increasingly come to take for granted that they go online to access the materials they need—including publications, data, tools for manipulating data, instruments, and the like—but their capacity to do so will be amplified, not revolutionized, by means of the new (more—or less—open) e-infrastructure.

Yet another limitation is that although accessing and contributing content, tools or other materials to the e-infrastructure may be open to all, organizational resources are needed to enable this access and contribution, and this will impose constraints: ensuring the quality of content, software formats and standards, and accessible portals—to name only a few—have limited resources that can be brought to bear on these tasks. And the electronic network of the Internet and Web, which consists of different parts, needs to have the capacity to allow access, provide storage and be maintained—again, a question of resources.

Finally, e-infrastructure is concentrated at the leading edge of scientific advance, but this constitutes a limit too: the major effect of e-research will be where research advances are highly dependent on more powerful research instruments; for example, in biotechnology with its impact on health, food and perhaps related areas such as energy. There will be impacts on other scientific areas of advance, but there are the well-known constraints in the intellectual and social organization of the sciences that have already been mentioned; whether scientific disciplines are closely coupled and mutually dependent, and to what extent they are driven by research instruments or machines. In other words, there is variation in the extent to which scientific advance relies on computing tools (or in this case, e-research tools), and how these more powerful instruments yield scientific advance.

Conclusion

As we have seen, openness comes in many varieties and it is still gaining strength among groups and organizations. The coherence and diffuseness of the different actors and components that are determining e-research infrastructures, however, are located on different levels: in terms of e-research policy, there is a common agenda that is driven to a large extent by national funding councils and consistently promotes openness. Openness is also widely promoted as a worldview and institutional characteristic of science. The organizational movement advocating openness more widely is proliferating but also diffuse. Technical e-infrastructures meanwhile are consolidating and expanding into a more robust physical electronic network in which all constituent parts are part of a larger whole, though the extension of the open capabilities of this large technological system depends as much on organizational capabilities as on technical ones.

Nevertheless, there continue to be tensions over the extent to which e-research infrastructure will be 'open' and how much will be 'proprietary'.[59] As we have seen, this is not strictly 'either/or'; nor, as we have seen, is the attention space or the impact of these new tools a zero-sum game, though both are limited. The movement promoting openness is bound to succeed in parts, but however much it gains in extending the bounds of openness, there will also continue to be closed parts, such as proprietary data, restricted publications and networks with limited access. In terms of e-infrastructure, there will continue to be a walled garden of proprietary access and use and exclusionary costs and standards on one side, and completely open access, gratis use and open standards on the other—and the e-research infrastructure will be dominated in one or the other direction, with some parts of functional overlap. Since much of this struggle concerns resources, the outcome is unlikely to be a clear victory for one side or the other. Instead, the scope of openness or closedness will be extended in particular directions, not just in reach or extent but also in the 'depth' (the degree of non-restrictiveness) of openness.

Before assessing prospects for the balance between them, it is worth spelling out an important contradiction (which has not so far, perhaps understandably, received much attention): why should national funding bodies fund an e-infrastructure to secure *national* competitive advantage when an open infrastructure will also benefit the world-at-large?[60] Similarly, why should a university or the research group of a private company make its data or findings freely available rather than securing economic gain through IPR—or at least exploiting this resource to gain academic status for itself before releasing a (less valuable) resource? This tension only needs to be stated baldly to recognize the force that works against it: first, that the major gain in academia or innovative research (including tools) in these cases may come from attention rather than economic gain; and second, that building the capacity or the expertise to develop these findings or data or tools within a group is itself part of the gain, including (again) the attention received by this work which secures further talent and/or resources.[61]

