TOWARDS INSTITUTIONAL INFRASTRUCTURES FOR E-SCIENCE:
The Scope of the Challenge

By

Paul A. David

Senior Fellow of the Oxford Internet Institute,
Professor Emeritus of Economics & Economic History in the University of Oxford
and Professor of Economics, Stanford University

and

Michael Spence

Faculty Associate of the Oxford Internet Institute,
Fellow of St Catherine's College
and Chair of the Law Board in the University of Oxford

15 September 2003

This document is the Final Report of the OII project on "The Institutional Infrastructure of e-Science: The Scope of the Issues." Work on that project was initiated with the support of a grant from the Joint Information Systems Committee (JISC) to the University of Oxford, and was carried out under the direction of P. A. David and W. H. Dutton.

THE OXFORD INTERNET INSTITUTE
1 St. Giles
Oxford OX1 3JS

Tel.: 0044+(0)1865+2878210
www.oii.ox.ac.uk

Acknowledgements

The interest and encouragement received from Tony Hey and Anne Trefethen (the Director and Deputy Director of the U.K. e-Science "Core Programme", respectively) were absolutely vital to the research project whose results are reported here. Many others from the e-Science community also were supportive: Paul Jeffreys, the Director of the Oxford e-Science Centre held an early orientation meeting met with members of the OII's research team, and James Fleming of the EPSRC arranged for the work of this project to be introduced to and elicit comments from the audience at a plenary session of the e-Science Core Programme's Pilot Projects Workshop, held on 30-31 January 2003 at the Edinburgh e-Science Centre. Helen Cordell, of Argonne National Laboratory, Chicago, kindly granted an extended interview which provided unique insights based upon her legal work concerning the development of the Globus Toolkit Public License.

Matthijs den Besten, the project's research officer, conducted that interview during a visit to Argonne and furnished excellent research assistance on many other matters in addition to those acknowledged explicitly in Appendices 1-4. Previous drafts of this document were improved by the penetrating and constructive comments received from W. Edward Steinmueller (University of Sussex) and Paul Uhlir (National Academy of Sciences). Also beneficial were the many points raised by participants in a day-long Workshop on "Institutional Infrastructures for e-Science," held by the Oxford Internet Institute on 2nd May 2003 at the Said Business School. The entire diverse group of experts who assembled on that occasion to lend their assistance to the project are too numerous to be thanked individually here; their names and affiliations appear in Appendix 7 of the Report. Jean-Michel Dalle and Rishab Ghosh, however, deserve special acknowledgement for the guidance they provided on the subject of free and open source software licensing.

Miranda Turner, Programs Assistant at OII was superbly efficient in handling all the logistical arrangements for the Workshop. Lastly, but not least, the authors are indebted to Bill Dutton, the OII's Director, for early and sustained encouragement of this project, as well as helpfully forthright advice on the organisation of the Workshop and the drafting of the Report.

The opinions expressed in this Report are those of the authors and do not necessarily represent the view of the Oxford Internet Institute, the Joint Information Systems Committee (JISC), nor should they be ascribed to the individuals whose assistance is acknowledged here with much gratitude.

Table of Contents

Summary

Purpose of the Report

This Report has a three-fold purpose:

to articulate the nature and significance of the non-technological issues that will bear on the practical effectiveness of the hardware and software infrastructures that are being created to enable collaborations in *e-Science*;

to succinctly characterise the fundamental sources of the organisational and institutional challenges that need to be addressed in regard to defining terms, rights and responsibilities of the collaborating parties, and to illustrate these by reference to the limited experience gained to date in regard to intellectual property, liability, privacy, and security and competition policy issues affecting scientific research organisations; and

to propose approaches for arriving at institutional mechanisms whose establishment would generate workable, specific arrangements facilitating collaboration in e-Science; and, that also might serve to meet similar needs in other spheres such as e-Learning, e-Government, e-Commerce, e-Healthcare.

The Role of Institutional Infrastructures in Realising the Potential of e-Science

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 potential for these advances in technology to support new levels of research collaboration and transform the conduct of science is a major driving force behind the UK's e-Science Programme. It also animates parallel efforts by other European Community member states, and the U.S., to greatly enhance the hardware and software network facilities that are available to support computation intensive research activities of all kinds.

