# How Open is e-Science?

Paul A. David
*Oxford Internet Institute*
*paul.david@oii.ox.ac.uk*

Matthijs den Besten
*Oxford e-Research Centre*
*matthijs.denbesten@oerc.ox.ac.uk*

Ralph Schroeder
*Oxford Internet Institute*
*ralph.schroeder@oii.ox.ac.uk*

This Version: July 6, 2006

## Abstract

*This paper examines various aspects of "openness" in research, and seeks to gauge the degree to which e-science is congruent with "open science." Norms and practices of openness, arguably, have been vital for the work of modern scientific communities, but concerns about the growth of stronger technical and institutional restraints on access to research tools, data and information recently have attracted increased notice – in part because of their implications for the effective utilization of advanced digital infrastructures and information technologies in research collaborations. Our discussion clarifies the conceptual distinctions between e-science and open science, and reports findings from a preliminary look at practices in U.K. e-science projects. Both serve parts serve to underscore the point that it is unwarranted to presume that the development of e-science necessarily promotes global open science collaboration. A programme of further empirical studies is outlined, aimed at establishing where, when, and to what extent "openness" and" e-ness" in scientific and engineering research may be expected to advance hand-in-hand.*

## 1. Introduction

Anyone enquiring about e-science is bound to come upon a quotation from John Taylor's (2001) introductory description of this movement's essence as being "about global collaboration in key areas of science and the next generation of infrastructure that will enable it." Although much that is said about e-science is occupied with the engineering and application of an enhanced infrastructure (Hey, 2005), this paper steps back to consider other requirements for attaining the ostensible goal of e-science programmes – augmenting the scale and effectiveness of global collaboration in scientific research.

Global collaboration takes many forms, but from the various initiatives around the world a consensus is emerging that collaboration should aim to be open, or at least that access to the underlying research and communication tools should be open. For example, the Atkins Committee, in a seminal NSF report that set the stage for research on "cyber-infrastructure" in the natural sciences and engineering in the US, advocated "open platforms" and referred to the grid as an "infrastructure for open scientific research" (Atkins, et al., 2003: 4, 38). In a follow-up report expanding that vision to include the social sciences, Berman and Brady (2005: 19) likewise stress the need for a "shared cyber-infrastructure." In the UK, the e-Science Core Programme has required that the middleware being developed by its projects be released under open source software licenses, and established an Open Middleware Infrastructure Institute (OMII). The e-Infrastructure Reflection Group (a high level European body formed in 2003 to monitor and advise on policy and administrative frameworks for easy and cost-effective shared use of Grid-computing, data storage, and networking resources) has gone further, issuing an "e-infrastructure roadmap" (Leenaars, 2005: 15-17, 22, 27) that calls for open standard grid protocol stacks, open source middleware, "transparent access to relevant [grid] data sources, and sharing of run-time software and interaction data including medical imagery, high-resolution video and haptic and tactile information. The e-IRG roadmap (p. 16) urges public funding for development of scientific software because "current Intellectual Property Right solutions are not in the interest of science."

Provision of enhanced technical means of accessing distributed research resources is neither a necessary nor a sufficient condition for achieving open scientific collaboration (David 2005). Collaboration technologies – both infrastructures and specific application tools and instruments – may be used to facilitate the work of distributed members of "closed clubs"—including government labs engaged in secret defense projects, and corporate R&D teams that work with proprietary data and materials, and guard their results as trade secrets. Nor do researchers' tools *as such* define the organisational character of collaboration. This is evident from the fact that many academic researchers who fully and frequently disclose their findings, and collaborate freely with colleagues on an informal, non-contractual basis, nonetheless, employ proprietary software and patented instruments, and publish in scientific journals that charge high subscription fees.

At the same time, it should be acknowledged that the availability of certain classes of tools, and the ease with which they may be used by researchers within and across scientific domains, is quite likely to affect organisational decisions and shape the ethos and actions of the work groups that adopt those tools. Some basic collaboration technologies -- notably e-network infrastructure such as grid services and middleware platforms -- particularly potent enablers of distributed multi-participant collaborations; they may significantly augment the data, information and computational resources that can be mobilized by more loosely organised, the "bottom-up" networks of researchers engaging in "open science."

In sections that follow further specify what we understand by the term "open science" and clarify the ways that this concept may be applied in examining structural features and working practices of a sample of emblematic e-science initiatives. We then present the results of a preliminary empirical enquiry, intended mainly to illustrate the implementation of our proposed conceptual framework. Although the selected sample of U.K. e-science projects is tiny, these findings support our contention that further research along the lines proposed here will be both feasible and illuminating.