The account given here so far has attempted to be neutral, pointing to the various social forces at work. There are also normative arguments[62] and pragmatic policy arguments[63] such as the effect on overall economic growth that can be made in favour of open e-infrastructures. In this respect it is important not to equate openness with ideas on the political left. Benkler's arguments, for example, are based primarily on liberal individualism, though some of his arguments are social democratic.[64] And the open source movement is to some extent informed by a libertarian 'hacker ethic'.[65] Further, openness, as we have seen, is not necessarily opposed to commercial exploitation or commercial gain. And finally, open e-infrastructures are being developed globally, regardless of political or economic systems.[66]

The normative ideas supporting openness, in other words, are varied. At the same time, the thrust of 'open science', including its origin in the Mertonian norms of 'communism' and 'universalism', resonates with ideas about a globally shared e-research infrastructure. Nevertheless, these normative or policy-driven ideas need to be put into sociological context; namely, how large technological systems and infrastructures develop and how they are shaped by the social environment of scientific research, including different social groups and their advocacy and the receptiveness of public opinion. The chances of success of 'openness' as an ideal and as policy and practice depends on advocacy among the research community and other social groups within a wider social environment, as well as on the technological momentum of the system itself and its extension in a more—or less—open direction.

As mentioned earlier, the various parts of openness, e-research and e-infrastructure cohere or connect only in overlapping agendas and intersecting organizations, as well as in the dynamic of a large technological system. This allows for a sociologically realistic assessment of its prospects in relation to the three arguments briefly stated in the introduction, and which we can now recognize as being interconnected: first, that despite their momentum, open e-research infrastructures cannot be assumed to be extensible without taking into account their limits, foremost how they compete on the research frontier for limited attention. Second, that research policy can attempt to steer infrastructures towards greater openness, but this will also require greater organizational coherence among the various actors and groups promoting it more broadly, and early support for open solutions in overcoming technical and social bottlenecks which will set the direction of the system on a particular track. Research policy will need to become aligned with as well as attempt to steer these wider forces. And finally, that an open technical infrastructure and openness as a research ideal are mutually constitutive; one without the other will narrow the social benefits and constrain knowledge production because of the way in which technical instruments drive scientific advance.

Apart from these general arguments at a 'global' level and in relation to an open e-research infrastructure as a whole, more detailed questions about the future of this system remain: which disciplines, types of e-science (i.e. access to instruments vs. access to data), national policies or other regulatory frameworks are most likely to extend open science practice? Where will the boundaries around open e-research be—between commercial and academic research, between the North and the South, or the humanities versus the sciences? Given its disparateness, how can the potential confluence of e-research and open science be strengthened in practice? And how, apart from greater organizational coherence, can the gaps between the rhetoric of open science, e-research and an open e-infrastructure and the reality of current practice be filled most effectively?

