

**Towards a *cyberinfrastructure*  
for enhanced scientific collaboration:  
Providing its 'soft' foundations may be the hardest part**

BY

**Paul A. David**

*The Oxford Internet Institute and Stanford University*

First draft: 30 May 2004  
This version: 12 August 2004

---

Prepared for presentation at the International Conference  
*Advancing Knowledge and the Knowledge-Economy*  
held at the National Academy of Science, Washington D.C., 10-11 January 2005  
under the sponsorship of  
the European Commission, the U.S. National Science Foundation  
and the Organization for Economic Cooperation and Development

---

**ACKNOWLEDGEMENTS**

This paper was first presented for discussion at the Workshop on Networks of Knowledge: Research and Policy for the Knowledge-Based Economy, a meeting held in Brussels, on 7-8 June 2004, under the co-sponsorship of the European Commission—DG INFSO, the Organization for Economic Cooperation and Development, and the U.S. National Science Foundation. In making the present version, the author has benefited from the comments made by numerous Workshop participants, including Carliss Baldwin, Jean-Michel Dalle, Dominique Foray, Brian Kahin, John King, Bronwyn Hall, and Ilkka Tuomi. He gratefully acknowledges the Oxford Internet Institute's support of his research and writing on scientific research cooperation in digital technology environments, including previous collaborative work with Michael Spence, upon which the present paper has drawn.

*Contact Author:* P. A. David  
Oxford Internet Institute  
1 St. Giles'  
Oxford OX1 3JS  
[pdavid@herald.ox.ac.uk](mailto:pdavid@herald.ox.ac.uk) or [pad@stanford.edu](mailto:pad@stanford.edu)

## ABSTRACT

A new generation of information and communication infrastructures, including advanced Internet computing and Grid technologies, promises to enable more direct and shared access to more widely distributed computing resources than was previously possible. Scientific and technological collaboration, consequently, is more and more coming to be seen as critically dependent upon effective access to, and sharing of digital research data, and of the information tools that facilitate data being structured for efficient storage, search, retrieval, display and higher level analysis. A recent (February 2003) report to the U.S. NSF Directorate of Computer and Information System Engineering urged that funding be provided for a major enhancement of computer and network technologies, thereby creating a *cyberinfrastructure* whose facilities would support and transform the conduct of scientific and engineering research. The articulation of this programmatic vision reflects a widely shared expectation that solving the technical engineering problems associated with the advanced hardware and software systems of the cyberinfrastructure will yield revolutionary payoffs by empowering individual researchers and increasing the scale, scope and flexibility of collective research enterprises.

The argument of this paper, however, is that engineering breakthroughs alone will not be enough to achieve such an outcome. Success in realizing the cyberinfrastructure's potential, if it is achieved, will more likely be the resultant of a nexus of interrelated social, legal *and* technical transformations. The socio-institutional elements of a new infrastructure supporting collaboration – that is to say, its supposedly “softer” parts -- are every bit as complicated as the hardware and computer software, and, indeed, may prove much harder to devise and implement. The roots of this latter class of challenges facing “e-Science” will be seen to lie in the micro- and meso-level incentive structures created by the existing legal and administrative regimes. Although a number of these same conditions and circumstances appear to be equally significant obstacles to commercial provision of Grid services *in inter-organizational contexts*, the domain of publicly supported scientific collaboration is held to be the more hospitable environment in which to experiment with a variety of new approaches to solving these problems. Having surveyed the nature of the challenges, this paper concludes by proposing several “solution modalities,” including some that also could be made applicable for fields of information-intensive collaboration in business and finance that must regularly transcend organizational boundaries.

**Keywords:** e-Science, cyberinfrastructure, research collaborations, information technology, GRID computing, GRID services, legal institutions, intellectual property rights, inter-organizational contracts

## Towards a *cyberinfrastructure* for enhanced scientific collaboration

A new generation of information and communication infrastructures, including advanced Internet computing and Grid technologies, promises to enable more direct and shared access to more widely distributed computing resources than was previously possible. The vision of the transforming and empowering consequences of pervasively networked and interoperable computing resources for the conduct of scientific research has been a major force driving public sector support for hardware and software engineering efforts to create the necessary technological infrastructure with the collaboration need of scientific research communities foremost in mind. The emergence of this point of focus is entirely understandable, as it reflects both demand-side and supply-side conditions that together make public sector research communities the most immediately attractive environment in which to experiment with and deploy the next discrete augmentation of the computer-mediated telecommunications infrastructure based upon the Internet. Furthermore, it resonates with the historical roles play by academic research communities in fashioning the architecture of the ARPANET, and the pioneering Web browsers (Mosaic and the World Wide Web), from which the Internet has evolved. But the world has not stood still, and even within the domain of research performed in publicly supported, non-commercial organizations, achieving the conditions needed to facilitate effective collaboration among spatially and institutionally separated parties presents formidable challenges. Some of the most difficult among these are non-technological in nature, and should be as much a subject of research and policy initiatives.

### The argument in brief

The most cursory review of modern sciences' dependence upon distributed digital data and information resources and their growing needs for distributed, pervasive computing resources suffices to reveal why so many distinct research communities view the success of technical efforts to provide an advanced "digital infrastructure" as common priority item on their respective requirements-lists. To be sure, there are differences in the degrees of enthusiasm expressed about this goal, and a number of valid questions that can be raised as to whether or not "the Grid" is really of equally critical importance for the conduct of twenty-first century research in all the principal domain sciences, let alone mathematics or the social sciences. But, that is only one, and perhaps not the most important of the "reality-checks" that should be undertaken before committing extensive resources to the quest for Grid-enabled collaborative science as the lead-user of the global cyber-infrastructure.

There are sound grounds for expecting that breakthroughs on the engineering front alone will not be enough to achieve the societal gains in knowledge-creation that could be made feasible by further reducing the marginal social costs of access to reliable processing, reproduction and transmission of data and information. Success in realizing the transformative potentialities of the cyber-infrastructure is likely to be the resultant of a nexus of interrelated social, legal *and* technical changes. The burden of the argument in this paper is that the supposedly "softer parts," that is to say, the socio-institutional elements are necessary complements of the technical components in the new digital information infrastructure that would support collaborative activities of many kinds. Curiously, the institutional infrastructure requirements have tended to be overlooked, as though fulfilling them will be easily arranged; whereas they are every bit as complicated as the hardware and computer software, and indeed

may prove much harder to devise and implement. This is particularly likely to be the case in regard to collaborative activities that are inter-organizational – the very sphere in which the vision of Grid-support seems to hold the greatest transformative potentialities.

By comparison with the pace of engineering advances, far greater uncertainties continue to surround the extent to which individuals, groups, and organizations engaged in scientific and technical research are able arrive at informal and formal contractual arrangements and institutionalized procedures to reduce the transactions costs of collaboration. The roots of this state of affairs lie in the micro- and meso-level incentive structures formed by familiar features of the established legal and administrative regimes. Mundane as these obstacles may be, those transaction costs, and the economic rents protected by intellectual property rights that now occasion greater difficulties in negotiating agreements governing inter-organizational research collaboration, cause private costs to greatly exceed the marginal social costs of effective access to data, information and information tools.

