SIEPR Discussion Paper No. 08-021

Collaborative Research in e-Science and Open Access to Information

By
Paul A. David
Stanford University

Matthijs den Besten
Oxford e-Research Centre

Ralph Schroeder
Oxford Internet Institute

January 2009

Stanford Institute for Economic Policy Research
Stanford University
Stanford, CA 94305
(650) 725-1874

A contribution to the INTERNATIONAL HANDBOOK OF eRESEARCH,
Jeremy Hunsinger, Matthew Allen and Lisbeth Klastrup, eds., Published by Springer Verlag, Berlin 2010

Collaborative Research in e-Science and Open Access to Information

Paul A. David
Stanford University & UNU-MERIT (Maastricht, Netherlands)
& All Souls College, Oxford
pad@stanford.edu

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: 22 January, 2009

Abstract

This contribution examines various aspects of "openness" in research, and seeks to gauge the degree to which contemporary" e-science" practices are congruent with "open science." Norms and practices of openness are held to 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 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 differences between e-science and open science, and reports findings from a preliminary look at practices in U.K. e-science projects. Both parts serve to underscore the point that it is unwarranted to presume that the development of e-science necessarily promotes global open science collaboration. As there is evident need for further empirical research to establish where, when, and to what extent "openness" and "e-ness" in scientific and engineering research may be expected to advance hand-in-hand, we outline a framework within which such a program of studies might be undertaken.

1. Introduction

Anyone enquiring about "e-science" is bound to be led to 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 has been written about e-science is occupied with the engineering and application of an enhanced technological infrastructure for the transmission, processing and storing of digital data and information (Hey, 2005), this paper steps back to consider other, non-technological requirements for attaining the ostensible goal of e-science programs – augmenting the scale and effectiveness of global collaboration in scientific research.

Global scientific 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 there should be a substantial measure of "open access" to the data and information underlying published research, and to communication tools. 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: pp. 4,38). In a follow-up report expanding that vision to include the social sciences, Berman and Brady (2005:pp.19) likewise stress the need for a "shared cyber-infrastructure." In the UK, the e-Science Core Program 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:pp.15-17, 22, 27). The e-IRG roadmap 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; and for public funding of scientific software development, because "current Intellectual Property Right solutions are not in the interest of science" (p. 16).

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, David and Spence 2008). 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, guarding their findings as trade secrets until they obtain the legal protections granted by intellectual property rights. Nor do researchers' tools *as such* define the organizational 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 commercial 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 organizational 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 -- are 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." The availability of access to those resources on "share-and-share alike" terms can induce researchers' participation in passive as well as active collaboration arrangements, acquainting them with benefits of cooperation and thereby reinforcing the ethos of open science.

The sections that follow present our understanding of the term "open science," its significance for epistemologists, sociologists and economists studying the relationships between institutional structures, working procedures and the formation of scientific knowledge, and discuss ways that this concept may be applied to assess the "open-ness" of certain structural features and organizational practices observable in programmatic e-science initiatives and particular projects. We then consider some results from preliminary empirical enquiries, intended primarily to illustrate the empirical implementation of our proposed conceptual framework. Although only a limited sample of U.K. e-science projects (to date) have been selected for study from this perspective, the recent findings based on structured interviews and responses to a targeted email survey of research project directors display noteworthy consistencies and support our contention that further investigation along the conceptual and methodological lines explored will prove to be both feasible and illuminating.

2. Open Science

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 organization and behavior across the diverse cognitive domains of academic endeavor. Whether in the mathematical and natural sciences, or the social sciences or the humanities, each discipline has its professional academies and learned societies, journal refereeing procedures, public and private foundation grant programs, peer-panels for merit review of funding applications, organized competitions, prizes and public awards. The outward forms are strikingly similar, even if the details of the internal arrangements may differ.

2.1. Ethos, norms and institutions

The norms of "the Republic of Science" that were so famously articulated by Merton (1942, 1973) are summarized compactly by the mnemonic device "CUDOS": communalism, universalism, disinterestedness, originality, and skepticism.[1] These five key norms constitute a clearly delineated ethos to which members of the academic research community generally subscribe, even though their individual behaviors 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, and they can be set out briefly, as in the following paragraphs.

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.

Full disclosure of data and information about the methods by which new findings were obtained is another aspect of communal cooperation, vital both to maintain the effectiveness of new entrants to particular research domains, and to speed validation of the results produced. The force of the norm of *universalism*, in turn, is required in order to keep 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. 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. Since s*kepticism* is the appropriate attitude towards all priority claims, those who seeking recognition and peer esteem for their contribution should take no offence when scientific peers subject their work close scrutiny, and, instead cooperate with the of process of establishing the validity of their research conclusions and the merits of their assertion of priority.

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 behaviors 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 programs. 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 organization 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

---

[1] The nmenonic *Cudos* was introduced by Merton's 1942 essay on the normative structure of science, but the association of the "O" with originality was a subsequent modification that has become conventional. See Ziman (994)the

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 organization) 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 organizational 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 organization 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 "Open science" norms, social communications, and collective cognitive performance

The relationship between the norms that Merton (1942) perceived as underlying the institutionalized organization and stratified social structure of scientific communities, and the way the scientific research process proceeds when viewed from the epistemological perspective, forms a question with which sociologists and philosophers of science have wrestled, and about which there has been not a little internal disciplinary struggle.[2] With a few notable exceptions, philosophers and epistemologists of science shared a primary interest in *normative* questions concerned with the nature of the beliefs that people *ought* to hold about their surroundings.[3] While the classical approach in such endeavors sought to settle questions on an *aprioristic* basis, through solely conceptual arguments, a significant departure from that tradition took shape in the movement to "naturalize" epistemology and the philosophy of science. Following the work of the philosopher W.V.O. Quine (1953), this development sought to inject into such accounts some facts about the actual epistemic situations with which human must contend. Quine (1969) coined the term "naturalized epistemology," arguing that the positivist distinction between contexts of "discovery" and "justification" could be "naturalistically" translated into questions of perceptual psychology and sociology of knowledge, respectively. The former involves studying people's behavioral dispositions to associate words with

---

[2] See, e.g., Popper (1959, 1963), Quine (1962), Kuhn (1962), Latakos (1970).

[3] For a sketch of the classical philosophy of science as a domain of concerns completely disjoint from the sociology and history of science, which is deliberately overdrawn for effect, see David (1998b). A more nuanced, but nonetheless compact account is provided by Fuller (1994).

situations, whereas the latter is studied by examining the communication patterns by which such dispositions become stabilized for a community.

In a broad sense it may be said that this latter branch of the "naturalized epistemology" program has been carried forward by the research movement associated with the "new economics of science." This emerged in the 1990's among social scientists informed by the analytical perspectives of modern micro-economic theory, which they undertook to apply to the study of patterns of resource allocation that would arise from the social communication behaviors among communities of researchers whose conduct conformed in varying degrees with the institutionalized norms and ethos of "open science."[4] To understand how this latter point of departure was arrived at, one must refer back to the second of the two domains into which the study of scientific activities was formerly divided. That comprised the territory in which sociologists of science following the leadership of Robert K. Merton sought to understand the social context of modern science, its characteristic institutionalized modes of organization, and the associated behavioral norms and cultural ethos which guided the participants' transactions and permitted them to succeed collectively, if not always individually, in their pursuit of knowledge about the external world.[5] There were influential contributors to the history of science and technology who pointed to the role of external, material conditions affecting the physical phenomena that scientists undertook to study, and the ways in which they sometimes went about it. But, the canonical investigations in the sociology of science focused upon generic behavioral and organizational problems that were represented to be largely independent of the specific substantive concerns of the scientific discipline in question, and orthogonal to its cognitive contents and claims at particular historical junctures.