## 2. Open Science

### 2.1. Ethos, norms and institutions

Many of the key formal institutions of modern science are quite familiar not only to specialists concerned with the economics and the sociology of science, technology and innovation, but equally to academic researchers of all disciplinary stripes. It is a striking phenomenon, well noted in the sociology of science, that there is high degree of mimetic professional organisation and behaviour across the diverse cognitive domains of academic endeavor. Whether in the social sciences, or the natural sciences, or the humanities, each discipline has its professional academies and learned societies, journal refereeing procedures, public and private foundation grant programmes, peer-panels for merit review of funding applications, organised competitions, prizes and public awards. The outward forms are strikingly similar, even if the details of the internal arrangements differ.

The norms of "the Republic of Science" that were so famously articulated by Merton (1973) are summarized compactly by the mnemonic device "CUDOS" (Ziman, 1994): communalism, universalism, disinterestedness, originality, and scepticism. These five key norms constitute a clearly delineated ethos to which members of the academic research community generally subscribe, even though their individual behaviours may not always conform to its strictures. It is important to appreciate both their separate and systemic effects as being conducive to the functional allocation of resources in an idealized research system. *Communalism* emphasizes the cooperative character of enquiry, stressing that the accumulation of scientific understanding emerges through the interactions among individual contributors; however much each may strive to contribute solely to the advancement of science, production of "reliable knowledge" cannot proceed far in isolation and so remains a fundamentally collective pursuit. Therefore, research agendas as well as findings ought to be under the control of personally (or corporately) *disinterested* agents: the specific nature and import of the new knowledge that is being sought should not be of such significant personal interest to the researchers involved that it risks skewing their methods or their reporting of "findings," and thereby rendering the results useless, or, worse, potentially detrimental to the research work of others. The force of the norm of *universalism*, in turn, is that it keeps entry into scientific work and discourse open for all persons of "competence," regardless of their personal and ascriptive attributes; equity aside, this preserves access to human talent and mitigates social pressures for conformity of opinion. Full disclosure of data and information about the methods by which new findings were obtained serves to maintain the effectiveness of new entrants to particular research domains, as well as to speed validation of the results produced. Ultimately, *originality* of intellectual contributions is the touchstone of acceptance of priority claims, and the source of collegiate reputations upon which material and non-pecuniary rewards are based. *Scepticism* is the appropriate attitude towards all priority claims, and, consequently, the claimant should be cooperative with the process (rather than taking offence) when scientific peers subject their work to trials of verification.

### 2.2. A Functionalist rationale for the norms of "Open Science"

It is thus possible to elaborate a functionalist explanation for the "open" part of the institutional complex of modern science, by focusing on its economic and social efficiency properties in the pursuit of knowledge, and the supportive role played by norms that tend to reinforce cooperative behaviours among scientists (Dasgupta and David 1987, 1994; David 1998, 2003). This rationale highlights the "incentive compatibility" of the key norm of disclosure within a collegiate reputation-based reward system grounded upon validated claims to priority in discovery or invention. In brief, rapid disclosures abet rapid validation of findings, reduces excess duplication of research effort, enlarge the domain of complementarities and yield beneficial "spill-overs" among research programmes. Without delving deeper into the details of this analysis, it may be noted that it is the difficulty of monitoring research effort that make it necessary for both the open science system and the intellectual property regime to tie researchers' rewards in one way or another to priority in the production of observable "research outputs" that can be submitted to "validity testing and valorization" – whether directly by peer assessment, or indirectly through their application in the markets for goods and services.

The specific functionality of the information-disclosure norms and social organisation of open science rests upon the greater efficacy of data and information-sharing as a basis for the cooperative, cumulative generation of eventually reliable additions to the stock of knowledge. Treating new findings as tantamount to being in the public domain fully exploits the "public goods" properties that permit data and information to be concurrently shared in use and re-used indefinitely, and thus promotes faster growth of the stock of knowledge. This contrasts with the information control and access restrictions that generally are required in order to appropriate private material benefits from the possession of (scientific and technological) knowledge. In the proprietary research regime, discoveries and inventions must either be held secret or be "protected" by gaining monopoly rights to their commercial exploitation. Otherwise, the unlimited entry of competing users could destroy the private profitability of investing in research and development.