Economic analysis tells us that efficient resource allocation can occur in a decentralized regime when the prices of the goods in question are set equal to their marginal social costs. This implies that under modern conditions the imposition of substantial costs of access to existing data and information-goods is tantamount to an inefficient tax, resulting in the wastage of society's resources. That burden is particularly difficult to justify on economic or ethical grounds where the initial, fixed costs of generating the information already has been borne by society through the provision of public funding for research and scholarship. Reducing the size of the transaction-cost “wedges,” and the rents are protected by intellectual property rights over scientific and technical data and information, is therefore a key challenge that must be met in order for global research communities, and society more generally to benefit from the novel “technologies of collaboration” that now are becoming engineering practicalities.

The same class of “soft” problems underlie the exacting technical challenges have emerged as serious obstacles to the commercial provision of Grid services *in inter-organizational contexts*. Although the private incentives for overcoming those problems in the commercial sphere may be stronger than those felt by policy-makers with responsibilities for public sector research, the latter domain – for all its complexities notwithstanding –remains the more hospitable of the two environments for experimentation with new approaches to solving these problems. This is the case both because the ethos of cooperation in the collective pursuit of knowledge and the informal norms of ‘open science’ still persist in many research communities, and because the public funding agencies still retain an important degree of policy-setting leverage over the relevant research organizations and institutions. Therefore, it is suggested here that serious efforts be made to explore some of the proposed modalities for the construction of an appropriate institutional infrastructure for collaborative e-Science. Not only may these yield direct benefits in terms of advancing the state of foundational scientific and engineering knowledge, but there can be significant spill-overs; experimentation with new institutional and organization arrangements may yield solutions that find application to other fields of collaborative production that are both information-intensive and regularly transcend the organizational boundaries.

The following sections of this paper provide some elaborations and substantiations for the key propositions in the foregoing argument, and conclude with a brief agenda for policy action and research.

## Science and cyberinfrastructure

Scientific research collaboration is more and more coming to be seen as critically dependent upon effective access to, and sharing of digital research data. Equally critical are the information tools that facilitate data being structured for efficient storage, search, retrieval, display and higher level analysis, and the codified and archived information resources that may readily located and reused in new combinations to generate further additions to the corpus of reliable scientific knowledge. The progress already made in these directions has enabled scientists to perform quantitatively and qualitatively new functions in the collection and creation of ever-increasing volumes and diverse forms of raw data pertaining to a wide array of natural objects and phenomena. It has compressed the space and time in which data and information can be made available for analysis and use in further research. It has opened up the practical possibilities of integrating and transforming scientific and technical data into virtually unlimited configurations of information, knowledge, and discovery.

These new capabilities have stimulated the emergence of entirely new forms of distributed research collaboration and information production. The idea that the potentialities of science and engineering research can in this way be greatly augmented has emerged as a driving force for publicly supported initiatives to create new, integrative technical elements of a global scientific infrastructure, such as the transport layers and networking protocols for the Grid, the e-Science “middle-ware” platforms and “virtual laboratories,” and, on the layer above, the Semantic Web. In the U.S., a report by a distinguished advisory panel to the NSF Directorate of Computer and Information System Engineering (in February 2003) envisages these enhanced computer and network technologies as forming a vital infrastructure – dubbed the *cyberinfrastructure* – the impact of which upon the conduct of scientific and engineering research would be akin to historical effects of super-highways, electric power grids, and other physical infrastructures in raising the economic welfare gains yielded by conventional physical production activities and commercial exchanges.<sup>1</sup>

### **The GRID, and the challenge of inter-organizational sharing of computing resources**

The vision animating much of the current interest in potential transformative effects of an enhanced cyberinfrastructure is the program to construct the GRID, a computer infrastructure that will not suffer from the technical deficiencies of the contemporary Internet -- unreliable connections, limited and unevenly distributed bandwidth, vulnerability of computers to intrusion and self-propagating malign programs, to name only a few among the more familiar. Akin to the electricity “grid” the computational GRID’s users would be able to plug in whatever information technology appliances they need, anywhere, and at any time; they will have at their instant disposal the GRID’s computing power, shared data, and shared instruments -- all without being forced to know, or worry about the underlying architecture that located and delivered these resources.<sup>2</sup> The vision thus projected of seamless access to ubiquitous or

---

<sup>1</sup> The potential to revolutionise science and engineering in the 21st century is set out at some length as the rationale for a major programmatic commitment by NSF. See D.E. Atkins, K.K. Koegemeier, S.I. Feldman, et al., *Revolutionizing science and engineering through cyberinfrastructure*: Report of the National Science Foundation blue-ribbon advisory panel on cyberinfrastructure. (February) 2003. [Available at: [http://www.communitytechnology.org/nsf\\_ci\\_report/](http://www.communitytechnology.org/nsf_ci_report/).] On of the transformative implication in the local, Oxford context, see also, P. Jeffries, “e-Science and the Grid: Why it will change Oxford”, Presentation by the Director of the Oxford University e-Science Centre to the Oxford BioInformatics Forum, 7 November, 2001. [Available at: <http://e-science.ox.ac.uk/>]

<sup>2</sup> General overviews of the Grid and related Internet computing are provided by I. Foster, (2000), “Internet Computing and the Emerging Grid”, *Nature*, 7 December 2000. [Available at:

“pervasive” computing resources, is somewhat utopian, to be sure. But, that is not an uncommon quality in conceptualizations of new technical systems; “technological presbyopia,” the condition of being able to envisage things more clearly the farther they are from present realization, seems to serve effectively as an coordinating mechanism for the mobilization of inventive efforts – even though the prospective users may grow weary and sceptical while waiting for the future to arrive.

#### **The GRID and Internet:**

The design goals for GRID engineering aim to provide *interoperable, ubiquitous, reliable and inexpensive* access to computational and computer-mediated resources.<sup>3</sup> Plainly, the GRID is not just another application that would run on the Internet. Rather, it is a sort of operating system for the Internet. It provides *middleware*, an abstraction from the peculiarities of the heterogeneous hardware which constitutes a network that allows applications to ignore these peculiarities and hence makes the development of such applications an easier task. But it is important to appreciate that a host of technical engineering issues have to be addressed before the GRID can take effect – management of distributed databases, communication between software across computing platforms, security systems that nonetheless permit (authorized) passage through protective firewalls, while preserving the privacy of those sharing networked resources, etc. Such are the formidable technical challenges with which the field of GRID computing is concerned, and to a realistic observer they suggest that full system will be long enough in emerging to allow an extended period to work on other non-technical requirements for its effective utilization.

#### **Web Services and GRID Services:**

Web services is a catch-all term for the current efforts of industry to come up with standards for interaction over networks. Web services can be thought of as the first evolutionary step from the Internet towards the GRID. A *service* is defined as a network-enabled entity that provides some capability. Entities are network-enabled when they are accessible from other computers than the one they are residing on. Capabilities to provide are computing, storage, programs, etc. The quintessential web service is Internet Banking. A *GRID service* is a web service that provides the interfaces and follows the protocols (interface conventions) spelled out by the Globus projects at the Argonne National Laboratory in Illinois, which aim to make it possible for software to discover which services are provided and/or compose them on the fly.

#### **Peer-to-Peer:**

The GRID and peer-to-peer are often lumped together, but they refer to different concepts. Peer-to-peer refers to the architecture of particular applications that are organized in a decentralized fashion (as opposed to the prevalent client-server model). Well-known examples of peer-to-peer applications are Napster, and in the sphere of distributed computing: Seti@home, and Climateprediction.<sup>4</sup> One may add to these the Internet itself, for it is a non-hierarchical,

---

<http://www.nature.com/nature/webmatters/grid/grid/html>]; I. Foster, “The Grid: Computing without Bounds,” *Scientific American*, April, 2003. For further detail, consult I. Foster, I. and C. Kesselman, eds., *The Grid: Blueprint for a New Computing Infrastructure*, San Francisco, CA: Morgan-Kaufmann, 2001; I. Foster, C. Kesselman, J. M. Nick, and S. Tuecke, “The Physiology of the Grid” (Version 2/17/2002) [available at: <http://www.globus.org/research/papers/ogsa.pdf>.]