These technical developments will not achieve their transformational potential, however, unless the institutional context in which they operate is supportive of the organisation of collaborative research programmes in science and technology. This Report reveals the challenges that e-Science currently faces, not from hardware and software engineering problems, but from the features of the existing legal and administrative regimes that cannot adequately accommodate the changing realities of scientific practice in the networked digital environment.

More positively, the Report offers a response to these "non-engineering" challenges, by articulating a general approach to creating institutional infrastructures that are better adapted to the emerging opportunities for global research collaborations in many scientific domains. Pursuing that approach, it indicates some specific steps that the e-Science Core Programme could take towards practical implementation. It envisages, as a first step, the creation of an interim Working Party on Collaborative Agreements that would report to the Joint Information Services Committee; and the eventual formation of an independent e-Science Advisory Board on Collaboration Agreements (ABCA).

The Legal Context of Collaborative Activities: the Shape of the Problems

There has been considerable discussion of the impact of intellectual property rights on scientific collaboration. Getting the balance wrong between the ownership of, and access to knowledge resources entails serious social costs that recently have been perceived more widely beyond the boundaries of the scientific community. 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. 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.

Four different types of legal problems that a collaborative project might encounter are considered by this Report. These problem-classes are concerned with:

(1) legal relationships among the parties to an e-Science collaboration, particularly where some of the parties are operating in different jurisdictions;.

(2) materials that each party brings to a collaboration;

(3) resources, if any, to which the collaborative project will give rise;

(4) apportionment (among the parties) of liability 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, 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 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.

This effect is exacerbated by a new, more competitive and more entrepreneurial spirit in at least some of the branches of public science and engineering – that is, research carried on in universities, non-profit institutes and other public research organisations (PROs). Accordingly, the rules under which individuals utilise their institution's resources, and the terms governing collaborations among researchers in different departments, and different institutions, need to be more explicitly articulated in reference to codified (legal) relationships. In that process, there is an opportunity for the parties to the collaboration to take control of determining those rules, rather than having them imposed by others without due reference to the particular scientific needs of their project. In particular, the parties involved can construct contractual arrangements that vary the general law so as to establish their own agreed terms of collaboration and minimise the possibility of unexpected disputes, or of unexpected outcome from the application of the general law.

Three problems arise from this approach. The first is that the law of contract is not always completely effective at delivering "what the parties want." Most often such failures occur because disputes arise over causes that were not anticipated, and because there are misunderstanding among the parties as to the implications of the terms of the agreement into which they entered. There is little that can be done about those possibilities, except to ensure that the tribunals that eventually interpret any agreement for scientific collaboration understand the norms that might generally be assumed as background in the relevant scientific communities. This may require that the contractual agreements involving scientific collaborations be made subject to the jurisdiction of specialist tribunals, such as arbitrators, and not to the jurisdiction of the general courts.

The second problem is that the culture of *ex ante* contracting is in some ways inimical to the establishment of effective relationships of trust and collaboration. Lawyers devise contracts by imagining everything that could go wrong with a particular relationship and either providing solutions to those possible disputes, or mechanisms for resolving them if and when they arise. But, scientific research – and particularly exploratory research of the sort carried on within academic institutions – inevitably takes unexpected turns. Scientists may well feel that it is counter-productive to address their imaginations to possible future disputes at the same time as they are trying to establish effective working relationships among their colleagues. This is particularly true because the researchers are not the only parties concerned: their host institutions, and the public and private funding bodies on each side of the prospective collaborative agreement, will have interests in common that are distinct from those of the research scientists. They also are likely to have some conflicting interests. In particular, the host institutions may take a different view from those of either the funding agencies or the research scientists in regard to the balance between the risks to which they may be exposed by a particular project and its potential private and social benefits.

Such problems are not as easily overcome by standard form agreements as public research administrators and some members of the scientific community would appear to believe. Standard form agreements carry a variety of legal risks of their own (which the Report details). But, the most obvious problem is that the structure of different types of collaborative agreement is going to vary widely, making it extremely difficult to determine an appropriate set of standard form agreements for the field of collaborative science. The situations in which standard form agreements operate most effectively are those in which the problems to be disposed of by agreement are highly predictable and the parties are in any case professionally well advised (for example, in domestic property conveyancing, and in certain types of international shipping.)