While adherents to these philosophical and sociological approaches to the study of science, therefore appear (from today's perspective) to have been collectively hesitant to assert any claims to the territory staked out by the other,[6] during the 1970's a new generation of sociologists took up the sociology of scientific knowledge -- "SSK", as it came to be styled. They insisted that the cognitive and the social dimensions of science should no longer remain compartmentalized as distinct, specialized fields of inquiry.[7] Instead, the subject matter had to be seen in reality to be inseparable, because in "discourse" -- within whose terms it was held possible to analyze everything -- social and cognitive contexts are thoroughly inter-penetrating and mutually interactive. For those within the SSK movement it appeared that there was little substance to the old sociology of science's preoccupation with macro-institutional structures, institutionalized reward systems, the social norms and the relationship of these to the organization of resource allocation within scientific communities. Encouraged to examine the practices of individual scientists and particular workgroups in academic research domains, they found that the pictures constructed by the Mertonian "normative structure" of science were highly idealized, and so could not readily account for many recurring micro-level realities that presented themselves to outside observers of "laboratory life" (Latour and Woolgar 1979); worse, the idealized normative account was held by critics to be an apologetic for "the scientific establishment" in that it accorded salience to institutionalize practices that were "functional" while glossing over dysfunctional aspects of the conduct of organized research. Criticism of its predecessors aside, the positive aspiration of this post-Mertonian generation of "sociologists of scientific knowledge" was to end the separation between epistemology and the sociology knowledge -

---

[4] Papers following Dasgupta and David (1987, 1994) in this genre include David, Mowery and Steinmueller (1992), Trajtenberg, Henderson and Jaffee (1992), Arora and Gambardella (1994, 1998); David (1994a, 1995, 1996); Gambardella (1994), David and Foray (1995), Arora, David and Gambardella (1997), Geuna (1999). Some of the foregoing are noticed in the wider survey of the economics of science by Stephan (1996).

[5] See, particularly, Merton (1942, 1968) among the writings assembled in Merton (1973,); see also, among the many works of Merton's students, Crane (1965, 1969, 1972), Cole and Cole (1967), Cole and Cole (1973), Cole (1978), Zuckerman and Merton (1971).

[6] In consequence, only occasionally, and rather tenuously, did their studies attempt to draw links between the ways in which communities of scientists were organized and rewarded, and the ways in which scientific statements were created and accepted within those communities. This is a gloss on a very extensive literature, for fuller discussion of which see Ben-David (1990), Callon (1995) and Leydesdorff (1995).

[7] See Barnes (1974, 1977), Bloor (1976), Mulkay (1979), Latour and Woolgar (1979), Knorr-Cetina (1981).

- by showing how societal forces translated into social processes that shaped the ways that scientists worked, the questions on which they chose to work, and the manner in which those questions were satisfactorily resolved (or left unresolved) by scientific communities.

Economists for the most part have been quite content to study the efficiency of resource allocation in the production of goods and services without stopping to inquire even superficially into the specific natures and concrete shapes of those commodities, much less entering into serious discussion of how they come to be differentiated in design and human perception.[8] Considered from that angle, it is perhaps not too surprising that contributors to the "new economics of science" (following Dasgupta and David, 1994) found it attractive to begin by reworking the terrain of organizational analysis already ploughed to good effect by sociologists in the Mertonian tradition. From that point of departure it was natural (being economists) to proceed to analyze how the structure of incentives created in a community whose members subscribed to the norms of open science would affect the efficiency of resource allocation in the production of generic information to which a significant measure of reliability could be attached; and then to comparisons of the workings of a academic, open science research regime with the alternative system commercially-oriented R&D grounded on the legal protections afforded to owners of trade secrets and intellectual property rights.[9]

This line of enquiry initially did not trouble itself too much about the nature of "scientific reliability," nor over the details of the ways in which that ("trustworthy") quality of information might be acquired, nor any of the other issues of socio-cognitive interaction that have occupied the sociology of scientific knowledge. But it was soon recognized that continuing to proceed in that fashion, which essentially ignored the foundation-level connections between the social organization of research and the nature of its information-products, was a mistake that was likely (sooner or later) to cause difficulties for an enterprise that aimed eventually to provide useful guidance for public policies affecting science.[10] Instead of continuing to focus exclusively upon the behavior of the individual researcher viewed in isolation, sociologists and economists have lately been giving greater attention to the systematic gathering and analysis of empirical evidence about role differentiation, specialization and cooperative patterns of resource allocation within and between teams, consortia, and institutionalized research networks.

As a consequence of the developments, contributions in the social sciences – from quantative sociology, from the bibliometrics of science (van Raan, 1988) and "scientometics Leydesdorff (1995,2001), as well as from the "new economics of science" – have become more closely aligned with the promising departure from older traditions in both epistemology and sociology that has marked the work of Kitcher (1993) on "social epistemology." The latter philosophical development explores the connections between the collective cognitive performance of scientific researchers, and the organization and conduct of social communications among them – by identifying the conditions under which a groups of individuals operating in conformity with various recognized procedures for modifying their individual research practices and informational transactions are able, through their interactions to generate "a progressive sequence of consensus practices."[11]

This is indeed a most welcome trend, because the accompanying movement away from empirical research conducted with the atomistic agent framework therefore may redirect the attention

---

[8] See Bacharach, Gambetta et al.(1994) for an exceptional departure from this tradition in economics.

[9] See, in addition to works cited above (fn. 4), e.g., Arora and Gambardella (1996), David (1996, 1998a, 2003, 2004), Turner and Mairesse (2002), Carayol (2003), Carayol and Dalle (2003), Carayol and Matt (2004, 2006).

[10] See, e.g., Brock and Durlauf (1999), David (1998b, 2002). Apart from the policy-relevant considerations noted in the following text, it may be argued that the cumulative, incremental development of propositions on which quasi-stable consensuses are formed among scientists do constitute grounds for insisting on the validity of drawing a cognitive distinction between the scientific and the other cultural pursuits. Because such resistance to "relativism" remains strongly contested in some quarters of contemporary "science studies," David develops an argument defending it on "evolutionary epistemological" grounds.

[11] See Kitcher (1993: Ch. 8) This chapter, on "The Organization of Cognitive Labor", opens by asking whether, in an "epistemically well-designed social system" it is possible that "the individual ways of responding to nature matter far less than the coordination, cooperation, and competition that go on among the individuals?"(p. 303)

of science policy toward considering the ways in which research communities' collective epistemological performance over that long-run could be improved by the provision of more adequate technical infrastructures and the reinforcement of norms and informal practices conductive to efficient resource-sharing, and collaborative interactions. If one asks whether social science research really can be informative on those matters, some signs of an affirmative answer are to be found in the recent quantitative work that is building upon the foundations of previous bibliometric studies. Such studies measure and explain patterns in the intensity of scientific co-publications not only involving authors affliated with different academic host institutions, but with different laboratories within the same institute or university.[12] Micro-level research on research in this vein have successfully identified factors that significant in promoting collaboration among scientists that lead to co-publications, and some that inhibit it.[13] Nevertheless, as promising as this is, the data sources on which such research have had to rely largely preclude one from saying anything very concrete about the effects upon productive collaboration of variations in incentive structures, institutional policies, and laboratory-level practices concerning dataflow controls, external releases of results, and the conditions for sharing access to materials and specialized facilities – institutional details that have remained sparsely observed and far from adequately documented.