<sup>3</sup> The corresponding terms favored in industry discussion of GRID engineering targets are “pervasive”, “consistent”, “dependable” and “inexpensive” computing. See, e.g., Keith Norman, “Grid Computing.” Tessella Scientific Software Solutions: Issue V1.R3.M0 (Abingdon, Oxon.:Tessella Support Services plc), February 2004. [<http://www.tessella.com>.]

<sup>4</sup> See <http://setiathome.ssl.berkeley.edu> ; <http://www.climateprediction.net> .

connectionless telecommunications system.

Although it is likely that the GRID will follow a peer-to-peer architecture, it appears to be feasible and perhaps is actually easier to implement the GRID as a central server that keeps track of what its clients are doing. This is a particularly attractive possibility where the application of the GRID architecture for distributed computer clusters would be “organizationally bounded,” that is to say, deployed “within the firewall” of a single organization. The firewall here carries both a technical and managerial control connotations: “inside” the firm means that many of the issues of restricted access to sensitive information, assignments of responsibilities, legal liabilities and divisions of gain will already have been addressed by other means. In the context of a large, transnational corporations with geographically dispersed facilities, the need to provide complex and collaboration-specific technological measures of security and control is far less exacting than is the case when the potentially conflicting interests of transient partners are involved.

#### ***e-Science and the GRID:***

Creating software platforms that can cope with the exacting data and information processing needs of geographically and institutionally distributed science and engineering research groups has been among the defining technical challenges of publicly funded programs aimed at building an enhanced infrastructure for “e-Science.” This neologism, evidently patterned on “e-Commerce”, has come into use chiefly in the United Kingdom.<sup>5</sup> In a weak interpretation, “e-Science” is the union of everything that is related to GRID enabled activities undertaken by science and engineering units (individuals or teams) or with collaboratories. Under a stronger (i.e., more restrictive) interpretation, e-Science encompasses the *intersection* of GRID and collaboratory research. In practice, within the present Pilot Projects funded in the United Kingdom by the e-Science Core Programme, there are few “non-GRID” projects, but quite a number of “non-collaboratory” projects.

For present purposes it is useful to classify collaborative research projects supported by digital networks by reference to their major purpose, rather by reference to the particular digital information tools and services they might employ. David and Spence (2003:Appendix 2) offer a taxonomy that distinguishes among the array of e-Science activities according to whether they are pre-dominantly:

- “community-centric” – aiming to bring researchers together either for synchronous or asynchronous information exchanges;
- “data-centric” – providing accessible stores of data captured or extracted from remote sources, and creating new information by editing and annotating them;
- “computation-centric” –providing high-performance computing capabilities either by means of servers accessing super-computers and parallel computing clusters, or making possible for the collaborators to organise peer-to-peer sharing of distributed computation capacity;
- “interaction-centric – enabling applications that involve real-time interactions among two or more participants, for decision-making, visualisation or continuous control of instruments.

---

<sup>5</sup> In a wave of Internet enthusiasms that also brought forth e-Government, e-Democracy (not the same as e-Government), e-Health – and hopefully, as one wit remarked, soon to be followed by e-Nough. For an overview of connections between the U.K. e-Science Programme, Grid services and high bandwidth middleware, by the Director of the e-Science Core Programme, see the presentation by: T. Hey, “Towards an e-Science Roadmap,” [http://umbriel.dcs.gla.ac.uk/nesc/general/news/ukroadmap180402/TonyHeyTowards\\_an\\_eScience\\_Roadmap.pdf](http://umbriel.dcs.gla.ac.uk/nesc/general/news/ukroadmap180402/TonyHeyTowards_an_eScience_Roadmap.pdf)

When this scheme is applied to classify the 23 Pilot Projects that have been funded under the U.K. e-Science Core Programme to develop “middleware” for the coming GRID, it turns out that the data-centric branch of the taxonomic tree emerges as far and away the most densely populated. The situation contrasts with the more uniform distribution that emerges from a comparable classification of much smaller number of pioneer collaborative projects that were organised under public funding programs in the U.S. during the late 1980’s and early 1990’s. That difference reflects in part the focus of the e-Science program on the creation of middleware platforms and tools, and in part the greater centrality of the roles that digital databases have more recently come to occupy in the work of science and engineering communities. Yet, a suspicion remains that some influence on the profile of the Pilot Project sample also has been exerted by consideration of the greater administrative complexities that would have to be overcome in order to organise thoroughly interactive modes of collaboration among research groups situated at various institutions within the U.K.

This suspicion is further reinforced by the observation that the number of distinct component products among the “deliverables” is often more or less the same as the number of “partnering” organizations. The natural supposition is that the pioneer e-Science projects have been partitioned among the collaborating parties so that cross-institutional interaction and joint responsibilities have been minimized. This could reflect an extreme division of labor along the lines of specialized expertise, but it would be surprising if such specialization ran strictly along university lines and so obviated the need to form teams by mixing researchers from different institutions. If the reality is masked by the outward appearances of the projects’ organization, one must again suspect that the latter was dictated by administrative considerations at the level of the host institutions.

In this connection it seems worth remarking on the observation that commercial developers of software for GRID services thus far have focused their efforts on intra-organizational applications, selling their commercial-off-the-shelf software (“COTS”) packages primarily as tools that will yield significant cost savings through the dependable sharing of the geographically dispersed and heterogeneous computer clusters and databases that are under the buyer’s control. The domain of true peer-to-peer inter-organizational sharing of computational resources among business entities therefore remains quite sparsely populated, and is sufficiently complex and idiosyncratic that the provision of GRID solutions there are left to consultant-developers of customized software systems. In a sense, this is the technical domain in which public sector scientific research projects involving distributed collaborating researchers and research instruments also find themselves working. Although it is still the case that many of the issues that surface inter-organizational conflicts among potential business partners are not found in the world of public research organizations, it will be seen that the challenges of negotiating formal arrangements govern cooperative research are far from trivial, and appear to be growing more and more burdensome.

## **Collaborations in e-Science: opportunities and institutional impediments**

The currently fashionable expectation, therefore, is that solving *technical engineering* problems associated with the advanced hardware and software systems of the cyberinfrastructure will unleash new scientific capabilities – leading to key discoveries, such as improved drug designs, deeper understanding of fundamental physical principles, and more detailed environmental models. But, in reality, engineering will not be enough to realize the societal gains in knowledge-creation that are being made feasible by the spectacular reduction of the marginal



social costs of information processing, reproduction and transmission, if it occurs, is likely to be the resultant of combined social, legal *and* technical transformations.

By comparison with the pace of engineering advances, progress has been slow in constructing social and legal arrangements enabling individuals, groups, and organizations to arrive at reliable and transparent agreements for the governance of collaborative work, and especially to do so in a dependably speedy fashion at affordably low transactions costs. Yet such costs, and the economic rents that otherwise cause private costs to greatly exceed the marginal social costs of effective access to data, information and knowledge in the possession of potential (and actual) collaborators. Many of the roots of the inefficiencies to which this situation gives rise will be seen to lie deep in the institutional structures that are the intermediary parties in the transactions between public funding agencies and scientific researchers, acting as agents of the public on the one hand, and as the intervening principals of the client research workers.