In addition, there are complications that arise from the element of novelty and fluidity in the present situation. If we now stand at the beginning of a new era of scientific collaboration, based upon high bandwidth telecommunications and grid-enabled computing, then standard form agreements put in place at this juncture carry the risk of ossifying the development of appropriate norms for e-Science. This is the non-technical form of the classic problem of "anticipatory standard-setting." The absence of extensive experience and the weight of precedence concerning arrangements in the new environment create an opportunity to exercise greater leverage over the future evolution of standards by setting them early and firmly. Yet, the very same conditions make it difficult to gauge what the new standards should be.

The third problem that arises in the contractual organisation 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 potential 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.

Furthermore, there is a danger that the contractual norms under which the project is to operate are at odds with the traditional research norms upon which the scientists might be relying. The emphasis that the Report places upon the importance of developing an institutional context for e-Science collaborations that is supportive of productive "open science" research, should be seen as fully consonant with the most recent and enlightened reconsideration of the impacts of intellectual property law upon the pursuit of reliable knowledge.

In this regard, its thrust is consistent with several of the detailed points recently advanced in recommendations for intellectual property law reforms. For example, the Royal Society's report, *Keeping science open: the effects of intellectual property policy on the conduct of science* (April 2003), criticises the introduction into U.K. law implementing the EC Database Directive (1996) and the Copyright Directive (2001) of narrow and ill-defined limitations on the "fair dealing" exceptions provided for research; in a departure from traditional practice, the statutes confine the exceptions to research that has "non-commercial purposes." Further, the Royal Society's report points out that, in line with the EC Database Directive, the U.K. statutory implementation confines the fair dealing exception for research (and education) to allowing extraction and not re-utilisation of the protected material. It thereby fails to address the scientific community's needs in using these increasingly important research facilities. Nevertheless, the policy of governments in the U.K., the E.U. and the U.S. continues to encourage universities to exploit their rights to intellectual property arising from publicly funded research in science and engineering.

Unfortunately, in a collaboration in which the participating institutions are contributing components that are complementary, the temptation for each of the parties to try to extract as large a part of the anticipated fruits as they can 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 be 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 understandable 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. This is the challenge that is addressed by the mechanisms and procedures proposed in this Report.

The Way Forward: Contractually Constructing Collaboration Norms

Appropriate institutional mechanisms for the organisation of e-science cannot simply legislated or put in place by administrative fiat. Similarly, the problems created by the international nature of collaborative e-Science cannot be solved by the international harmonisation of formal legal rules. Legislation and the harmonisation 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 behaviour. In any case, the international harmonisation of legal rules is unlikely to be effective. The international harmonisation of law is a slow and frustrating process. Harmonisation 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 harmonisation 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 harmonisation.

The Report recommends a more “bottom up” approach to constructing appropriate institutional infrastructures for e-Science, one that calls for the creation of a co-ordinating and facilitating mechanism in the shape of a novel public agency. We envisage the establishment of an independent body which is referred to as 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 that 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 utilise in so-called "dual licensing" of GNU General Public Licence 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 organisation 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.

The activities of the study committees organised under the auspices of the ABCA would focus upon framing appropriate standard contractual clauses that could be readily assembled into a variety of alternative collaboration agreements, much in the same way that software sub-routines and modules can be assembled into functionally more comprehensive software systems that are suited for particular applied tasks. As part of its supervisory and co-ordinating role, the ABCA would have to consider the underlying principles that will be implemented through the contractual clauses of those agreements. It would also need to determine the best ways of organising the accumulation and dissemination of information and analyses concerning the actual formulations and manner of implementation of contractual agreements. These principles for the establishment of e-Science collaborations and model contractual clauses could then be put into effect in individual cases by universities and research bodies.

Although the proposed Board would be something of a novelty on the U.K. and international public sector research scene, a number of interesting institutions operating in other spheres currently provide concrete and successful models for an entity charged with carrying out some of the key functions envisaged for the ABCA. The part of the Board's activities that involved promoting principles of best practice in establishing institutional arrangements might find a parallel in the work of the Basel Committee on Banking Supervision, which assists the Governors of the Central Banks of thirteen countries to develop common principles for banking supervision. The Basel Committee does not make laws of any kind, but it builds consensus between important actors in the international banking community and its recommended standards are given appropriate effect by relevant actors in local contexts.