2.4. Empirical questions about the degrees of "openness" of the organization of research

The major point underscored by the preceding review of the treatment of scientific communication and collaboration in the evolving literature on the philosophy and social science of scientific research is the emergence of a generally shared recognition of the reciprocal interdependence between the ethos and normative structures of research communities, on the one hand, and, on the other hand, the informal arrangements and institutionally reinforced conditions of access to research findings, underlying data and methodologies. Together, the norms and rules affecting communications through personal networks and broadcast channels, and the interchange of personnel among scientific workgroups, shape the possibilities of coordination and effective collaboration.[14] They thereby impinge upon the efficiency of scientific projects internal use of resources, and of resource allocation among the members of separate projects who constitute "invisible colleges" that are distributed across academic departments, institutes, universities, transcending national and regional boundaries – extending even into research laboratories of business corporations and government agencies.

The foregoing considerations have given us strong reason to regard the formal and informal *institutional* arrangements governing access to scientific and technical data and information, no less than to physical research facilities, instruments, materials, as critically influential among the factors determining how fully e-science will be able to realize the potentials for the advancement of reliable knowledge that are being created by advances in digital information technologies.

Questions concerning the actual extent of "openness" of research processes identified with contemporary e-science, therefore ought to address at least two main sets of issues pertaining to the conduct of "open science." The first set concerns the terms on which individuals may enter and leave research projects. 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

---

[12] See Blau (1973), Beaver and Rosen (1978, 1979), Katz (1994) for studies of scientific collaboration in the sociological and bibliometric tradition using co-publications. While economists have followed Jaffee, Trajtenberg and Henderson (1993) in used patent co-citations, and cross-citations between scientific publications and patents to try to gauge the extent of knowledge flows and inferred collaboration between academic and university scientists, Mairesse and Turner (2006) recently have broken new ground in measuring and econometrically explaining variations in the intensity of co-publication among the members of the large community of physicists in the condensed matter section of the CNRS.

[13] The statistically significant positive factors include both characteristics of the individual researchers (e.g. age, professional status, publication history), and their structural or institutional situations (e.g., department or laboratory size and disciplinary specialization, institutional reputation, ease of face-to-face interactions gauged by spatial proximity to other laboratories with specialization in the same or complementary research field).

[14] See Fry, Schroeder and den Besten (2008), and den Besten, David and Schroeder (2009) for further discussions of the distinction between generic research-technologies and narrowly defined research tools, and its bearing on the potential for openness in e-science.

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 the communication of 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 about research methods and results. 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? Are findings held back, rather than being disclosed in order to first obtain intellectual property rights on a scientific project's research results, and if so, then for how long is it usual for publication to be delayed (whether by the members or their respective host institutions)? Can research partners in university-business collaborations require that some findings or data not be made public? And when intellectual property rights to the use of research results have been obtained, will its use be licenses to outsiders on an exclusive or a non-exclusive basis? Do material transfer agreements among university-based projects impose charges (for cell lines, reagents, specimens) that require external researchers to pay substantially more than the costs of making the actual transfers? In the case of publicly funded research groups, the rights to use such legally "protected" information and data conditional on payment of patent fees, copyright royalties over which the members of the research group has any discretionary control, or is control exercised by external parties (in their host institution, or the funding sources)?

Ideally, these and still other questions might 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. The Stanford checklist, however, having initially been designed primarily to implement rules against secrecy in sponsored research, actually is too limited in its scope for our present purposes, and a fuller, more specific set of questions (inspired by this source) has been designed for data-gathering data in the context of contemporary U.K research projects. This empirical framework has been "field-tested" both in a small number of structured interviews, and a subsequent more extensive email-targeted survey of e-science project-leaders.[15] It is not intended to be comprehensive, and, instead, focuses on salient aspects of "openness and collaboration in academic science research" that could be illuminated by implementing systematic surveys of this kind on a much wider scale.

Of course, to pursue a substantially expanded program of inquiry into evolving e-science practices along these lines would necessitate some substantive modifications of the questionnaire in order appropriately "customize" the interview protocols and the survey template, which been designed for exploratory, "proof-of-concept" investigations. Conducting research of this kind across a widened international survey field certainly would require adjustments to allow for the greater diversity of institutional and organizational forms, research cultures, languages and technical nomenclatures. Furthermore, practical considerations might call also for abridging the questionnaires, so as to reduce the burden upon respondents and obtain a reasonably high response rates from an internationally administered survey – while avoiding costly individual email-targeting and follow-up requests for cooperation from potential respondents.

3. e-Science as Open Science

Researchers in public sector science and engineering organizations historically have been at the forefront of many basic technological advances underlying new paradigms of digital information

---

[15] For a report on the structured interviews, see Fry, Schroeder and den Besten (2008). David, den Besten and Schroeder (2006) presents a preliminary version of the framework of questions from which were developed both the structured interview protocol and subsequent on-line survey questionnaire, the result of which are resported by den Besten and David (2008). The complete set of survey questions may be consulted in den Besten, David and Schroeder (2009: Appendix Figures 1-11); see Fig. 4 (Q.8) for questions directly inspired by the Stanford checklist.

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.[16]

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 – which, like the "free encyclopedia" *Wikipedia,* make use of voluntary collective efforts to produce and maintain new information artifacts that, being produced by and for expert researchers, can constitute critical resources for those in specialized scientific fields. Repositories such as MIT's D-Space and OpenCourseWare, Southampton University'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 agency actions to support the open science conduct of research; nor can it transform the incentive and "reward systems" affecting the behaviors 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 organization 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 programs and projects.

3.2. e-Science infrastructure engineering

To examine the realities of e-science practices regarding "openness" it is both appropriate and feasible to focus on the undertakings initiated in 2001 by the U.K.'s e-Science Core Program for the development middleware platforms. Most of the university-based projects funded by this program were exploratory tool-building activities, and would seem to be good candidates for "open science practice" in addition to having been the earliest of the programs to be launched in the present century under the banner of "e-science."[17] According to the program's Director, Tony Hey, a basic policy

---

[16] See David, 2005; Dalle et al., 2005; David and Uhlir, 2005, 2006; Uhlir and Schröder, 2006; Schroeder 2006.

[17] See David and Spence (2003, 2008), for descriptions of the pilot projects, and discussion of their relationship to the U.S. "collaboratories" projects of the 1990's, and the more recent NSF "Cyberinfrastructure" Program. The discussion in this section draws upon David, den Besten and Schroeder (2006).

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."[18]

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 organizations to share enhanced computing resources; it too released code under an open source software license.[19] 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 program," (d) "tracking and engagement with the standards processes," and, (e) building a "sustainable business" by attracting "partnerships and new investors."

As an organization 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

---

[18] See http://www.omii.ac.uk/about/index.jsp, emphasis added. The foregoing, however, is a rather different rationale than the one offered previously by Program 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…."

[19] The idiosyncratic features of the Globus license are discussed by David and Spence (2003) :pp.32-33, and Appendix 4.

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."[20]

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 (2008) entry for 'OASIS'). 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 program 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 standardized 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

---

[20] 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. David, den Besten and Schroeder (2006) reported that an attempt to download version 1.0 of OMII's 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.