It is important to appreciate also that many of the technical challenges of creating a GRID services infrastructure for scientific research stem from existence of organizational boundaries that have to be respected in computational transactions because the parties involved have chosen, in effect, to protect their respective interests in that fashion—rather than by constructing a common “research space” through prior inter-organizational agreements. This situation will be seen to be quite understandable in view of the sheer complexity of the multi-organizational institutional environment and the complications that are entailed in arriving at reasonably comprehensive cooperative research agreements among them.

### ***The institutional and organizational environment of e-Science***

The institutional and organizational ‘environment’ of public sector e-Science encompasses a wide and diverse array of interrelated social, economic and legal factors that shape the utilisation, consumption, governance and production of e-Science capabilities and artefacts. Principal amongst these are the following three:

- (1) the rules and regulations of the agencies that provide grant and contract funds to researchers in public research organizations (e.g., universities, public institutes);
- (2) public research organizations’ own rules and administrative procedures governing formal relationships with their employed staff (faculty, research students and technical staff, in the case of universities), which typically will refer to elements of the external legal system (such as the statutes governing contracts, liability, privacy and intellectual property);
- (3) informal epistemic community norms and conventions, which will be recognized (if not always adhered to) by members of the various scientific and technological professional groupings, as well as some particular ‘local social norms’ that are likely to emerge among colleagues engaged in recurring or extended research collaborations.

Thus, any systematic approach to the transformation of the conduct of scientific and technological research hardly can avoid directing attention to these ‘institutional infrastructures.’ Their features are likely to turn out to be quite crucial for ensuring that the technical capabilities of advanced Internet computing and the Grid actually will be accessed, effectively applied and exploited thoroughly by researchers organizing collaborations in a variety of fields. The foregoing non-technological elements are depicted in Figure 1 (below) along with the middleware platforms and supporting layer of computer mediated communications hardware and software, as providing key infrastructural and regulatory supports of the ‘e-Science collaboration domain’. It will be noticed that each the four ‘facets’ of the tetrahedron in Figure 1 makes contact with, and hence is both bounded and supported by three other elements of the

‘infrastructure’. None of the elements exists in isolation, and hence in the long run it is appropriate to view all of them as endogenously.

The functional domain of institutional arrangements supporting scientific collaboration is thus both extensive and complex. These arrangements will govern the terms of access to and control over instruments and other physical facilities, and the data-streams generated in the research process. They will, in effect, apportion the scientific recognition and the disposition of ownership rights in collective work products created in cyberspace. They must also assign responsibilities for errors of commission and omission in those research outcomes, as well as liabilities for damages and legal infractions of various kinds arising from the actions of participants in the joint activities.

Generic collaborative arrangements of these kinds involve issues whose solutions naturally may appear quite familiar, and altogether tractable in the context of a co-located research team. Yet, the same issues quickly can become dauntingly complex when collaboration is extended to a multiplicity of geographically distributed teams and physical facilities, each of whose members have contractual relationships as employees of, or consultants to one or another among several different corporate entities. The latter, moreover, may well mix both public and private sector institutions and organizations all of which are not situated within and hence under the governance of a single legal jurisdiction and political authority.

### ***Collaborative e-Science -- promises and realities***

As has been pointed out, the e-Science label often is applied liberally (one might say, indiscriminately) to all research involving Internet communications, rather than being restricted to refer to those activities that are supported by a conjunction of GRID and “collaboratory” technologies. A defining feature of the latter is that they involve ‘virtual presence’: researchers and their research instruments and data at spatially remote locations can work together interactively, in real time. For present purposes it is useful to distinguish among collaborative research projects that can benefit from the support digital networks according to the main forms of interchanges that they involve, rather than by reference to the particular digital information tools and services they might employ. David and Spence (2003:Appendix 1.2) offer a taxonomy that distinguishes among the array of e-Science activities according to whether they are predominantly:

- “community-centric” – aiming to bring researchers together either for synchronous or asynchronous information exchanges;
- “data-centric” – providing accessible stores of data captured or extracted from remote sources, and creating new information by editing and annotating them;
- “computation-centric” –providing high-performance computing capabilities either by means of servers accessing super-computers and parallel computing clusters, or making possible for the collaborators to organise peer-to-peer sharing of distributed computation capacity;
- “interaction-centric – enabling applications that involve real-time interactions among two or more participants, for decision-making, visualisation or continuous control of instruments.

On this basis, activities belonging to the *synchronous community-centric* and the *interaction-centric* category could be deemed to come closest to realizing the proximate goals of the builders of an infrastructure for “collaborative e-Science.”<sup>6</sup> A sense of the size of the gap between the

---

<sup>6</sup> Since it is possible for interaction-centric activity to involve no more than two agents, whereas “community” implies number at least in excess of two, one can separate our asynchronous community-centric activity as a pure category, and consider as another category the combination of dyadic and polyadic forms of interactive research. This is done in applications of the taxonomy by David and Spence (2003: Appendix 1.3,Figure 2).

promise and the reality emerges immediately when the foregoing scheme is applied to classify the 23 Pilot Projects that have been funded under the U.K. e-Science Core Programme to develop “middleware” for the coming GRID environment. Middleware support for interaction-centric activities is featured by only 2 among the 23 Pilot Projects, one of which is restricted to dyadic interactions. It turns out that the data-centric branch of the taxonomic tree emerges as far and away the most densely populated, holding more than two-thirds of all the project (16 to be precise). This state of affairs may be contrasted with the more uniform distribution that is found when the same taxonomic exercise is repeated for the much smaller number of pioneer “collaboratory” projects that were organized under public funding programs in the U.S. during the late 1980’s and early 1990’s.<sup>7</sup> The difference reflects in part the focus of the U.K. e-Science program on the creation of middleware platforms and software tools, and in part the greater centrality of the roles that digital databases have more recently come to occupy in the work of science and engineering communities. Nevertheless, a suspicion remains that some influence on the profile of the Pilot Project sample also has been exerted by consideration of the greater administrative complexities that would have to be overcome in order to organize thoroughly interactive modes of collaboration among research groups situated at various institutions within the U.K.

This suspicion is further reinforced by the observation that the number of distinct component products among the “deliverables” of these e-Science Pilot projects is often more or less the same as the number of “partnering” organizations. The natural supposition is that projects forming this vanguard of the e-Science movement tended to be organized in ways that partitioned their tasks among the collaborating parties in order to minimize cross-institutional interaction and joint responsibilities. This could reflect an extreme division of labor along the lines of specialized expertise, but it would be surprising if such specialization ran strictly along university lines and so obviated the need to form teams by mixing researchers from different institutions. If the reality is masked by the outward appearances of the projects’ organization, one must again suspect that the latter was dictated by administrative considerations at the level of the host institutions.<sup>8</sup>

As collateral support for the foregoing interpretive speculations, it is relevant to observe that commercial developers of software for GRID services have to date focused their efforts quite exclusively on *intra*-organizational applications. They market their commercial-off-the-shelf software (“COTS”) packages primarily as tools that will yield significant cost savings through the dependable sharing of the geographically dispersed and heterogeneous computer clusters and databases that are under the buyer’s control.<sup>9</sup> The domain of commercially-provided software tools for true peer-to-peer inter-organizational sharing of computational resources among business entities therefore remains quite sparsely populated, as Figure 2 indicates. It is sufficiently complex and idiosyncratic that the provision of GRID solutions there has been left to the consultant-developers of customized software systems. In a sense, this is the same technical domain in which public sector scientific research projects building the means to work with colleagues, databases and equipment at other laboratories and field research sites, find themselves engaged. Although it is still the case that many of the issues bring to the surface inter-

---

<sup>7</sup> For further details, including descriptions and characteristics of the project involved, see David and Spence (2003: Appendix 1.3).