This approach has the advantage of flexibility and that principles develop slowly rather than being imposed at what might be an inappropriate stage in their articulation.

Establishing a U.K. body in this mould would be an important step in establishing socially desirable rules for the organisation of e-Science at the national level. Indeed, it is likely that the proposed Advisory Board could set a lead in the international organisation of e-Science collaborations; its principles and contractual clauses could well become a *de facto* standard that would be more widely adopted. To be sure, greater initial momentum would be imparted to such a process were the national institutions receiving grants and contracts from public funding agencies to be required to construct their collaboration agreements using the clauses recommended by the ABCA.

The part of the Advisory Board's work that involved the development of standard contractual clauses might also find a parallel in the work of the Basel Committee and, with one important reservation, in that of the Grain and Feed Trade Association ("GAFTA"). The Basel Committee makes recommendations for contractual clauses in certain areas of banking practice as a means of ensuring that the principles it articulates are given appropriate and certain effect. Central amongst the clauses that the Board might be expected to suggest would be those reflecting a consensus as to the appropriate *fora* for the resolution of disputes under e-Science collaboration agreements, particularly those involving parties from different jurisdictions. A model of how such choice of forum clauses might operate is found in the standard form agreements established and maintained by GAFTA, which has 80 different standard contracts that contain arbitration clauses allowing parties to make use of the GAFTA Dispute Resolution Service. This service provides for the speedy and final resolution of disputes in an expert forum that is cheaper and quicker than traditional legal systems; further, it has the advantage of being outside the legal system of any of the parties to the international collaboration. In the event that arbitration fails for some reason, disputes in international commodities contracts of this type are by agreement usually referred to the courts of a small number of jurisdictions, in particular those of England and New York. The establishment of standard arrangements for the resolution of disputes in trans-national e-Science collaborations would greatly reduce the uncertainty surrounding such projects.

The establishment of an Advisory Board on Collaboration Agreement is not going to be an overnight solution to the problems of reducing the legal barriers to effective Internet-based scientific collaboration. Nevertheless, it would represent an important step in the right direction at a time when the natural course of the institutional pressures seems to be in the opposite direction. Once established, it could serve as kind of public conscience, pushing universities and funding institutions and their lawyers to back the commitment that collaborative science is an essential part of the future success of scientific advance.

TOWARDS INSTITUTIONAL INFRASTRUCTURES FOR e-SCIENCE:
The Scope of the Challenge

1. INTRODUCTION – THE OPPORTUNITIES AND CHALLENGES OF e-SCIENCE

1.1 BACKGROUND, MOTIVATION AND PURPOSE

The opportunity exists today for unprecedented connections between scientists, information, data, computational services, and instruments through the Internet. A new generation of information and communication infrastructures, including advanced Internet computing and Grid technologies, is beginning to enable much greater direct and shared access to more widely distributed computing resources than previously has been possible.[1] The term 'e-Science' usually is applied in reference to large scale science that, increasingly, is being carried out through distributed global collaborations enabled by the Internet.[2] Such collaborative scientific enterprises typically require access to very extensive data collections, very large scale computing resources, and high performance visualisation of research data and analysis of results by the individual users. The potential for these advances in technology to support new levels of collaborative activity in scientific and engineering, and ultimately in other domains, is a major driving force behind the U.K.'s Core e-Science Programme.[3]

A growing number of those acquainted with these technological developments anticipate that they will have transformative effects on the organization and conduct of 'knowledge work' – particularly scientific and engineering research. Thus, the recent report by a distinguished advisory panel to the NSF Directorate of Computer and Information System Engineering envisages the enhanced computer and network technologies supporting those connections as forming a vital infrastructure – dubbed the *cyberinfrastructure* – whose

---

[1] 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: 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.]