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 software 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) to build 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 program and includes several departments and related projects. Only a few organizational 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: 29.22.3005].

4. Structured interviews and on-line survey Findings

The questions about degrees of openness that we outlined in section 2.3 could be answered, at least in some part by the people involved in the research and development projects associated with the e-Science program in the UK. Fry, Schroeder and den Besten (2008) have carried out a series of structured interviews with a small group of the principle investigators of U.K. e-Science projects, designed to assess perceptions and practices relating to aspects of "openness" of the projects for which they had leadership responsibilities. A related questionnaire, suitable for implementation in an on-line Internet survey was developed on the basis of this experience and implemented in an email targeted survey of a larger population of U.K. e-science projects P.I.'s. The results obtained from the latter survey by den Besten and David (2008a) are broadly congruent with the detailed and more nuanced impressions drawn from the structured interviews.

4.1 Open science in e-science -- policy or contingency?: insights from in-depth interviews

Fry, Schroeder and den Besten (2008) report the findings from their use of a structured interview in conducting in-depth interviews about the relationships between collaboration in 'e-science' and 'open science,' with 12 individuals who had roles as principal investigators, project managers and developers engaged in UK e-Science projects during 2006.[21] The interview questions focused on research inputs, software development processes, access to resources, project documentation, dissemination of outputs and by-products, licensing issues, and institutional contracts. A focal interest of the approach in this study was the authors' juxtaposition of research project leaders' perceptions and views concerning research governance policies at the institutional level, with the responses describing local practices at the project level. As a detailed discussion of the responses (along with related documentary evidence drawn from the respective project's websites) is available elsewhere, its will be sufficient here to summarize briefly the main thrust of Fry, Schroeder and den Besten's (2008) findings.

Their interviews suggest that the desirability of maintaining conditions of "openness" in "doing (academic) science" is part of a generally shared research ethos among this sample of university-based project leaders. More specifically, the latter were not only cognizant of but receptive to the U.K. e-Science Pilot Program's strong policy stance favoring open source software tools and sharing of informational resources. Nevertheless, there were many uncertainties and yet-to-be resolved issues surrounding the practical implementation of both the informal norms and formal policies supporting open science practices. Making software tools and data available to external users might mean simply putting these research outputs on-line, but that need not be the same thing as making them sufficiently robust and well-documented to be widely utilized.[22] It seems that for those with leadership responsibilities at the project level, the most salient and fundamental challenges in resolving issues of openness in practice and operating policies, and thereby moving towards coherent institutional infrastructures for e-science research, involve the coordination and integration of goals across the diverse array of e-science efforts.[23]

By comparison, much less concern is voiced about the resolution of tensions between IPR (intellectual property rights) protections and the provision of timely common-use access to research tools, data and results. This is not really surprising, when the context is considered, even though these issues have been very much at the center of public discussions and debates about the effects of the growth of "academic patenting" on the "openness" of publicly funded research.[24] The programmatic

---

[21] Fry, Schroeder and den Besten (2008)'s structured interview protocol elaborated and modified the extended questionnaire proposed by David, Schroeder and den Besten (2006).

[22] As Fry, Schroeder and den Besten (2008) point out: "The effort to make the tools or data suitable or robust enough to make them into a commonly used resource may be considerable, and thus represents a Catch-22 situation for researchers: a large effort can be made, which may not be useful, but if it is not attempted, then it cannot be useful in the first place. Nevertheless, all projects expressed the aspiration to contribute to a common resource, even if this was sometimes expressed as a hope rather than a certainty or foregone conclusion."

[23] Coordination and integration problems calling for solutions that take the form of interoperability standards posed particularly difficult challenges for on-going projects in the UK e-Science Pilot Programme, according to the Fry, Schroeder and den Besten (2008): whereas some new software tools required compatibility with existing tools (for example, CQeSS needed to be interoperable with Stata) and this might be technically difficult to implement, achieving integration with other tools that are currently under development confronts more fundamental uncertainties about the requirements for compatibility or interoperability. The same applies to complying with standards, ontologies and metadata that are still in the process of development.

[24] See, e.g., David and Hall (2006); David (2007). Much of that discussion, however, has focused on the implications of the patenting of research tools, and *sui generis* legal protection of database rights (in the EU) in the areas of genomics, biogenetics and proteinomics, the patenting computer software (in the US) and computer implemented inventions (in the EU), and extensive patenting of nanotechnology research tools. While those have been very active fields of academic science research, and growing university-ownership of patents, they

focus U.K. e-science has not been on biomedical and biotechnology, life science fields in which patenting is especially important for subsequent commercial innovation. Furthermore, EU policy has circumscribed the patenting of software (without eliminating the patenting of embedded algorithms and a wider class of so-called "computer implemented inventions"), and in the U.K. itself, government agencies funding e-science projects have explicitly prohibited university grant and contract recipients from filing software patents that would vitiate the open source licensing of their outputs of middleware and applications software.

Most of the foregoing observations, although drawn from structured interviews conducted with a only a very small and non-random sample of project leaders, turn out to be quite informative -- in that these impressions are reinforced by the findings of a subsequent on-line survey that sought responses from the entire population of principal investigators on U.K. e-science projects.

4.2 Contract terms and "open-ness in research": survey findings on e-science projects

Systematic and detailed data at the individual project level about the openness of information and data resources remains quite limited, both as regards actual practices and the priority assigned to these issues among project leaders' concerns. A glimpse of what the larger landscape might be like in this regard, however, is provided by the responses to a recent on-line survey of issues in U.K. e-science that was conducted among the principal investigators that could be identified and contacted by email on the basis of National e-Science Centre (NeSC) data about the projects and their principal investigators. (See den Besten and David, 2008; den Besten, David and Schroeder, 2009).[25] Out of the 122 P.I.'s that were contacted, 30 responded with detailed information for an equal number of projects.[26] A comparison of the distribution of the projects for which responses were obtained with the distribution of the whole population of NeSC projects showed remarkable similarities along the several dimensions on which quantitative comparisons could be made -- including project grant size, number of consortium members and project start dates. This correspondence is reassuring, providing a measure of confidence in the representativeness of the picture that can be formed from this admittedly very restricted survey sample.

Formal agreements governing the conduct of publicly funded university research projects may, and sometime do involve explicit terms concerned with the locus and nature of control over data and publications, and the assignment of intellectual property rights based upon research results, especially when there are several collaborating institutions and the parties include business organizations. The survey sought to elicit information about project leaders' understandings of these matters and the importance they attached to such bearing as the terms of their respective project's agreement might have upon information access issues. It did so by posing various questions intended to probe the extent of participant's knowledge of the circumstances of the contractual agreement governing their

---

are not represented in the U.K.'s e-Science core program and so do not appear among the projects included in either the structured interview or the survey samples discussed here.

[25] This questionnaire instrument (which was posted on SurveyMonkey (www.surveymonkey.com) is reproduced in den Besten, David and Schroeder (2009): Appendix Figures 1-11. Particular questions among the 19 in the survey's substantive portion (many of which were comprised of several sub-items) are referred to in the following text and footnotes by their number in source, thus: [deBDS: App. Fig.1, Q.1].