One may then say, somewhat baldly, that the regime of proprietary technology (*qua* social organisation) is conducive to the maximization of private wealth stocks that reflect current and expected future flows of economic rents (extra-normal profits). While the prospective award of exclusive "exploitation rights" have this effect by strengthening incentives for private investments in R&D and innovative commercialization based on the new information, the restrictions that IP monopolies impose on the use of that knowledge perversely curtail the social benefits that it will yield. By contrast, because open science (*qua* social organization) calls for liberal dissemination of new information, it is more conducive to both the maximization of the rate of growth of society's stocks of reliable knowledge and to raising the marginal social rate of return from research expenditures. But it, too, is a flawed institutional mechanism: rivalries for priority in the revelation of discoveries and inventions induce the withholding of information ("temporary suspension of cooperation") among close competitors in specific areas of ongoing research. Moreover, adherents to open science's disclosure norms cannot become economically self-sustaining: being obliged to quickly disclose what they learn and thereby to relinquish control over its economic exploitation, their research requires the support of charitable patrons or public funding agencies.

The two distinctive organisational regimes thus serve quite different purposes that are complementary and highly fruitful when they co-exist at the macro-institutional level. This functional juxtaposition suggests a logical explanation for their co-existence, and the perpetuation of institutional and cultural separations between the communities of researchers forming 'the Republic of Science' and those who are engaged in commercially-oriented R&D conducted under proprietary rules. Yet, these alternative resource allocation mechanisms are not entirely compatible within a common institutional setting; *a fortiori,* within same project organisation there will be an unstable competitive tension between the two and the tendency is for the more fragile, cooperative micro-level arrangements and incentives to be undermined.

### 2.3. Questions about the Degrees of "Openness" of the Organisation of Research

Questions concerning the "openness" of the research process therefore should address at least two main sets of issues pertaining to the conduct of "open science" projects. The first set concerns the terms on which individuals may enter and leave the project. Who is permitted to join the collaboration? Are all of the participating researchers able to gain full access to the project's databases and other key research resources? How easy or hard is it for members and new entrants to develop distinct agendas of enquiry within the context of the ongoing project, and how much control do they retain over their findings? What restrictions are placed (formally or informally) on the uses they may make of data, information and knowledge in their possession after they exit from the research collaboration?

The second set of questions concerns the norms and rules governing disclosure of data and information. How fully and quickly is information about research procedures and data released by the project? How completely is it documented and annotated--so as to be not only accessible but useable by those outside the immediate research group? On what terms and with what delays are external researchers able to access materials, data and project results? When intellectual property is taken (by the project's members or their respective host institutions), will its use by outsiders be licensed on an exclusive or a non-exclusive basis? Are the rights to use such "protected" resources made conditional on payment of patent fees, copyright royalties? Do material transfer agreements impose charges (for cell lines, reagents, specimens) that require external researchers to pay substantially more than the costs of making the actual transfers?

Ideally these and still other questions could be formulated as a simple checklist, such as the one devised by Stanford University (1996) to provide guidelines for faculty compliance with its "openness in research" policy. That particular checklist was designed to implement rules against secrecy in sponsored research, so its scope is much too limited for our purposes. The Appendix (below) sets out a fuller set of questions, which practical considerations may abridge before it is used in structured interviews with the leaders and members of e-science projects.

## 3. e-Science as Open Science

Researchers in public sector science and engineering organisations historically have been at the forefront of many basic technological advances underlying new paradigms of digital information creation and dissemination. Their pressing needs for more powerful information processing and communication tools have led to many of the key enabling technologies of the "Information Society," including its mainframe computers, packet-switched data networks, the TPC/IP protocols of the Internet and the World Wide Web, its proliferation of markup languages, the Semantic Web and many more recent advances that facilitate distributed conduct of collaborative research.

### 3.1. Collaboration infrastructures, tools and materials, and open access dissemination

For essentially the same reasons, scientific and engineering research communities throughout the world now are active in developing not only technical infrastructure tools like the grid and middleware platforms, but a new array of shareable digital data and information dissemination resources, including public-domain digital data archives and federated open data networks, open institutional repositories, "open access" electronic journals, and open-source software applications. (David, 2004; David and Uhlir, 2005, 2006; Uhlir and Schröder, 2006; Schroeder 2006).

Peer-to-peer sharing of computing facilities – from SETI@Home to the International Virtual Observatory, focus on cooperative arrangements for distributed support of long-term exploratory research projects. Collaborative community-based open source software engineering projects, like the "free encyclopedia" Wikipedia, exemplify voluntary collective efforts that produce and maintain new information artifacts of use to researchers. Repositories such as MIT's D-Space, and OpenCourseWare, and Southampton's "*e*Prints", the physics arXiv pre-print repository, and GenBank are emblematic products of academic pro-activity in providing "open access" to research data and information; they complement the major scientific database development work of public institutes,' such as the U.S. NCBI and the Europe's EBI in bioinformatics.