<sup>8</sup> At present these conjectures are wholly speculative. An effort is underway to obtain support for an interview-based study that would elicit information about the considerations entering into the Pilot projects’ designs and organization structures.

<sup>9</sup> See, e.g., Keith Norman, *Grid Computing* (Issue V1.R3.MO). Tessela Support Services plc, February 2003).

organizational conflicts among potential business partners are so acutely present in the world of public research organizations, the challenges of negotiating formal arrangements governing cooperative research there are hardly trivial. Indeed, they are growing more complicated and more burdensome.

It should be evident that the complex collaborative undertakings in view here – those that are meant to be enabled, indeed, empowered by e-Science facilities and services – cannot be supposed to arise and function automatically as ‘perfect teams,’ expressing some primitive cooperative impulse among the human actors. Quite the contrary: even non-commercial research collaborators will need to find solutions for non-technological issues of resource allocation and governance that involve conflicts arising from the divergent interests of the individuals and organisations involved. Moreover, to sustain extended programs of research that continue to build upon and utilize the specialised knowledge that they generate, those solutions must be sufficiently flexible to accommodate the high order of uncertainty that inevitably surrounds research activities. That is especially so for fundamental, exploratory research programs of the sort for which public support is particularly warranted. Only the satisfactory resolution of those conflicts will permit realisation of the gains from cooperation. But, it is important not to lose sight of the reality that ‘conflict resolution’ is not a costless process. Consequently, the means by which such solutions are arrived at ought not impose heavy ‘transactions costs’ upon the parties, thereby draining resources from the conduct of research itself, or, worse still, undermining whatever cooperative spirit and ethos of common purpose initially animated the collaborative enterprise.

Achieving the aims and aspirations of e-Science is thus not just a matter of breakthroughs in hardware or software engineering, or system design improvements to provide tools that will be readily useable by individual researchers and their organizations – as challenging as those engineering tasks may be. Nor is it matter of providing equitable access to research tools and networked computer equipment to scientific and engineering personnel in many regions of the world where lack of such facilities renders it difficult to make use of the enormous amount of data and information that presently can be readily access in digital form. The informal norms and formal rule structures for collaboration on the ground also set the conditions and costs of effective access. So too do the agreements governing data-stream management and control in “virtual laboratories,” in federations of annotated dynamic data bases, in the on-line publication archives, and in repositories” of software tools needed to search, display and manipulate digital information. To the extent that these must be collectively constructed and maintained, as well as collaboratively used and refined, a web of formal and informal policies and understandings among individuals, their host institutions and the agencies that support their research is required to enable effective research collaboration on a global scale to take place.

An “institutional infrastructure” already exists, comprised of the public and private policies, administrative arrangements, and legal rules that both constrain and facilitate the formation of various research collaborations that are forming within and across disciplinary, university, and national boundaries. The questions that need to be addressed are whether the existing institutional infrastructure is congruent with the aspirations of those who are fashioning an enhanced technical infrastructure of collaborative research; whether the directions of change occurring in the several elements of that infrastructure will remove serious impediments in a timely fashion; what remedial actions, if any, might be called for; and how best to identify and motivate the most effective agents of institutional change in this complex multi-actor domain.

### The legal framework for scientific collaboration—a brief overview

One part of the existing institutional infrastructure reflects the legal framework within which formal, contractual agreements among public agencies and research performing organizations, and among such organizations, will be constructed. Recently there has been attention devoted to the role of intellectual property law in the formation and conduct of scientific and technical research collaborations.<sup>10</sup> Getting the balance wrong between the ownership of, and access to knowledge resources entails serious social costs that recently have begun to be perceived more widely, beyond the boundaries of the scientific research communities that are immediately involved.<sup>11</sup> But, it is surprising is how few people have recognised that intellectual property rights are only one among the many kinds of legal issues that need to be successfully resolved to facilitate collaborative work.<sup>12</sup> Collaboration among researchers can be affected by the entire complex of legal norms and informal professional conventions. It is important that institutional arrangements are made so as to minimise the extent to which the law becomes an impediment to cooperation among researchers, whether directly, or indirectly by undermining informal mechanisms of trust and dispute resolution

A collaborative project might encounter and one or all the following four principal classes of legal problems, and consequently these become the issues to which responsible counsel, prior negotiations and formal contracting will need to address:

- a) legal relationships among the parties to an e-Science collaboration, particularly where some of the parties are operating in different jurisdictions;
- b) materials that each party brings to a collaboration;
- c) resources, if any, to which the collaborative project will give rise;
- d) apportioning (among the parties) liabilities for potential harms arising from the collaborative project.

In relation to each category of issues, the law offers “solutions” to the problems or procedural mechanisms that may be more or less satisfactory from the point of view of the researchers and organisations that are involved. These answers flow from the general law in areas as diverse as contract, conflict of laws, arbitration and civil procedure, data protection,

---

<sup>10</sup> For one of the few empirical studies that presents information on the influence of intellectual property rights considerations on the negotiation of inter-organizational research agreements among business firms, and between firms and universities, see Henry R. Hertzfeld, Albert N. Link, and Nicholas S. Vonortas, “Intellectual Property Protection Mechanisms in Research Partnerships,” forthcoming in *Research Policy* (Special Issue on ‘Property and the Pursuit of Knowledge’, Guest-edited by P. A. David and B. H. Hall), Winter 2005.

<sup>11</sup> Concerns about the recent thrust of public policy on this score have emerged more strongly in recent years among academic lawyers and economists in the U.S. See, e.g., James Boyle, ed., “The Public Domain,” *Law and Contemporary Problems*, 66(1&2), Winter/Spring, 2003 (Special Issue of the Collected Papers from the Duke University Conference, held November 2001); J. H. Reichman and P. F. Uhler, “A Contractually Reconstructed Research Commons for Scientific Data in a Highly Protectionist Intellectual Property Environment,” *Law and Contemporary Problems*, 66(1&2), Winter/Spring, 2003: pp.315-462; P. A. David, “Can ‘Open Science’ be Protected from the Evolving Regime of Intellectual Property Protections?” *Journal of Institutional and Theoretical Economics*, 123 (2004). [Pre-publication version available at: [http://siepr.stanford.edu/papers/papersauth\\_D-H.html](http://siepr.stanford.edu/papers/papersauth_D-H.html)].) For views within the scientific community, see, e.g., The Royal Society, *Keeping Science Open: The Effects of Intellectual Property Policy on the Conduct of Science*, Policy document 02/03, April 2003 [Available at:<http://www.royalsoc.ac.uk>].