Notes and References

1. The author would like to thank Matthijs den Besten, Bill Dutton, Grace de la Flor and Wolf Richter for helpful comments on this paper. The research is funded by ESRC grant RES-149-25-1022 for the Oxford e-Social Science (OeSS) project: Ethical, Legal and Institutional Dynamics of Grid–enabled e-Sciences (http://www.oii.ox.ac.uk/microsites/oess/).
2. It is important to distinguish between access and use since use can entail contributing to the development of the shared resource, rather than just accessing it. A different way to make this distinction is to talk about the input to—and output from—research.
3. R. Schroeder and J. Fry, 'Social science approaches to e-science', *Journal of Computer-Mediated Communication*, 2007 (forthcoming).
4. D. Atkins *et al.*, *Revolutionizing Science and Engineering through Cyberinfrastructure*, Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure, Directorate for Computer and Information Science and Engineering, National Science Foundation, Arlington, VA, 2003.
5. *Ibid*, p. 4.
6. *Ibid*, p. 38.
7. *Ibid*, p. 53.
8. For example, *Ibid*, p. 57.
9. F. Berman and H. Brady, *Final Report: NSF SBE–CISE Workshop on Cyberinfrastructure and the Social Sciences*, 2005. Available at: www.sdsc.edu/sbe/, accessed 28 September 2006.
10. *Ibid*, p. 19.
11. American Council of Learned Societies, *The Draft Report of the American Council of Learned Societies' Commission on Cyberinfrastructure for Humanities and Social Sciences*, 2005, pp. 47–8. Available at: http://www.acls.org/cyberinfrastructure/acls-ci-public.pdf#search=% 22Cyberinfrastructure %20arts%20and%20humanities%22, accessed 28 September 2006.
12. M. Lennars *et al.*, *e-Infrastructures Road Map*, e-IRGSP, The Hague, 2005, p. 26. Available at: http://e-irg.eu/roadmap/, accessed 28 September 2006.
13. *Ibid*, p. 27.
14. *Ibid*, p. 52.
15. The openness of individual e-research projects is part of an ongoing research project by David, den Besten, Fry and Schroeder: see P. David, M. den Besten and R. Schroeder, 'How open is e-science?', *Proceedings of IEEE e-Science*, Amsterdam, December 2006.
16. P. David, 'Understanding the emergence of "open science" institutions: functionalist economics in historical context', *Industrial and Corporate Change*, 13, 4, 2004, pp. 571–89.
17. R. Merton, 'The normative structure of science', in *The Sociology of Science*, University of Chicago Press, Chicago, 1973 [1942], pp. 267–78.
18. *Ibid*, p. 273.
19. D. Hess, *Science Studies. An Advanced Introduction*, New York University Press, New York, 1997, p. 57.
20. See David *et al.*, *op. cit.*

21. S. Fuchs, 'What makes sciences scientific?', in J. Turner (ed.), *Handbook of Sociological Theory*, Kluwer Academic/Plenum Publishers, New York, 2002, pp. 21–35; T. Becher and P. Trowler, *Academic Tribes and Territories: Intellectual Inquiry and the Culture of Disciplines*, 2nd edition, Open University Press, Milton Keynes, 2001.
22. Merton's other two norms, disinterestedness and organized scepticism, are also relevant and all four are interrelated, but these two are perhaps less directly relevant to e-infrastructures.
23. S. Weber, *The Success of Open Source*, Harvard University Press, Cambridge, MA, 2004.
24. One lesson from open source for open e-research and e-infrastructure is that it is necessary to protect research and tools in order to release them for free access and use. For open source, this is done by means of licensing, either to prevent commercial use or to ensure that even commercial use stays 'open'. For academic research, one way to safeguard 'ownership' is by means of copyright, and developing various forms of licensing modelled on open source as a means of disseminating research (not only publications, but also data and tools). This is a key element in ensuring the 'open science' character of e-research; see J. Willinsky, 'The unacknowledged convergence of open source, open access, and open science', *First Monday*, 10, 8, 2005, available online; D. Burk, 'Intellectual property in the context of e-science', *Minnesota Legal Studies Research Paper No.06-47*, 2006. Available at http://ssrn.com/abstract=929479.
25. For the distinctions between different types of licenses see Weber, *op. cit.* and Wikipedia for an accessible overview, http://en.wikipedia.org/wiki/Open_source. This section will focus on more general issues and not only licensing.
26. See E. Deibel, 'Common genomes: open source in biotechnology and the return of common property', *Tailoring Biotechnologies*, 2, 2, 2006, pp. 49–84.
27. But Willinsky provides an inventory; see J. Willinsky, *The Access Principle: The Case for Open Access to Research and Scholarship*, MIT Press, Cambridge, MA, 2006.
28. Willinsky, 2005, *op. cit.*, p. 7.
29. *Ibid*, p. 10.
30. The umbrella term 'intellectual property' conflates several phenomena that are separate from a legal perspective: copyright, patents, and database protection (trademarks and trade secrets, two additional areas of IP, are less relevant in this context). There is no space to analyze the implications of these differences, especially as the distinction between them, especially in the digital realm, is currently in flux. But it can be mentioned briefly that the implications of these distinctions will be different if one takes a legal and economic perspective (to what extent, for example, does the expression as against the use of ideas, which is roughly the distinction between what is covered by copyright and patents, foster economic innovation); or if, as here, one takes a user perspective (to what extent are content and tools free of charge, accessible and usable, and do not lock the user in to a particular tool or resource or incompatible ones).
31. See for example Y. Benkler, *The Wealth of Networks: How Social Production Transforms Markets and Freedom*, Yale University Press, New Haven, 2006, pp. 380–2.
32. S. Sell, *Private Power, Public Law. The Globalization of Intellectual Property Rights*, Cambridge University Press, Cambridge, 2003.
33. P. David and B. Hall, 'Property and the pursuit of knowledge: IPR issues affecting scientific research', *Research Policy*, 35, 2006, pp. 767–71; see also Burk, *op. cit.*
34. R. Collins, 'Ethical controversies of science and society: a relation between two spheres of social conflict', in T. Brante, S. Fuller and W. Lynch (eds), *Controversial Science: From Content to Contention*, State University of New York Press, Albany, 1993, pp. 301–17.
35. P. Haas, 'Introduction: epistemic communities and international policy coordination', *International Organization*, 46, 1, 1992, pp. 1–35, quote on p. 3.
36. M. Keck and K. Sekkink, *Activists beyond Borders: Advocacy Networks in International Politics*, Cornell University Press, Ithaca, 1998.
37. A.–M. Slaughter, *A New World Order*, Princeton University Press, Princeton, 2004.
38. Collins, 1993, *op. cit.*
39. See M. Bauer and G. Gaskell, 'The biotechnology movement', in M. Bauer and G. Gaskell (eds), *Biotechnology: The Making of a Global Controversy*, Cambridge University Press, Cambridge, 2002, pp. 379–404. Bauer and Gaskell's functionalist model is in fact more