[26] This number represented just over 10 percent of the projects listed by NeSC, implying a "project response rate" of 25 percent. The number of individual responses to this survey was larger, because P.I.'s receiving the email request were asked also to send it on to non-P.I. members of their project (which yielded an additional 21 responses that are not discussed here; also, in 3 cases more than one P.I. for a single project returned the questionnaire. The present analysis used only the one with the lowest frequency of "don't know" responses. The low apparent response rate from P.I.'s and projects may be due in some part to the relatively short time interval allowed for those who submitted survey replies to be eligible to receive a book-token gift. The existence of projects that appear more than once in the NeSC database and had multiple (co-) P.I.'s also would contribute to reducing the apparent rate of "project" responses.

project, namely, the identities of the parties responsible for its initial drafting and subsequent modifications (if any), as well as some of the contract's specific terms.[27]

The overall impression one draws from these survey responses is, once again, quite broadly congruent with the impressions that Fry, Schroeder and den Besten (2008) report on the basis of their 12 in-depth interviews. That applies also in regard to their vagueness as to the way that their project's governing agreement(s) had been arrived at. More than one third of the projects' P.I. and non-P.I. members either could not or would not say whether it was the lead scientists, or university administrative and service offices, or funding agency staff that had framed the initial project agreement; nor could they say who -- if anyone -- subsequently had sought contract modifications, whether before or after the funding contract(s) had been signed and the project was launched officially. The latter aspect of the results, predominantly reflect the reality that in many instances a university-based project's scientific activities already are underway well before of the completion of the initial template of a legal agreement, let alone the signing of a contract. Furthermore, the responsibility for producing an agreement that will fund and govern the collaboration, typically will be in the hands of actors that are not directly engaged in the project or involved in any way with its scientific work: staff in the host universities' research services offices (sometimes their legal counsel's offices), or officers of public funding agencies, or both. When multiple partners are involved, the role of the funding agency in the formal framing of the project – and hence in the framing much of its governing agreement, tends to be augmented *vis-à-vis* that of both the academic host institutions and sponsoring business companies.[28]

With a few notable exceptions, involving restrictions on the uses of proprietary data and publication of findings (where a collaboration had industrial partners), the terms of the agreement governing their project about the which respondent P.I.'s could were not such as would breach "openness in research" guidelines modeled on those of Stanford University (1996). Excluding the respondents who either found the question "not applicable" to their project or "did not know" the answer, between 96 and 98 percent of the replies reported that the terms of the agreement governing their project neither restricted research participation on the basis of country of origin or citizenship, nor required participation in EU-citizens-only research meetings, nor prohibited the involvement of research personnel from outside the EU.[29]

When asked whether their project agreement gave a sponsor the right of pre-publication review for purposes other than the preparation of a patent application, or the exclusion of proprietary data – i.e., the right to suppress findings that (presumably) were simply deemed "commercially sensitive" -- 92 percent among those replying definitively said "No." Although approximately one-quarter of all the respondents did not give a definitive reply because the found this was not applicable to their project (one may suppose there was no sponsor that would have such interest), 19 percent of those who accepted the question as relevant did not know whether to give a "yes", or a "no" answer. Almost as high a proportion (87 percent) among the definitive (yes or no) responses, reported that their project placed no restrictions on access to proprietary data that would have the effect of significantly blocking the work of a participating researcher. But, in the latter case, their was a considerable lower fraction of "don't know" responses (11 percent) among all those who accepted the question as pertinent for their respective projects.

---

[27] Unlike other survey findings discussed in sections 4.2 and 4.3, the results obtained from the 7 items in Survey Question 8 [see dBDS-Appendix Fig.4], are based on the complete tabulation of the answers from all 54 survey respondents – including both the 3 cases of reports on multiple projects by a single P.I., and the 21 non-P.I.'s.

[28] Funding bodies sometimes seek to form larger joint projects by bringing together academic that have submitted separate (competing) proposals, especially where there are opportunities to exploit differences the applicants' respective areas of special expertise. For further discussion of the formal legal context of collaborative e-science, see David and Spence (2008: sect. 2.3), and Fitzgerald (2008: Chs. 6, 11, 12).

[29] Among all the respondents who found these 3 questionnaire items (in Q 8) applicable to the circumstances of their respective projects, approximately 11 percent said they did not know the answer to the question. each of those questions.

The highest proportions of "don't know" responses were elicited by the questionnaire items concerning the existence of project contract terms and sponsorship agreements that were to be keep confidential, or provision that mandated project compliance with government regulations restricting the export of material or software (deemed sensitive for national "defense" purposes). The latter represented between 26 and 28 percent of those respondents who did not dismiss these specific issues as irrelevant to the circumstances of their project. Of course, it is to be expected that quite a few participants would not be uninformed about contract provisions that were supposed to be confidential, *a fortiori* when a substantial share of them were not project P.I.'s. Nonetheless, among those who thought they could give a definitive answer to the question declared that their project's agreement contained no such restrictive provisions.

The survey results just reviewed suggest that these e-science projects generally are free from positive, contractually imposed restrictions the participation of qualified researchers and significant restraints upon participants' access to critical data resources, and ability eventually to make public their research results. That a substantial fraction of project members appear not to be informed about the specifics of the project agreements under whose terms they are working is not very surprising, as many scientists express disinterest if not impatience with such matters, wishing to get on with their work without such distractions, and therefore leaving it to others -- including some among their fellow P.I.'s -- to deal with legal aspects of governance if and when problems of that nature intrude into the scientific conduct of the project. That more between 20 and 30 percent of participants remain uninformed about the details of contract terms that appear germane to the conduct of their research projects therefore could be taken as a healthy indication, namely, that issues involving restrictive provisions projects' contractual terms intrude upon the researchers' work only very infrequently, and so have remain little discussed among them.

Encouraging as that would be, the absence of formal, contractually imposed restraints on disclosure and access to scientific information and data resources leaves a substantial margin of uncertainty as to how closely the norms of "open science" are approximated by the operating practices and informal arrangements that are typically found within these projects. To probe into those important areas of "local" policy and practice, it is possible to examine the results obtained from a different set of the survey's questions.

4.3 Provision of information access in e-science projects: practices and policy concerns

The survey respondents were asked (see [dBDS: App. Fig.2, Q.6])) to classify their respective projects with regard to two taxonomic principles. Firstly, with which of the following functional scientific tasks was the project mainly engaged?: (1) generic tool development, (2) application development, (3) end-use application. Secondly, towards which among the main collaborative e-science forms was their project's work principally oriented to furthering? (i) grid access to distributed computing capacity, (ii) access to remote hardware instruments, (iii) access to specialized software, (iv) access to linked datasets or federated databases, (v) collaborative research with non-co-located teams. Although with these two axes and the resulting fifteen taxonomic combinations a more elaborate taxonomy may be constructed (den Besten and David, 2008b), for purposes of empirical analysis of the present small survey, the project classifications were collapsed into 3 broader purpose-engagement categories: (I) developing generic middleware tools for access to distributed computing resources and instruments (8 projects), (II) combining application development with database resources (11 projects), and (III) combining end-use for collaborative research (7 projects). A residual category absorbed (4) projects characterized by mixed purposes and activities that resisted simple summary description. In the following, we therefore focus on findings relating to the project-purpose clusters that can be concisely labeled as (I) *middleware-*, (II) *database-*, and (III) *end-user community*-oriented.[30]

From responses to survey question about measures actually undertaken to provide access to data and information relating to project results to researcher within the project and to outside researchers

---

[30] Considering only the "classifiable" group of projects, their percentage distribution among the broad "purpose-engagement" clusters is seen to be: 31 percent with *middleware* (I), 42 percent with *database applications* (II), and 27 percent with *end-user communities*.