<sup>12</sup> This discussion draws upon a more extensive discussion of the legal context of collaborative activities in sect. 1.4 of P.A. David and M. Spence, *Towards Institutional Infrastructures for e-Science: The Scope of the Challenge* [Final Report to the Joint Information System Committee of the UK Research Councils], Oxford Internet Institute Reports, No.2, September 2003. [Available at: [http://www.oii.ox.ac.uk/resources/publications/OIIRR\\_E-Science\\_0903.pdf](http://www.oii.ox.ac.uk/resources/publications/OIIRR_E-Science_0903.pdf)].

intellectual property, competition law and torts. Broadly speaking, all of these problems stem from the potential for disputes among the various parties to the collaboration (i.e., among the participant scientists, their respective host institutions and the private and public funding bodies that are sponsoring the project). It must be recognised that, in addition, disputes may arise between the parties to the collaboration as a group and any of a variety of “outsiders” – private individuals, other universities and institutes, public regulatory authorities. Often there is little that the parties involved in a collaborative project can do in relation to disputes with outsiders, except to be aware of the law in planning their own internal relationships. They may therefore decide how the risks of liability suits by outsiders are to be apportioned by using devices such as indemnity clauses); and take out appropriate insurance, insofar as it is made available. By careful planning, parties to collaborations can avoid the sanctions of competition law and can allocate the risk of liability in, for example, tort to parties harmed in some ways by the collaborative research.

Parties to a given collaboration have other means of controlling the terms of their own relationships, even where the latter have been constructed with forethought and suitable legal guidance. Relationships among academic researchers traditionally have been governed by informal norms operating within particular scientific communities. The workings of these norms and conventions (for example, in relation to ordering of names in multi-authored papers, and more, generally, the attribution of credit for research findings) might not always have been perfectly just; but they were well understood and broadly accepted. As collaborative science has come to involve larger teams of people operating in more diverse contexts – researchers in different national communities, scientists in different scientific disciplines, researchers who are primarily publicly funded and those who are primarily privately funded – the clarity of these informal norms tend to become blurred and their force in guiding individual behaviours weakens.

The core of many of the difficulties arising in the contractual organization of scientific collaboration is that the actual work is to be done by individuals in laboratories, but the agreements that underpin collaborations are usually made by the institutions which employ them. It is appropriate that scientists should be relieved of the burdens of negotiating contract details. Yet, taking the contracting process out of their hands presents a number of dangers. One likely difficulty is that the process of setting the terms of inter-institutional collaborations might be affected by the conflicting interests of the university or other host institution. This problem often is very real and may be exacerbated by the structures for obtaining legal advice that operate in most universities: legal counsel have the responsibility to protect the institution from the hazards of entering into collaborations, “hazards” that include emerging from a collaborative undertaking with a visibly smaller share of the gains than other parties have enjoyed. In the calculus of “due diligence” the lawyers are predisposed to protect the immediate and palpable interests of their client, the university, whereas the researchers are left, less comfortably, having to decide whether to argue for their own career interests or for the more transcendent and speculative benefits that society at large might derive from the proposed project.

In a collaboration in which the participating institutions are contributing components that are complementary, there is an understandable temptation for each of the parties to try to extract as large a part of the anticipated fruits as they can. But this is likely to result in reducing the efficiency of the project design, as well as in a protracted and costly bargaining process. Inter-institutional conflicts over research credits and intellectual property rights can only become more difficult if the parties try to anticipate the consequences of the increasingly mobile pattern of employment among academic researchers in the sciences. Yet, perhaps the most formidable problems are likely to stem from the fact that the universities will be entering into agreements about matters (such privacy of personal data) on which their powers to assure

delivery are highly uncertain, and which can expose them to considerable risks. The quite reasonable nervousness on the part of responsible administrators and their respective legal counsels may adversely affect the traditional structure of the institutional relationships under which academics work. The effect of each party to the collaboration seeking to protect itself at the expense of the others tends to raise the costs of the entire undertaking.

The challenge in designing appropriate legal arrangements for collaborative e-Science is, therefore, to construct agreements that are adequately clear and determinative without damaging the trust and informal norms essential to the day-to-day conduct of collaborative research; and to provide processes for constructing those agreements that involve the scientists without unduly burdening them with negotiations over legal complexities. Some adverse consequences of the introduction of formal, contractual norms may not be avoidable, since these may displace the efforts that the parties might otherwise devote to resolving conflicts informally. But, the goal must be to avoid the worst outcomes.

### **Broadening the “information commons” and contractually constructing “collaboration spaces”**

Growing awareness of the encroachments that have been made into the public domain in scientific and technical data and information over the past two decades --primarily as the unintended consequence of the privatization of government data and research functions, and the pressures to extend and strengthen legal protections of intellectual property rights – is now stimulating a growing counter-movement. This has been marked by new initiatives to preserve and in some areas significantly enlarge the domain of “open access” and reduced costs of data exchanges through the institutionalization of “open standards.” Much, but by no means all of the effort to explore and apply new paradigms for the organization of virtual knowledge-based communities, and the distributed production of new data and information, have roots in the historical practices and habits of mind that developed in public science. Examples include the open-source software movement, “libre source” tools for free and open source software development, open public-domain data archives and federated data networks, community-based open peer review, collaborative research Web sites, collaboratories for virtual experiments, virtual observatories, and open access on-line journals.

Open access to the research literature produced from public funding is a major issue that has received considerable scrutiny in the past few years, particularly as the rising prices of commercially published scientific journals collided with the constricted resources of university libraries. There are now over 1000 scholarly journals provided under open access conditions on the Internet. This has been made possible by numerous “open access publishing” initiatives, including notable initiatives such as the Public Library of Science and BioMed Central. Policy principles on open access to journal articles reporting findings from publicly funded research were issued in both the United States and Europe in 2003 through the "Bethesda Principles" and the "Berlin Declaration." In 2004, many professional society journal publishers produced the "DC Principles," which also recognized the imperative of broad access to the scholarly literature produced from publicly funded research. Experiments with variety of new business models for scholarly and scientific publishing have been encouraged, and the flexibilities of differential pricing that are inherent in the traditional, “subscription” model also has been utilized in efforts to reduce the costs of information access to researchers in the developed and the developing and transition countries alike. New initiatives also have been established for pre-prints and e-prints of journal articles (e.g., Stanford University’s HighWire Press, and the Cornell arXiv, originally created for high-energy physics and now expanded to include other areas of physics, mathematics, computer science, and computational biology), for individual research articles and

other information resources (e.g., the Social Science Research Network, the MIT D-Space initiative), and for university educational material (e.g., MIT's OpenCourseWare).

Taken together, these initiatives and emerging capabilities can be seen to form a broader trend toward both formal and informal peer production of information in a highly distributed, volunteer, and open networked environment. Such activities are imbued with and reflect the cooperative ethos of rapid and complete disclosure of new knowledge that traditionally guided the organization and conduct of publicly supported scientific research. They are indeed based on principles that can be characterized as those more suited to the governance of a scientific and technical “information common,” rather than the rules, regulations and behavioral norms for commercial transactions in (intellectual) property.<sup>13</sup> References to “the digital commons,” and the “information commons” now abound, evoking in allusive, metaphoric terms the idea of “the common” – a collectively held and managed resource, to which access by cooperating parties is open and subject to minimal transactions costs. It is important to clarify the connotations of this term, so that the nature of the challenge of broadening the information commons will be grasped from outset to be one of building new social and legal structures, and not confused with a utopian dream of returning to some imagined golden age when property did not exist. The metaphoric allusion to “the common” is quite apposite where the resources in question take the form of information, which is not like ordinary tangible commodities, but instead possesses inherent properties that economists associate with the so-called “public goods”.<sup>14</sup> On the other hand, if the contrast between “common” and “private” is helpful, the juxtaposition of “common” with “private property” can be misleading: historically, the “common lands” of Europe’s agrarian communes were neither a wilderness nor an unregulated part of the settled domain; non-villagers did not enjoy access rights and collective possession did not translate into egalitarian distribution of use-rights. Moreover, the modern example of free and open source software, shows how the legal framework of copyright may place in the hands of the owner the power to set contractual terms that emulate desirable features of the public domain in data and information.