complicated as it deals with an area (biotechnology) where there has been considerable controversy and thus also includes the media and commercial interests. e-Research is more contained within the research community and among NGOs and advocacy groups.

40. For the role of the state in the globalization of science, see G. Drori, J. Meyer, F. Ramirez and E. Schofer, *Science in the Modern World Polity: Institutionalization and Globalization*, Stanford University Press, Stanford, 2003. The global nature of e-research is discussed by R. Schroeder, 'e-Sciences: infrastructures that reshape the global contours of knowledge', paper presented at the Second International Conference on e-Social Science, Manchester, 28–30 June 2006. Available at: http://www.ncess.ac.uk/events/conference/2006/papers/, accessed 28 September 2006. On the continuing strength of the state and its role as a target of social movements, see M. Mann, 'Has globalization ended the rise and rise of the nation-state?', in T. V. Paul and J. A. Hall (eds), *International Order and the Future of World Politics*, Cambridge University Press, Cambridge, 1999, pp. 237–51.
41. T. Hughes, 'The evolution of large technological systems', in W. Bijker, T. Hughes and T. Pinch (eds), *The Social Construction of Technological Systems*, MIT Press, Cambridge, MA, 1987, pp. 51–82.
42. T. Hughes, 'Technological momentum', in L. Marx and M. R. Smith (eds), *Does Technology Drive History? The Dilemma of Technological Determinism*, MIT Press, Cambridge, MA, 1994, pp. 101–13.
43. Van der Vleuten discusses the relation between them: E. Van der Vleuten, 'Infrastructures and societal change. A view from the large technical systems field', *Technology Analysis and Strategic Management*, 16, 3, 2004, pp. 395–414. In what follows, 'infrastructure' will sometimes be used in the singular to indicate the whole ensemble of e-research tools, and sometimes in the plural when referring to the different components of which it is made.
44. R. Collins, 'Why the social sciences won't become high-consensus, rapid-discovery science', *Sociological Forum*, 9, 2, 1994, pp. 155–77.
45. P. Galison and B. Hevly (eds), *Big Science: The Growth of Large-Scale Research*, Stanford University Press, Stanford, 1992.
46. I. Hacking, *Representing and Intervening*, Cambridge University Press, Cambridge, 1983, p. 146.
47. The point is overstated here: the key to biotechnological advances has been based on stand-alone computers for processing data from high-throughput machines for analyzing genetic material. Nevertheless, being able to access results via networks—scientific communication in the broad sense described earlier—has in recent years arguably become equally important. For the contribution of computing to biotechnology, see T. Lenoir and E. Gianella, 'The emergence and diffusion of DNA microarray technology', *Journal of Biomedical Discovery and Collaboration*, 1, 11, 2006, available online; and for how tools (in this case, PCR techniques) have been 'open' in the sense that patents did not hinder the dissemination and use of these tools, see J. Fore, I. Wiechers and R. Cook-Deegan, 'The effects of business practices, licensing, and intellectual property on development and dissemination of the polymerase chain reaction: case study', *Journal of Biomedical Discovery and Collaboration*, 1, 7, 2006, available online.
48. R. Whitley, *The Intellectual and Social Organization of the Sciences*, 2nd edition, Oxford University Press, Oxford, 2000.
49. Hughes, 1987, *op. cit.*
50. P. David and M. Spence, 'Towards a cyberinfrastructure for enhanced scientific collaboration', *OII Research Report No. 4*, Oxford Internet Institute, University of Oxford, Oxford, 2006. Available at: http://www.oii.ox.ac.uk/microsites/oess/papers.cfm, accessed 28 September 2006.
51. Whitley, *op. cit.*, pp. ix–xliv.
52. Collins, 1994, *op. cit.*; and R. Collins, *The Sociology of Philosophies: A Global Theory of Intellectual Change*, Harvard University Press, Cambridge, MA, 1998, pp. 532–8.
53. The literature on these phenomena is extensive, but van der Vleuten (*op. cit.*) provides a good overview in relation to large technological systems and infrastructures.
54. This user-base points to an important distinction between a bottom-up approach to infrastructures by means of sharing computing resources (as with the Seti@home research