(specifically from [dBDS: App. Fig.9, Q.10; Fig.5, Q.13, and Fig. 9, Q.13]) it is possible to form some sense of the relative importance of these goals among the projects. What emerges is that when projects are grouped by main purpose category (I, II, or III), the distributions of responses differs noticeably from group to group. One simple measure of relative importance is the ratio for the group between "yes" (Y)responses, signifying that specific access-enhancing facilities were being provided, and "no"(N) responses.[31]

The overall pattern in this (Y/N) response ratio displays systematic variation along two axes. Along the first axis, rather wider attention to providing external researcher with information access, compared with concerns about within-project access provision by means of working paper and publication repositories, databases, and regular data-stream access. Thus, the external-vs-internal access differences in the Y/N ratio holds within each of the main project-purpose categories: for the 3.0 vs 1.0 for projects in the *database* group, 0.91 vs. 0.44 for those in the *end-user community* group, and 0.35 vs. 0 in the *middleware* group. These figures also display the second axis along which there is systematic variation: attention to providing information access (both to outside and to inside researchers) is relatively more widespread among the database projects, less so among the end-user community projects, and least evident among the middleware development projects.

The existence of a separate institution created by the U.K. e-Science project that is dedicated to improving robustness and distributing open middleware, namely the OMMI (discussed previously), may well account for the latter feature of the pattern. That comparatively lower priorities appear to be attached to the internal provision of formal information access facilities among all 3 project- purpose categories, may well reflect the fact that only two-fifths of the survey responses pertain to projects that involved more than 2 consortium members, and another two-fifths of them had no other participating team. The management of inter-team information flows and data exchanges therefore may not be perceived among these projects as presenting major challenges.

Looking at the project start dates for the surveyed projects, one may group them into three cohorts whose relative sizes in the aggregate reflect the marked recent deceleration in the funding dynamics of the the U.K.'s e-science program as a whole: the pre-2003 cohort accounts for about 40 percent of the survey sample, the 2003-2004 cohort another 40 percent, leaving 20 percent in the post-2004 cohort. Within that temporal framework, something further may be said in regard to the specific information access repositories that have provided to members of these projects, the extent to which the latter are required to deposit materials therein, and also about the trends in the diffusion of these particular information management practices. The survey inquired about two main types of depositories: "common respositories"--for project-generated working papers and memos, for software code, and for data; "open access" repositories for project publications (department- or university-wide), for project preprints, project-generated software (distinguish version-controlled code, middleware, and applications software), and project-generated datasets. [32]

Common repositories for projects' research outputs in the form of working papers and software code appear to have been established quite wide from the early days of the U.K. e-Science program: at least, those for working papers and memos are reported by 77 percent of the projects among the pre-2003 cohort, and almost as large a proportion of the post-2004 cohort; whereas 69 percent of the projects forming the early cohort provide common repositories for software code, they are universally present among the most recent cohort. Comparison of the pre-2003 and post-2004 cohorts shows a rise also in the proportion of projects that are requiring the deposit of software code in these common repositories. In the case of data, however, common repositories are found only about half as frequently, and there is no evident secular movement on the part of projects that do provide them also to require that participants require their data.

---

[31] In compiling the results reported in the text, counts of instances where respondents said the particular question was not applicable, or that they did not know, have been omitted.

[32] See den Besten, David and Schroeder (2009), Table 2, for fuller presentation of the survey results reviewed here, than the survey questions in [dBDS: App. Fig.6, Q.10 and Fig. 7, Q.11].

It should be clear that access to the "common" repositories these e-science projects maintain may be restricted in many ways, and it therefore is of particular interest to turn to the available data about "open access" repositories. One finds in the case of data there are essentially no "open-access" repositories in the sense in which that term is understood currently. On the other hand, the spread of institutionally maintained (department or university-wide) repositories for "OA publications" is noticeably strong, although the proportion of cases in which participants are required to deposit material has not increased. The opposite pattern of change appears for pre-print repositories – their ubiquity has risen less markedly, but where these facilities have been set up, deposit requirements have become universal.

With regard to the various types of repositories for software, it seems clear that the proportion of open access repositories has approached the 30 percent share of middleware development projects in the total, and the relative frequency of adoption of version-control systems (with their archives) has more-or-less matched the relative share of middleware projects in the total – at least among the initial and most recent cohorts. Open access repositories for applications software have been established less frequently among the projects in the survey sample, but, where they do exist among the more recent product cohorts, the requirement mandating deposit of project-created computer code is widespread as is in the case among projects engaged in developing middleware.

What stands out most clearly from the findings reviewed in this section is that high level policy guidelines, set by the funding agency, can exert a potent influence on the pattern of adoption of open access archiving of scientific research products. In this instance there was an important early policy commitment by the U.K. e-Science core programme that middleware "deliverables" from its pilot projects would be made available as open source code, and this requirement for the research projects has been maintained (as has been noted above, in section 3.2) – even through there has been an evolution away from the original expectations of open source release of these output under GNU General Public Licences once they had passed through the OMII's enhancement and repacking process.

The extent to which the provision of access to data and information is perceived *at the project level* to be matters of explicit policy concern varies with the projects' roles in e-Research. This is only to be expected, particularly in view of the varied nature of these projects' "deliverables" and the existence of higher level policy regarding the software that is being created. A clear pattern of co-variation is evident in the responses to the question "Was the provision of access to data and information to members of the project a matter of particular concern and discussion in your project?"; and a parallel question referring to "external researchers" (see Fig. 11, Questions 16,17).[33] Among the projects engaged in *middleware development*, none expressed a concern for access within the project – presumably because the organization of the project and the ubiquity of open access code repositories meant that the matter one that largely had been settled. In contrast, however, the issue of external access was seen to be an important project concern by a third of the respondent P.I.'s from the projects developing *middleware*. That concern was expressed also by one-third respondents from projects involved with *user-communities* and *database resources,* especially the latter group.[34]

The responses concerning "obstacles encountered by the project in achieving "openness" (see Fig. 11, question 18) are consistent with the survey finding regarding actual practices and policy concerns at the project level, for they indicate that providing access to information to people *within* the project not found to be a problem deserving mention. All but two of the P.I.'s indicated at least one type of common repository to which participants were given access. Open access repositories are almost only provided where access for external research is seen as a concern within the project, which is the case for about one-third of the projects for which survey data is available. Project participants

---

[33] Over half of the projects having more diffuse purposes-- that is, purposes not preponderantly oriented toward either construction of middleware, research community usage, or applications and database resources -- failed to provide clear answers to questions 16 and 17. Responses from the "other purposes" group are not included in the analysis whose results are described in the text.

[34] Specifically, providing access to researchers outside the project was a significant concern for almost two-thirds of the *data*-centric projects and a third of *community*-centric projects.

are not always instructed to contribute to the repositories when the latter are provided, and it appears to be generally assumed that they will do so. On the other hand, none of the respondents indicated that their project was paying fees for the maintenance of an institutional or external repository to which their researchers would be given access.[35] Among the respondents who stated that the provision of access to outsiders was an important project goal, almost two-thirds listed one or more obstacle that had been encountered in achieving it; whereas among those who stated that such provision was not a project concern, almost half volunteered that they had encountered practical obstacles to external dissemination of their research outputs.[36]

.