But the availability of this expanding array of facilities for scientific cooperation and coordination, does not substitute for public and private support for open science research; nor can it transform the incentive and "reward systems" affecting the behaviours of researchers and research managers. Moreover, policies that involve disclosure of research findings can be consonant with the pursuit of goals quite removed from those traditional open science communities. Business corporations may encourage publication of employees' R&D results in peer-reviewed "open access" journals and conference proceedings for various strategic reasons: greater freedom to publish their results may prove effective in recruiting talented young scientists and engineers; disclosing results also can be a way to pre-empt a "frontier" research area that might otherwise be contested by rival firms. Yet, because the corporate research lab's ultimate purpose is to improve the "bottom line" of a profit-seeking business, R&D managers must be discriminating when setting the research agendas that employees are allowed to pursue, and when deciding which results may be disclosed to whom, when, and with what degree of completeness. .

The goals of the organisation conducting the research, rather than its selection of supporting tools and managerial techniques, thus set the balance of the configuration of its project's policies and practices; that overall configuration is what must be assessed in order to distinguish "open science" conduct from other institutional arrangements that may be found in contemporary e-science programmes and projects.

### 3.2. e-Science Infrastructure

Our examination of the "openness" of e-science practices is focused on the undertakings launched in 2001 by the U.K.'s e-Science Core Programme for the development middleware platforms. Most of the university-based projects funded by this programme were exploratory tool-building activities, and would seem to be good candidates for "open science practice" [see David and Spence (2003), for descriptions of the pilot projects]. According to the programme's Director, Tony Hey, a basic policy decision was taken at the outset to make these projects' results available as open source software. Implementation of this policy was pursued subsequently by the founding of the Open Middleware Infrastructure Institute (OMII), the stated purpose of which is to "leverage the wider development community through open source software development" (http://www.omii.ac.uk). That would appear to meet basic disclosure norms of open science, because a peculiar property of the output of software engineering research is that the artifacts produced (i.e., the code) also reveal the method of their construction.

The OMII's description of its mission, however, points to a strategic purpose behind the open source release policy. To become "the source for reliable, interoperable and open-source Grid middleware, ensuring the continued success of Grid-enabled e-Science in the UK," the Institute is promoting adoption of an open web services standard adopted widely among UK-funded e-science projects (http://www.omii.ac.uk/about/index.jsp; Atkinson et al., 2005). Provision of an open source reference implementation thus is seen as a means of making the Institute's web services standard more attractive to a broader array of potential users, including those in the industrial sector. OMII's mission statement supports this view: "Key members of the OMII have specific experience in industrial software development. We will also leverage the wider development community through open source software development. The OMII intends to lead the evolution of Grid Middleware through an open consultative process with major software vendors and major UK, EU and US projects." [http://www.omii.ac.uk/about/index.jsp, emphasis added]. That is a rather different rationale than the one offered by Programme Director Hey's statement at the September 2004 e-Science All Hands Meeting (http://www.allhands.org.uk/proceedings/proceedings/introduction.pdf): "…Web Services still are 'work in progress' so we must adopt conservative strategies to safeguard our UK investments and ensure that we converge on the standards that eventually emerge…."

In some significant respects the development of OMII into OMII-UK, like the web services standard that the Institute now is promoting, appears to parallel the course of evolution of the GlobusToolkit grid service protocols that were developed by a joint public-private project of the Distributed Systems Laboratory (U.S. Argonne National Laboratory). The initial goal of the Globus project, similarly, was to enable scientific research organisations to share enhanced computing resources; it too released code under an open source software license [whose idiosyncratic features are discussed by David and Spence (2003:32-33, and Appendix 4)]. But, in December 2004, the leaders of the Globus project launched the Univa Corporation as a provider of commercial software, technical support and professional services for the construction of grids based on Globus open source software. A number of the major hardware and software systems companies presently are aligned with this venture in the Globus Alliance. (http://www.univa.com/pdf/Univa-Launch_Release.pdf).

Manifestly, the dissemination of software engineering products as open source code has been quite compatible with both projects' evolution into "e-science" infrastructure and grid/web services providers whose activities are diverging from the traditional open science conduct of science and engineering research. OMII-UK now describes itself in business terms as creating an integrated "e-Science value chain" by providing infrastructure, components and solutions for "the e-science end user" (De Roure, 2006). Being neither a multi-site research collaboration nor a public entity supporting exploratory ('blue-sky') software engineering, it focuses on (a) "forming partnerships with targeted user communities," (b) sourcing code provided other grid service and middleware developers, (c) coordinating the "quality-assured software engineering" carried out by OMII-UK partners and its "managed programme," (d) "tracking and engagement with the standards processes," and, (e) building a "sustainable business" by attracting "partnerships and new investors."