In somewhat the same spirit, it has been proposed to utilize the lever of contract law and the fulcrum of legally enforceable property rights to lift from would-be collaborators in pursuit of knowledge the burdens of excessively high transactions costs and oppressive charges for access to public goods in the form data and information.<sup>15</sup> Appropriate institutional

---

<sup>13</sup> Emblematic of these spontaneous, bottom-up developments is the international conference series of the Wizards-of-OS (Operating Systems), the foundations of which are described by the organizers as rooted in “the grand liberation movements in the realm of knowledge: free software, free content, free science, free networks, free hardware.” The “WOS 3” conference to be held in Berlin on 10-12 June, 2004 on “The Future of the Digital Commons,” will feature presentations and discussions of a wide array of working initiatives ranging from a variety of open access publishing and alternative copyright licensing arrangements (particularly those provided by “Creative Commons” which will be launched in Germany at this event), open standards and open source software, and still other virtual community projects such as Wikipedia (the free online encyclopedia) and Simputer, the free and open hardware design project. See: <http://wizards-of-os.org/index.php?id=50&L=3>

<sup>14</sup> Information is not exhausted by use, may be utilized concurrently by many, and require significant additional resource expenditures to prevent them from becoming ubiquitously accessible. The lattermost among these properties reflects a general condition that the progress of digital information technologies has rendered manifest: the marginal costs of reproducing and distributing information today are negligibly small, both absolutely and in relation to costs of creating “the first copy.”

<sup>15</sup> The particular proposals that are briefly indicated here are those put forward by David and Spence *Towards Institutional Infrastructures for e-Science* (2003). But, save for the details, share the same perspectives and approach to constructing “collaboration spaces” and broadening the information commons as that found in J.H. Reichman and P.F. Uhler, “A Contractually Reconstructed Research Commons for Scientific Data in a Highly



mechanisms for the organization of e-science cannot simply be legislated or put in place by administrative fiat, even if the policy climate was more receptive to the notion that this is an important matter to which political leaders should attend. Similarly, the problems created by the international nature of collaborative e-Science cannot be left to be solved by the international harmonization of formal legal rules. Legislation and the harmonization of legal rules have a potentially stultifying effect on the development of new and more appropriate institutional mechanisms. When legislation is enacted and international conventions are agreed, they tend to have the effect of petrifying the norms regulating a given area of behavior. In any case, the international harmonization of legal rules is a slow and frustrating process, which in the end is not likely to be effective. Harmonization would be a particularly daunting task given the range of legal issues that might impact upon the conduct of collaborative on-line research. Further, the harmonization of legal norms is only even partially effective in assuring that disputes determined under the same norms will find the same result in different courts. (The history of the European Patent Convention, for example, shows that the same norms can lead to different outcomes in different courts with different interpretative traditions.) To establish norms that can facilitate collaborative e-Science, we must therefore look elsewhere than to formal law reforms and legal harmonization.

Acknowledging these realities, David and Spence (2003) have argued for a more “bottom up” approach to constructing appropriate institutional infrastructures for e-Science, one that calls for the creation of a coordinating and facilitating mechanism in the shape of a novel public agency. Their report to the Joint Information Systems Committee of the U.K. research councils envisages the establishment of a new independent body to be called the Advisory Board on Collaboration Agreements (ABCA). Its remit would be to guide, oversee and disseminate the work of producing, maintaining, evaluating and updating standard contractual clauses, those being the constituent elements from which formal agreements may be more readily fashioned by the parties undertaking particular ‘Grid-enabled’ collaborations in science and engineering research. This advisory body would, of necessity, play a leading role in enunciating a set of fundamental principles to guide the formulation of those contractual clauses, and thereby ensure that the effects of the agreements into which they are introduced will not be inconsistent with the intent underlying those principles.

In other words, what is proposed is the establishment of a new “public actor”, an independent entity with on-going powers to initiate, co-ordinate and provide resources required to support and, above all, articulate principles for developing, an array of model contractual clauses, each of which would treat some specific problem (among the myriad legal issues that have been seen to arise from the formation of research collaborations). Included among these specific problems would be such questions as those concerning appropriate forms of licensing for middleware and higher level software applications; and terms of the private contracts that holders of copyrights might utilize in so-called “dual licensing” of GNU General Public License software in order to permit third party commercial exploitation of publicly funded software systems. Much of this detailed work could be entrusted to specialised task force-like “study committees” comprising individuals with diverse expertise: scientists and engineers familiar with the organization and conduct of collaborative projects, legal scholars and practitioners, social scientists with expertise regarding the workings of academic research institutions, and others with detailed knowledge of the policies and administrative rules of pertinent funding agencies in the U.K. and abroad.

## Collaboration institutions for the cyberinfrastructure: a policy and research agenda

The practical challenge is to develop and adapt new mechanisms for the governmental agencies and public research institutions to more rationally manage publicly (and charitable, quasi-public) funded data and information production and distribution in the rapidly progressing digitally networked research environment.<sup>16</sup> Combined with that it is necessary to address the myriad obstacles that raise the transactions costs and restrict the terms of collaborations that require inter-organizational agreements involving universities and other non-profit research institutions.

Recent proposals suggested for adoption by government science agencies have included the following:

- Provide funding for public domain or open access data centers and active archives of foundational data sets derived from publicly funded research as an essential component of the public research infrastructure.
- Review and enforce research contract and grant clauses regarding open data availability and use.
- Develop guiding principles and contractual provisions for licensing data products and services to or from the private sector, or for privatizing such government activities, that protect research user interests.
- Provide open access to the scientific literature produced by the government, and promote open access to the scientific literature produced through government-funded research.
- Encourage universities to establish broad inter-institutional agreements and specific contractual provisions for maintaining open access to, and use of data in the academic sector (for non-commercial purposes).
- Work with foreign governments and their science agencies through multilateral and bilateral *fora* to promote public domain status or open access to their public data resources on a reciprocal basis.

Recommendations that have been directed specifically towards facilitating research collaborations based in universities, institutes and other public research organization, including the recent proposals by David and Spence (2003) and Reichman and Uhler (2003) for an immediate start on the work of “contractual reconstruction and expansion” of the scientific research commons, call for :

- Developing inter-institutional agreements and cooperative institutional approaches to ensure unimpeded access to, and liberal uses of scientific data and information in a non-profit research commons, while allowing for commercial exploitation of those resources vis-à-vis the private sector.
- Formulating model contractual provisions for inter-university and inter-researcher relationships in order to both protect access to and unrestricted use of publicly funded research data and information, and reduce the costs and delays of negotiating agreements governing the terms of projects requiring effective inter-organizational collaboration.