5. Conclusion

We have described both the rationale and key identifying characteristics of collaborative "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 projects that conduct research in ways that accord more or less closely to open science norms, this does not assure that such is or will be the case where-ever collaborative research is pursued under the name of "e-science." Even academic e-science projects who leaders subscribe to the ethos of "open-ness in research" and institute some concrete "open access" practices, fall short of those norms in one or more respects, especially in regard to effective sharing of data resources and timely external disclosure of research findings. But, as has been shown, e-science projects are far from homogeneous, and in order to understand the variations in their information sharing policies and practices it is necessary to take into account the diversity of their scientific purposes, the technical nature of their tasks and the details of their organizational structures. The review presented here of the empirical evidence pertaining to U.K.-funded e-science projects, has been able to draw upon recent studies that carried out a small number of in-depth (highly insightful) interviews with selected P.I.'s, and obtained quantitative data from the responses to an on-line survey of e-science project leaders and other participants. These efforts in data collection and analysis represents only a trial steps in what is envisaged as a far broader and longer term program of systematic inquiries into the evolving global.

Acknowledgements

Thanks to Anne Trefethen, Jenny Fry, Jeremy Frey and Mike Fraser for their contributions. The preparation of this paper, and most of the research on the underlying data, has been supported by ESRC grant RES-149-25-1022 for the Oxford e-Social Science (OeSS) Project on Ethical, Legal and Institutional Dynamics of Grid-enabled e-Sciences. We gratefully acknowledge the institutional support provided by the Oxford Internet Institute and the Oxford e-Research Centre, and claim sole responsibility for the judgments expressed, as well as and mistakes and misinterpretations that may be found herein.

References

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

A. Arora, P.A. David, A. Gambardella, "Reputation and competence in publicly funded science: estimating the effects on research group productivity." *Annales d' Economie et de Statistique*, No. 49/50 (Numero Exceptionelle, Ed. J. Mairesse), 1998.

Arora, A., and A. Gambardella, The changing technology of technological change: general and abstract knowledge and the division of innovative labour," *Research Policy* 23, 1994:523-32.

A. Arora, A. Gambardella, Public policy towards science: picking stars or spreading the wealth?, *Revue d'economie industrielle,* 1997.

---

[35] Perhaps this question should have been phrased differently, e.g.: "Would the project be willing to pay repository changes, and for the inclusion of open access journals?".

[36] 11 respondents said external access listed external access among their project goals, 9 said it was not an important concern, and another 9 respondents left this particular question unanswered.

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.

M. Bacharach, D. Gambetta et al. (1994), "The economics of salience: A research proposal," Unpublished paper, Oxford Institute of Economics and Statistics.

B. Barnes, *Scientific Knowledge and Sociological Theory*, London: Routledge and Kegan Paul, 1974

B. Barnes, *Interests and the Growth of Knowledge*, London: Routledge and Kegan Paul, 1977.

J. Blau, Patterns of communication among theoretical high energy physicists, *Sociometry*, 37, 1973: pp.391-406.

D. Beaver, R. Rosen, Studies in scientific collaboration, Part 2: Scientific co-authorship, research productivity and visibility in the French scientific elite, *Scientometrics*, 1, 1979: pp.133-149.

J. S. Katz, Geographical proximity and scientific collaboration, *Scientometrics*, 31(1), 1994: pp.31-43.

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

D. Bloor, *Knowledge and Social Imagery* ,London: Routledge & Kegan Paul, 1976.

W. A. Brock. S. N. Durlauf, A formal model of theory choice in science, *Economic Theory*, 14(1), 1999: pp.13-130.

M. Callon, Four models for the dynamics of science.In *Handbook of science and technology studies,* Eds., S. Jasanoff, G.E. Markle, J.C. Petersen and T. Pinch (London: Sage Publications), 1995.

M. Callon, J.-P. Courtial, *Co-word analysis: a tool for the evaluation of public research policy* (Paris: Ecole Nationale Supérieure des Mines), 1989.

N. Carayol, The incentive properties of the Matthew Effect in academic competition, *Working Papers of BETA* 2003-11 ULP, Strasbourg, 2003.

N. Carayol, J -M. Dalle. Sequential problem choice and the reward system in Open Science, BETA *Working Papers of BETA* 2003-12, ULP Strasbourg, 2003. Revised and published in *Structural Change and Economic Dynamics*, 18(2), 2007:pp. 167-191].

N. Carayol, M. Matt, Does research organization influence academic production?: Laboratory level evidence from a large European university, *Research Policy*, 33(8), 2004: pp. 1081-1102.

N. Carayol, M. Matt, Individual and collective determinants of academic scientists' productivity, *Information Economics and Policy*, 18(1), 2006: pp. 55-72.

J. Cole, S. Cole, *Social Stratification in Science*, Chicago: Chicago University Press, 1973 .

S. Cole, Scientific reward systems: a aomparative analysis," in *Research in Sociology of Knowledge, Sciences and Art*, Ed.,R. A. Jones (Greenwich, Conn.: JAI Press), 1978: pp. 167-190.

S. Cole, J.Cole, Scientific output and recognition, *American Sociological Review*, 32, 1967:pp.377-390.

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(a).

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(b).

D. Crane, "Scientists at major and minor Universities: a study in productivity and recognition," *American Sociological Review*, 30, 1965: pp.699-714.

D. Crane, *Invisible colleges: diffusion of knowledge in scientific communities*, (Chicago: The University of Chicago Press) 1972.

J.-M. Dalle, P. A. David, R. A. Ghosh and W. E. Steinmueller, Advancing economic research on the free and open source software mode of production. In *How Open Will the Future Be? Social and Cultural Scenarios based on Open Standards and Open-Source Software*, .Eds. M. Wynants and J. Cornelis, (Brussels: VUB Press), 2005. [Preprint available at: http://siepr.stanford.edu/papers/pdf/04-03.html.]

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

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

P. David, Positive feedbacks and research productivity in science: reopening another black box. In *Economics and Technology*, O.Grandstrand, Ed., (Amsterdam: Elsevier), 1994.

P. A. David, Science reorganized? Post-modern visions of research and the curse of success, *Proceedings of the 2nd International Symposium on Research Funding*, Ottawa, 1996.

P. A. David, Common agency contracting and the emergence of "open science" institutions, *American Economic Review*, 88(2), May 1998a.

P. A. David, Communication norms and the nollective nognitive performance of 'invisible colleges'," in *Creation and the Transfer of Knowledge: Institutions and Incentives*, G.Barba Navaretti et.al., eds, New York: Springer, 1998b.

P. A. David, Cooperation, creativity and closure in scientific research networks: nodelling the simpler dynamics of epistemic communities, *Institutions et innovation: de la recherche aux systPmes sociaux d'innovation,* Ed., Jean-Phillipe Touffut, Paris: BibliothPque Albin Michel (Oconomie), 2002 [December 2002 revision available at: http:/siepr.stanford.edu/programs/SST_Seminars/David_All.pdf].

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, Can "open science" be protected from the evolving regime of intellectual property rights protections," *Journal of Theoretical and Institutional Economics*, 160, 2004: pp.1-26 [Available at: http://siepr.stanford.edu/papers/pdf/02-42.html.]

P. A. David, Toward a cyberinfrastructure for enhanced scientific collaboration: providing its "soft" foundations may be the hardest part. In *Advancing Knowledge and the Knowledge Economy*, ed. B. Kahin, Cambridge: MIT Press, 2006. [Pre-print available at: http://www.oii.ox.ac.uk/resources/publications/RR4.pdf.]