As an organisation intermediating between university researchers and business clients, the Institute maintains a repository that can "ingest" code contributions from external sources -- if these match the OMII criteria. In practice, such donations come from U.K. academic research groups, and especially from the coordinated software "production pipelines" operating at three partner institutions—Southampton, Manchester, and Edinburgh. Getting software into the OMII repository is one thing, however, and accessing the "quality-assured" middleware code is something else again. Given OMII-UK's "sustainable business" goal, it is perhaps not surprising that unauthorized outsiders are not allowed to download the evaluation version of OMII client/server code; only the older stable versions can be accessed from the website. But, the terms on which even those versions are available from the Institute's repository disappoint expectations of easy "open access." Non-client researchers, after registering and obtaining a login name and password, may proceed to download software packages, but they will not necessarily obtain the underlying source code. [An attempt to download version 1.0 of the certificate management tool yielded a tar-ball within which was a jar-file containing java byte-code; procedures for extracting the corresponding java source code from that file are far from straightforward.]

Lastly, it should be remarked that the web-service standardization efforts of IBM and Microsoft– the big contenders for that potentially important business market--have been moving toward OASIS, and therefore away from the W3C's open standards approach. Since the OMII-UK and the Globus Alliance appear to be aligning themselves and thereby reinforcing this shift, they may be creating a serious impediment to the future emergence of open web service standards. The source of this threat is OASIS's policy of allowing publication of standards that require payments of licensing fees to patent holders (see Wikipedia, on 'OASIS', 2006). Aside from the drawbacks of proprietary standards, this could well have the added effect of foreclosing a volunteer-based open source implementation of web service standards.

The U.K.'s OMII initiative thus appears to have "morphed" into something other than a conventional academic research programme to build an enhanced open science infrastructure. By transforming "research-level" code created by e-science projects into tested and well-documented "production-level" middleware and grid software solutions, it is likely that the Institute will contributed substantially to facilitate the work of future e-science researchers. Whether it will prove to have promoted global expansion of "open science" e-collaborations, as much as proprietary R&D and e-business, however, remains much less clear.

### 3.3. e-Science Research Projects

We turn now to examine current collaborations that have emerged in several key domains of research that the infrastructure is intended to enable: (1)e-DiaMoND, a Grid-enabled prototype system intended to support breast cancer screening, mammography training, and epidemiological research; (2) MiMeG, which currently aims to produce software for the collaborative analysis and annotation of video data on of human interactions; and (3) Comb*e*-chem, an e-science test-bed that integrates existing sources of chemical structure and properties data, and augments them within a grid-based information and knowledge environment. Although none of these quite different projects have developed income-generating activities that might conflict directly with their adherence to open science norms, it is striking that all three have confronted other difficult issues related to "control rights" over data and information.

For e-DiaMoND the problem of control of mammography images remained unresolved when this "proof of concept" project reached it scheduled end. The researchers' original intentions to distribute standardised images for research and diagnostic purposes over electronic networks, clashed with the clinicians' concerns about their professional responsibilities to patients, protecting patient privacy, and assuring ethical uses of the data. Convincing clinical practitioners to trust the researchers, and engineering a comprehensive, adequately flexible security system proved to be less straightforward than had been expected (Jirotka et al., 2005). Even "to develop a clear legal framework that fairly accounts for the needs of patients, clinicians, researchers and those in commerce"— one with which the projects diverse partners would be able to work – has been surprisingly difficult (Hinds et al., 2005).

MiMeG, an ESRC funded e-social science project, encountered similar problems: the researchers who employed the tool for collaborative analysis of video-streams felt that the trust of the persons whose images they were studying would be violated by archiving the collaboration's data and making it available for re-use by other researchers, possibly for purposes other than the one for which consent originally had been obtained. It remains to be seen whether or not the ethical *desiderata* of privacy and informed consent of experimental subjects can be satisfied in future projects of this kind that plan sharing research data via the gird...

For the present, however, MiMeG has abandoned the project's initial intention to analyze video collaboratively via e-networks, and is focusing on the development of video analysis tools that other researchers can use. In that connection it is significant that the research ssoftware created by MiMeG is being released under the GNU GPL license (and hence distributed at minimal cost for non-commercial use). This policy resulted, at least in part, from the use of some GPL components (such as the MySQL relational database) in building the project's software tools. In addition, however, MiMeG's is encouraging external users to participate in further developing it recently released video analysis software tools. In these respects, the project has been able to go forward in the collaborative "open science" mode.