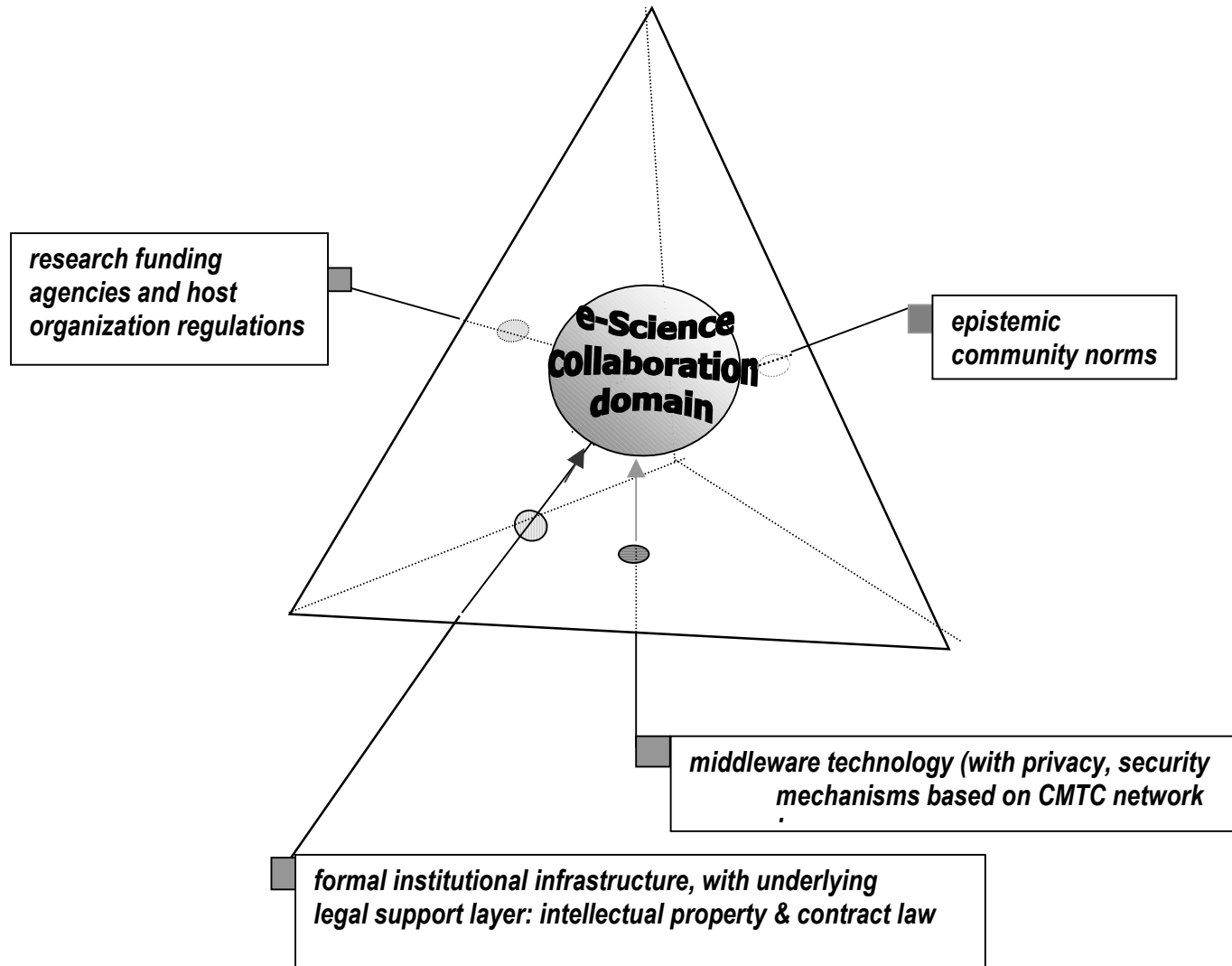
---

<sup>16</sup> The material in section is drawn directly from P. A. David and P. F. Uhler, “Broadening the Information Commons for Science and Innovation: Strategic Institutional and Public Policy Approaches,” Draft Proposal for the Planning Committee on the 2005 CODATA-ICSTI-U.S. NAS Workshop ( May 18, 2004)

- Vigorously promoting non-exclusive licensing by authors of their scientific articles to science, engineering and medical journals, rather than assigning exclusive copyrights.
- Seeking the introduction of exceptions to existing and pending legislation for the legal protection of database rights, restricting the renew-ability of such rights and allowing “fair use” extraction of contents by non-profit entities engaged in research and teaching.
- Approaching these initiatives through strategic international initiative that take account of the international scope of digital computer-mediated networks, research collaborations, and the prevailing international conventions protecting intellectual property rights.

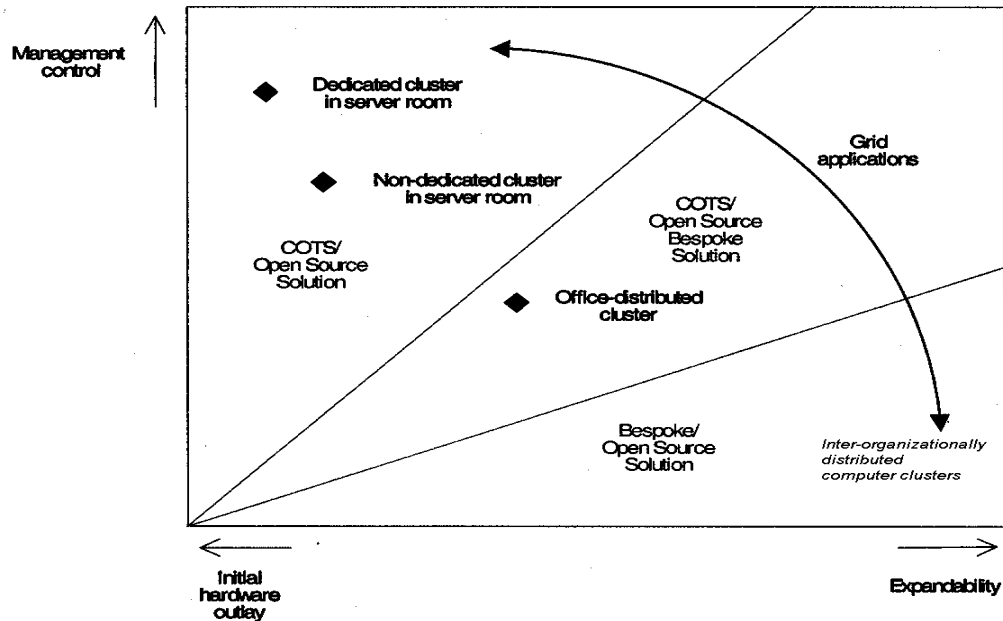
Lastly, as bromidic and predictable as the academic’s closing plea for “further research” may be, surely it will be accepted as thoroughly justified here. There is a largely unmet need for empirical assessments of the nature and severity of the varied impediments to an effectively functioning infrastructure for publicly supported scientific and technological collaborations. Intrinsically interesting methodological challenges, as well as difficult data collection tasks lie along the path to developing systematic measures of the effects of the incentives and constraints of such undertakings that are created by prevailing organizational norms, institutional rules and governmental policies. A better understanding of their differential impacts upon the direction and conduct of research projects in the various domain sciences, and on exploratory work in emerging transdisciplinary fields would be of real value in identifying specific targets for remedial attention. Only on the basis of such knowledge will it be practical to formulate and implement coordinated strategies of private and public action that have a good prospect of freeing distributed collaborative research from the persisting constraints of a mal-adapted institutional infrastructure.

**Figure 1**  
***e-Science Collaboration Domain and Infrastructural -Regulatory Supports***



**Figure 2**  
**Distributed computing modes and the domain of “grid applications”**

**COTS**, i.e., *commercial off-the-shelf software* packages implementing distributed computing architectures currently are not available for “geographically distributed clusters” outside the organization’s firewall.



Source: Keith Norman, *Grid Computing* (Issue V1.R3.MO, Tessela Support Services plc, February 2003), with modifications.