P. A. David, Innovation and europe's universities: second thoughts about embracing the Bayh-Dole regime. In *Perspectives on Innovation*, F. Malerba and S. Brusoni, Eds. Cambridge U.P. , 2007. [Preprint available at: http://siepr.stanford.edu/papers/pdf/04-27.html.]

P. A. David, M. den Besten and R. Schroeder, How 'open' is e-science? In *e-Science '06: Proceedings of the IEEE 2nd International Conference on eScience and Grid Computing*, Amsterdam, v. Iss December 2006: 33ff. [at: http://ieeexplore.ieee.org/iel5/4030972/4030973/04031006.pdf?isnumber=4030973â̂ =STD&arnumber=4031006&arnumber=4031006&arSt=33&ared=33&arAuthor=David%2C+P.A.%3B+den+Besten%2C+M.%3B+Schroeder%2C+R].

P. A. David, B. H. Hall, Property and the pursuit of knowledge: an introduction, *Research Policy*, 35 (June-July) [Special Issue, Guest-edited by P. A. David and B. H. Hall] 2006.

P.A. David, D.R. Mowery, W.E. Steinmueller, Analyzing the payoffs from basic research, *Economics of Innovation and New Technology*, Vol. 2 (4), 1992.

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

P. A. David and M. Spence. "Designing Institutional Infrastructures for e-Science," Ch. 5 in *Legal and Policy Franework for e-Research: Realizing the Potential*, Brian Fitzgerald, ed., Sydney, Australia: University of Sydney Press, 2008.

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). Unpublised 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.]

M. den Besten, P. A. David. Mapping e-science's path in the collaboration space. Paper presented to the conference on Oxford e-Research Conference 2008, 11-13 September 2008 [http://www.oii.ox.ac.uk/microsites/eresearch08/].

M. den Besten, P. A. David, R. Schroeder. E-science research and "open science": realities of information access policies and practices at the project level. In *World Wide Research*, W. H. Dutton and P. Jeffreys, Eds., Cambridge MA: MIT Press, forthcoming in 2009.

B. Fitzgerald, ed., *Legal and Policy Franework for e-Research:Realizing the Potential*, B. Fitzgerald, ed., Sydney, Australia: University of Sydney Press, 2008.

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.

J. Fry, R. Schroeder, and M. den Besten. Open science in e-science: Contingency or policy? *Journal of Documentation*, forthcoming in 2008.

S. Fuller, A Guide to the Philosophy and Sociology of Science for Social Psychology of Science, Ch. 17 in *The Social Psychology of Science,* W. Shadish and S. Fuller, Eds.(New York: The Guilford Press), 1994.

A.Gambardella, A. (1994), *Science and innovation,* Cambridge: Cambridge University Press, 1994.

A. Geuna, *The Economics of Knowledge Production: Funding and the Structure of University Research*, (Cheltenham,Eng.: Edward Elgar), 1999.

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.

J. Hirshleifer, The private and social value of information and the reward for inventive activity," *American Economic Review,* 61, 1971: pp:561-74.

A.B. Jaffe, M.Trajtenberg, R. Henderson, Geographic localization of knowledge spillovers as evidenced by patent citations," *Quarterly Journal of Economics*, 108(3), 1993: pp.557-598.

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.

P. Kitcher, The Advancement of Science: Science Without Legend, Objectivity without Illusions (Chicago: University of Chicago Press),1993.

Knorr-Cetina, K. (1981), *The Manufacture of Knowledge: An Essay on the Constructivist and Contextual Nature of Science,* Oxford: Pergamon Press.

T.S. Kuhn, *The Structure of Scientific* Revolutions, First/ Second Edition (Chicago: University of Chicago Press), 1962/1970

I.Lakatos, Falsification and the methodology of scientific research programs, in I. Lakatos and A. Musgrave, Eds., *Criticism and the growth of knowledge* (Cambridge: Cambridge University Press), 1970.

B. Latour, S. Woolgar, *Laboratory life* (Beverly Hills: Sage Publications), 1979.

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

L. Leydesdorff, *The challenge of scientometrics: the development, measurement, and self-organization of scientific communities*. Leiden: DSWO Press, 1995; 2nd Edition. Universal Publishers: uPUBLIS.com [available at: http://www.upublish.com/books/leydesdorff.htm].

J. Mairesse and L. Turner, Measurement and explanation of the intensity of co-publication in scientific research: an analysis at the laboratory level. Ch. 10 in *New frontiers in the economics of innovation and new technology*: *essays in honour of Paul A. David*. Eds., C. Antonelli et al. (Cheltenham, Eng.: Edward. Elgar. 2006.

. R. K. Merton. The normative structure of science [1942]. In R.K. Merton, *The sociology of science: theoretical and empirical investigations*, Edited by N. W. Storer (Chicago: University of Chicago Press), 1973: pp. 267–278.

R. K. Merton, The Matthew Effect in science, *Science*, 159(3810), January 5, 1968:pp. 56-63.

R.K. Merton, *The sociology of science: theoretical and empirical investigations*, Edited by N. W. Storer (Chicago: University of Chicago Press), 1973.

R. K. Merton, The Matthew Effect in science, II: cumulative advantage and the symbolism of intellectual property," *Isis*, 79, 1988: pp. 606-623.

D. R. Mowery, R. R. Nelson et al., *Ivory Tower and Industrial Innovation: University–Industry Technology Transfer Before and After the Bayh–Dole Act*. Stanford, C.A: Stanford University Press. 2004.

M. Mulkay, *Science and the Sociology of Knowledge*, (London: George Allen and Unwin), 1979.

K. R Popper, *Logic of scientific discovery* (London: Hutchinson), 1959.

K. R Popper, *Conjectures and refutations* (London: Routledge & Kegan Paul), 1963.

W.V.O Quine, Two Dogmas of Empiricism.In W.V.O. Quine, *From A Logical Point of View*, (Cambridge, Mass.: Harvard University Press), 1953.

W.V.O Quine, Carnap and logical truth. In W.V.O. Quine, *Logic and language: studies dedicated to Professor Rudolf Carnap on the occasion of his seventieth birthday* (Dordrecht: Reidel), 1962.

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.

S. Shapin, *A social history of truth* (Chicago: University of Chicago Press), 1994.

Stanford University, 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].

P. Stephan. "The Economics of Science," *Journal of Economic Literature*, 34(3),1996: pp.1199-1262.

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.

M. Trajtenberg, R. Henderson, A.B. Jaffe "Ivory tower versus corporate lab: an empirical study of basic research and appropriability," National Bureau of Economic Research Working Paper No. 4146 (August) 1992.

L. Turner and J. Mairesse, Explaining individual productivity differences in public research : how important are non-individual determinants ? An econometric analysis of French Physiciists (1986-1997). Working Paper-Cahiers de la MSE 2002-66. Université Paris-I, 2002.

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

A.F.J van Rann, ed., *Handbook of quantitative studies of science and technology*.(Amsterdam:Elsevier), 1988.

Wikipedia, Entry for “OASIS” Available at: http://en.wikipedia.org/wiki/OASIS_(organization). (Last accessed on 14 November) 2008.

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

H. Zuckerman, R. K. Merton, Institutionalization and Patterns of Evaluation in Science [1971], reprinted in R K. Merton*, The sociology of science: theoretical and empirical investigations*, N.W. Storer, ed., (Chicago: University of Chicago Press), 1973.