project, http://setiathome.berkeley.edu/, whereby the spare computing capacity of Internet-connected PCs was harnessed for the Search for Extraterrestrial Intelligence) or user-generated content (for example, Wikipedia, see Benkler, *op. cit.*) in a decentralized way; as against top-down approaches which are characteristic, for example, of national communication technology infrastructures (or electricity grids). Note, however, that both require more or less centralized coordination and control, rather than being at one or other extreme.

55. Whitley, *op. cit.*
56. For the different social science perspectives on e-Science, see Schroeder and Fry, *op. cit.*
57. Willinsky, 2006, *op. cit.*
58. In this respect, perhaps the most critical controversy, which may mean the difference between an e-infrastructure that will come to be multi-tiered rather than being part of a single network, is the current challenge to 'network' or 'net neutrality' (Benkler, *op. cit.*, pp. 396–7): whether the Internet will continue to operate on the 'end-to-end' principle whereby content is delivered regardless of its origin and destination, or if different parts come to operate according to other principles. This remains to be seen.
59. Part of the reason that the tensions are not simply for or against open e-infrastructures is that they operate on different levels or in different realms; the realm of legal regulation, of the technical standard-setting possibilities for open systems, of research funding bodies and advocacy movements, for example, will push and pull the system on these different levels; although the point of large technological systems, of course, is that they will ultimately have to become aligned or gel into a stable whole.
60. This tension has been noticed for basic research: why should one country support basic research only to see the exploitation of this research take place in another?
61. This is akin to the 'shadow of the future' in open source software of potential commercial reward in the longer term if there are many users (see Weber, *op. cit.*).
62. Benkler, *op. cit.*
63. David, *op. cit.*
64. Benkler, *op. cit.*, pp. 301–10.
65. Weber, *op. cit.*
66. Though whether scientific e-research can be independent of political borders for countries like China, which is developing an e-research infrastructure and yet imposes restricted access on the Internet/Web, remains to be seen.