The Comb*e*-chem project at Southampton University is funded under the EPRSC's e-science programme and includes several departments and related projects. Only a few organisational features of this complex collaboration can be considered here, but several important aspects of its activities clearly are "open". One utilises the pre-existing EPSRC National Crystallographic Service, which has allowed remote "users" from UK universities to submit samples of chemical compounds to the laboratory at Southampton for x-ray analysis. Comb*e*-chem accepts submitted samples and returns them via a Globus-based grid and web services infrastructure (see Coles et al. 2005: appendix B). A present this service has some 150 subscribers who submit more than 1000 samples per annum (Frey 2004: 1031).

In addition to demonstrating and developing this grid implementation, a major project goal is to increase the archiving of analysed samples, thereby averting the loss of un-archived information and the consequently wasteful repetition of crystallographic analyses. Formerly, chemical analysis results yielded by these techniques were "archived" by virtue of their publication in research journals, most of which were available on a "subscription only" basis. Now it is possible to make results available in open access repositories via the open archive initiative (OAI), and deposited in *e*-BankUK archives and *e*Prints publications (Coles et al., 2005). Because they are put into RDF (Resource Description Framework) and other standard metadata formats, the archived results are searchable via the Semantic Web. With only 20 per cent of data generated in crystallographic work currently reaching the public domain (Allen 2004), and not all of it beings readily searchable, this service extension is an important open science advance.

Combe-chem's interrelated e-science activities thus illustrate four facets of open science practice: (a) using the Globus and web services open source grid software, (b) providing web access to shared resources for a diverse research community, (c) open access archiving and dissemination of results through an open repository, and (d) formatting of information using open standards. Like other publicly funded academic research, the project interacts easily with the world of commercial scientific publishing: fee charging journals that adhere to “subscriber only access” policies provide readers with links to the Comb*e*-chem data archive Moreover, as is the case in other collaborative projects that fit the traditional open science model quite closely, Comb*e*-chem has been able nonetheless to draw some sponsorship support from industry -- IBM having been interested in this deployment of a grid service [Interview with J. Frey, P.I.,Combe-chem: 29.11.2005].

## 4. Conclusion

We have described both the rationale and key identifying characteristics of open science, and have begun to explore ways to map the regions of practice where e-science and open science coincide. Although there are many e-science tools that could support distributed open science projects, this does not necessarily mean that all collaborative research carried out in the name of e-science qualifies as open science. Even academic e-science projects that intend to be open fall short in one or more aspects, particularly in regard to the information disclosure norms. The analysis presented here represents only a few trial steps in what is envisaged as a longer term programme of empirical research of the evolving conduct of e-science (David and Spence, 2003, Schroeder 2006). The list of questions set out by the Appendix is meant to serve as primary guidance for that enquiry. Our planned next step will approach a larger sample of the principal investigators of U.K. e-science projects for their help in constructing an inventory of policies and practices. We expect that this to provide a sufficient basis for producing more systematic answers to the question posed by the title of this paper.

### Acknowledgements

Thanks to Anne Trefethen, Jenny Fry, Jeremy Frey and Mike Fraser for their contributions. This paper is supported 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. All the judgments, mistakes and mis-interpretations herein are ours alone.

### References

F. H. Allen. High-throughput crystallography: the challenge of publishing, storing and using the results. *Crystallography Reviews*, 10:3–15, 2004.

D. E. Atkins, K. K. Droegmaier, S. I. Felman, et al. Revolutionizing science and engineering through cyberinfrastructure. Technical report, *National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure*, Washington D.C.: NSF, 2003.

M. Atkinson, D. DeRoure, A. Dunlop, et al.. Web service grids: An evolutionary approach. *Concurrency and Computation: Practice and Experience*, 17(2):377–390, 2005.

F. Berman and H. Brady. NSF SBE-CISE workshop on cyber-infrastructure and the social sciences. Final report, San Diego Supercomputing Centre, 2005.

S. J. Coles, J. G. Frey, M. B. Hursthouse, et al., The end-to-end crystallographic experiment in an e-science environment: From conception to publication. In *Proceedings of the Fourth UK e-Science All Hands Meeting*, Nottingham, UK, 2005.

S. J. Coles, J. G. Frey, M. B. Hursthouse, et. al., ECSES - examining crystal structures using ’e-science’: a demonstrator employing web and grid services to enhance user participation in crystallographic experiments. *Journal of Applied Crystallography*, 38(819-826), 2005.

P. Dasgupta and P. A. David, Information disclosure and the economics of science and technology. In G. Feiwel, editor, *Arrow and the Ascent of Modern Economic Theory*, chapter 16, pages 519–542. New York University Press, New York, 1987.

P. Dasgupta and P. A. David. Toward a new economics of science. *Research Policy*, 23:487–521, 1994.

P. A. David, Communication Norms and the Collective Cognitive Performance of ‘Invisible Colleges’,” in *Creation and the Transfer of Knowledge: Institutions and Incentives*, G.Barba Navaretti et.al., eds, New York: Springer, 1998.

P. A. David. The economic logic of ‘open science’ and the balance between private property rights and the public domain in scientific data and information,” in *The Role of the Public Domain in Scientific and Technical Data and Information: An NRC Symposium*, J. Esanu and P. F. Uhlir, eds., Washington, D. C.: Academy Press, 2003.

P. A. David, "Towards a cyberinfrastructure for enhanced scientific collaboration," forthcoming in *Advancing Knowledge and the Knowledge Economy*, ed. B. Kahin, Cambridge: MIT Press, 2006. Available at http://www.oii.ox.ac.uk/resources/publications/RR4.pdf.

P. A. David and M. Spence. Towards institutional infrastructures for e-science. Research Report 2, Oxford Internet Institute, Oxford, 2003 [Available at: http://www.oii.ox.ac.uk/research/project.cfm?id=26.]

P. A. David and P. F. Uhlir, Creating the global information commons for e-Science: Workshop Rationale and Plan. UNESCO, Paris, September 1-2, 2005.[Available at:
http://www.codataweb.org/UNESCOmtg/workshopplan.html

P. A. David and P. F. Uhlir. Creating global information commons for science: An international initiative of the committee on data for science and technology (CODATA). Unpublished prospectus, 17 April 2006.

D. De Roure, The OMII Experience. Presentation to the OSSWatch Conference on *Sustainability and Open Source*, Said Business School, Oxford 10-12 April, 2006. [Available at:http://www.oss-watch.ac.uk/events/2006-04-10-12/presentations/davidderoure.pdf.]

J. G. Frey. Dark lab or smart lab: The challenges for the 21st century laboratory software. *Organic Research and Development*, 8(6):1024–1035, 2004.

T. Hey. e-Science and open access. In *Berlin 3 Open Access: Progress in Implementing the Berlin Declaration on*

*Open Access to Knowledge in the Sciences and Humanities*, UK, University of Southampton, 2005.

C. Hinds, M. Jirotka, M. Rahman, et al., Ownership of intellectual property rights in medical data in collaborative computing environments. In *First International Conference on e-Social Science*, 2005.

M. Jirotka, R. Procter, M. Hartswood, et al., Collaboration and trust in healthcare innovation: The eDiaMoND case study. *Computer Supported Cooperative Work*, 14(4):369–398, August 2005.

M. Leenaars, e-Infrastructures Roadmap. *e-Infrastructure Reflection Group Technical Report*, 2005 [available at http://www.e-irg.org/roadmap/eIRG-roadmap.pdf].

. R. Merton. The normative structure of science. In *The Sociology of Science*, pages 267–278. University of Chicago Press, Chicago, 1973 [1942].

Openness in research. In *Stanford University Research Policy Handbook,* ch. 2.6, Stanford, CA, 1996; Openness in research checklist at: http://www.stanford.edu/dept/DoR/C-Res/ITARlist.html.

R. Schroeder. e-Sciences: Infrastructures that reshape the Global Contours of Knowledge, in *Proceedings of the Second International Conference on e-Social Science*. National Centre for e-Social Science, Manchester 28-30 June, 2006.

J. Taylor, Presentation at e-Science Meeting by the Director of the Research Councils, Office of Science and Technology, U.K., 2001. Available at: http://www.e-science.clrc.ac.uk.

P. F. Uhlir and P. Schröder. Promoting access to public scientific data for global science. *Data Science Journal*, (forthcoming) 2006.

Wikipedia, Entry for "OASIS" (organisation). http://en.wikipedia.org/wiki/OASIS_%28organisation%29

J. M. Ziman. *Prometheus Bound*. Cambridge University Press, 1994.

## Appendix: A Structured Interview Protocol

| |
|---|
| **Participation: Entry** |
| Do the terms of the award of support or the agreements among institutions and/or researchers on the project (I) Restrict research participation based on country of origin or citizenship; (II) Restrict research participation differentially (faculty, students, others – UK, EU, others); (III) Require institutional partners to be UK-based; (IV) Prohibit the hiring of non-UK/EU research staff? |
| Does the project (I) Restrict research participation to (a) members of the original group of research collaborators; (b) members of organisations that received an award of funding for the project; (c) members of the universities and other public research organisations that are parties to the formal agreement governing the project (whether or not they are awarded funding); (II) Accept all research personnel designated by a non-academic partner organisation in which they will be employed during the period of their participation? |
| **Participation: Exit** |
| Are any of the participants in the project bound by a confidentiality agreement? (I) Are academic researcher (faculty or student) participants bound by confidentiality agreements? (II) Are participants free to use research project outputs/by-products in their own future research and publications? (III) Are they allowed to continue working on the same topic directly after leaving the project? If not (a) What is the duration of the delay; (b) What is the sanction? |
| **Participation: Researchers' Access to Internal Resources** |
| Does this project or agreement (I) Provide that any part of the sponsoring, granting, or establishing documents may not be disclosed; (II) Limit access to confidential data so centrally related to the research that a member of the research group who was not privy to the confidential data would be unable to participate fully in all of the intellectually significant portions of the project? |
| If accepting proprietary information as part of the project (I) Is the information clearly defined; (II) Can the information be appropriately protected by security measures; (III) Can proprietary information be removed from research results, so that results may be freely published; (IV) Can student dissertations based on this project be publicly examined, and published in their entirety? |
| Does everyone involved in this project or agreement have access to all the information used in this project or agreement (within own institution, within partner institutions - data, software, facilities)? |
| Who sets permissions for data access? (I) Project director; (II) Principal investigators within the project; (III) Officer(s) of each participating institution; (IV) Government funding agency or regulatory authority; (V) Industry sponsor(s) that contribute (a) funding or facilities, (b)data, (c) personnel? |
| **Derivatives/By-Products: Research Materials** |
| Does this project or agreement have a standard material transfer policy? If yes, does the policy stipulate that material be made available according to a standard material transfer agreement? Do material transfer fee exceed incremental costs? |
| **Derivatives/By-Products: Tools (*software/techniques/devices*)** |
| In what forms are new research tools disclosed? (I) in patent specifications – (a) after filing, (b) when patent has issued? (II) In copyright software - (a) distributed as open source code; (b) distributed only as compiled machine code?; (III) Described in research publications – (a) in general, abstract terms; (b) specifically documented? |
| **Derivatives/By-Products: Data** |
| Does this project maintain a closed platform on a server, where data and other digital resources (applications programmes, memoranda) can be posted? What kind of platform (e.g. wiki/ftp)? How is access controlled (e.g. encryption, passwords, ip-based)? Are passwords and encryption key information available to all project members? If not, who sets permissions and security levels? |
| Can all project members document and annotate the databases? |
| Are participating researchers required to maintain lab-books? Can these be removed from the lab/project facilities? Does the project director have access to individuals' notebooks? Are lab notebooks deposited in an open archive? |
| **Research Findings: Control of disclosure** |
| Does this project/ agreement (I) Grant the sponsor a right of prepublication review and deletion of data and/or findings? (II) Restrict dissemination in compliance with – (a) law, external executive authority to classify, (b) its own project guidelines, (c) researchers' discretion? Are these restrictions based on considerations of (a) privacy, (b) confidentiality (c) commercialisation exploitation? May individual members of this project freely post their pre-prints of papers and data on the web, or submit them to open repositories? |
| **Research Findings: Deletions** |
| Are all project working papers internally reviewed before publication? Do reviewers have authority to delete material? Is deletion usual? |
| **Research Findings: Discretionary delays of publication** |
| May publication of research reports be embargoed for fixed periods? (I) only while a patent application is being prepared and filed? (II) At the request of industry sponsors? What is the maximum number of days of delay permitted? |
| **Research Findings: Cost of access to published results** |
| Is intellectual property claimed on project pre-prints by host institution(s)? Are standard copyright releases automatically signed with commercial publishers (a) assigning all rights requested, (b) reserving some rights? Are pre-prints usually archived in a public repository? Are project researchers encouraged to publish in 'open access' journals? Does project policy encourage self-archiving of electronic copies of papers, on an institutional web-servers. Are the data used in published articles deposited concurrently in public (open) data repositories? When data is deposited post-publication, what is the maximum delay allowed? |
| **Disputes** |
| Have there been disputes within institutions or among institutions, or with industry? How were these resolved? |
| **Policies and Governance** |
| If governance policies were formulated, was this done at the start of the project, around the first milestone, or subsequently? Were important matters of governance (on the issues discussed above) not properly anticipated? |