[2] See Appendix 1, below, on this and related terminology found in the text. 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

[3] In November 2000 Dr John Taylor (the Director General of the Research Councils) announced £98M funding for a new U.K. e-Science programme: £15M was allocated by the Office of Science and Technology (OST) to the Core e-Science Programme, a cross-Council activity to develop and broker generic technology solutions and generic middleware to enable e-Science and form the basis for new commercial e-business software. OST funding for the core e-Science Programme has been augmented by a further £20M from the Department of Trade and Industry (DTI), which is to be matched by £15M from industry. See U.K. Research Council e-Science Programme [available at: http://www.research-councils.ac.uk/escience/ ; also, for allocations to specific science domains, http://www.escience-grid.org.uk/docs/briefing/funding.htm].

effects would be analogous to the historical impacts of super-highways, electric power grids, and other physical infrastructures in raising the productivity of conventional work.[4]

The expectation is that solving *technical* problems associated with advanced 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. In reality, these gains are likely to be the *combined effect of social and technical* transformations. One might be reasonably confident about the pace and scope of future technical advances in computing that will follow from the dynamics of 'Moore's law' and the plummeting price-performance ratio of micro-processors. But, whereas as far greater uncertainties continue to surround the extent to which individuals, groups, organisations, and institutions will be able adapt to and benefit from the novel technological systems that may be engineered on those foundations. Even enthusiastic advocates of heavy investment in hardware, middleware and software components of the coming Grid-enabled cyberinfrastructure have come to recognise that there may be a profound gap between 'raw' performance capabilities (based on bandwidth, storage capacity, processor speed, and interconnection protocols), and its realised performance (based upon the 'usability' properties of the constituent system designs). Some of that awareness has been heightened by systematic evaluations of the pioneering *collaboratory* projects that were mounted in the U.S. during the early 1990's to explore the potentialities of the 'virtual laboratory' concept.[5] That, however, is not the only gap that can significantly limit the transformational potential of 'cyberinfrastructure investments'.

Achieving the aims and aspirations of e-Science is 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 organisations – as challenging as those engineering tasks may be. The informal norms and formal rule structures for collaboration on the ground as well as in cyberspace, that is to say, the 'institutional' contexts within which the work of communities of scientific and technical researchers is carried on also will matter. These, too, will constrain as well as facilitate improvement in the effectiveness of the variety of research collaborations that actually are formed within and across disciplinary, university, and national boundaries.

The institutional and organisational '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 organisations;

---

[4] 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/.] 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/].

[5] See, e.g., Thomas A. Finholt, "Collaboratories as a new form of scientific organization," *Economics of Innovation and New Technology*, 12 (January) 2003: pp. 5-25. Appendix 1 and 2, below, present some information about the characteristics of the pioneer "collaboratories."

(2) the latter organisations' own rules and administrative procedures governing formal relationships with their employed research staff (and research students, 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 recognised 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 extended research projects.

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 organising collaborations in a variety of fields. In Figure 1 (below) the foregoing non-technological elements are depicted, 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.[6]

In shifting the focus of attention from questions of technical engineering to institutional design, is it particularly important to bear in mind that the goals and requirements of the research organisations and host institutions that are likely to emerge as the eventual users of these facilities may well diverge significantly different from those found among the projects which today are pioneering the development of hardware and software systems for e-Science. Some forward-looking exercise of the imagination, therefore, is in order at this time, contemplating the likelihood of e-Science collaborations that will not bear close resemblance to the projects that currently are proceeding under that banner.

To hope to avoid, or even to significantly postpone the effort of critically thinking through the likely needs of projects that have yet to be conceived of, may well prove be a costly strategy. Very substantial resource costs can be entailed when societies try to utilise technological systems the immediate applicability of which turn out to be unexpectedly limited outside the immediate conditions of their genesis. Figure 1. The e-Science Collaboration Domain and Its Infrastructural & Regulatory Supports A concrete, pertinent and not unfamiliar illustration to support that proposition is available in the story of the evolution of the ARPANET into the Internet.[7] The ARPANET and its direct successor, the NSFNet, were communications infrastructures that had been developed under public auspices to serve the needs and circumstances of university-based research groups. This was an environment of application in which individual and organisational behaviours generally are

---

[6] That general perspective informs the approach taken by this report, but is not explicitly elabortaed. For further discussion, see P.A. David, D.Foray and J. Mairesse, 'Public dimensions of theknowledge-driven economy: an analytical framework'. Working Group on the Knowledge Economy, Center for Education Research and Innovation (CERI), OECD, 21 June 2001.

[7] This draws (briefly) on P. A. David, "The Evolving Accidental Information Super-Highway*," Oxford Review of Economic Policy* (Special Issue: "The Economics of the Internet"), 17(2), Fall 2001, pp. 159-187.

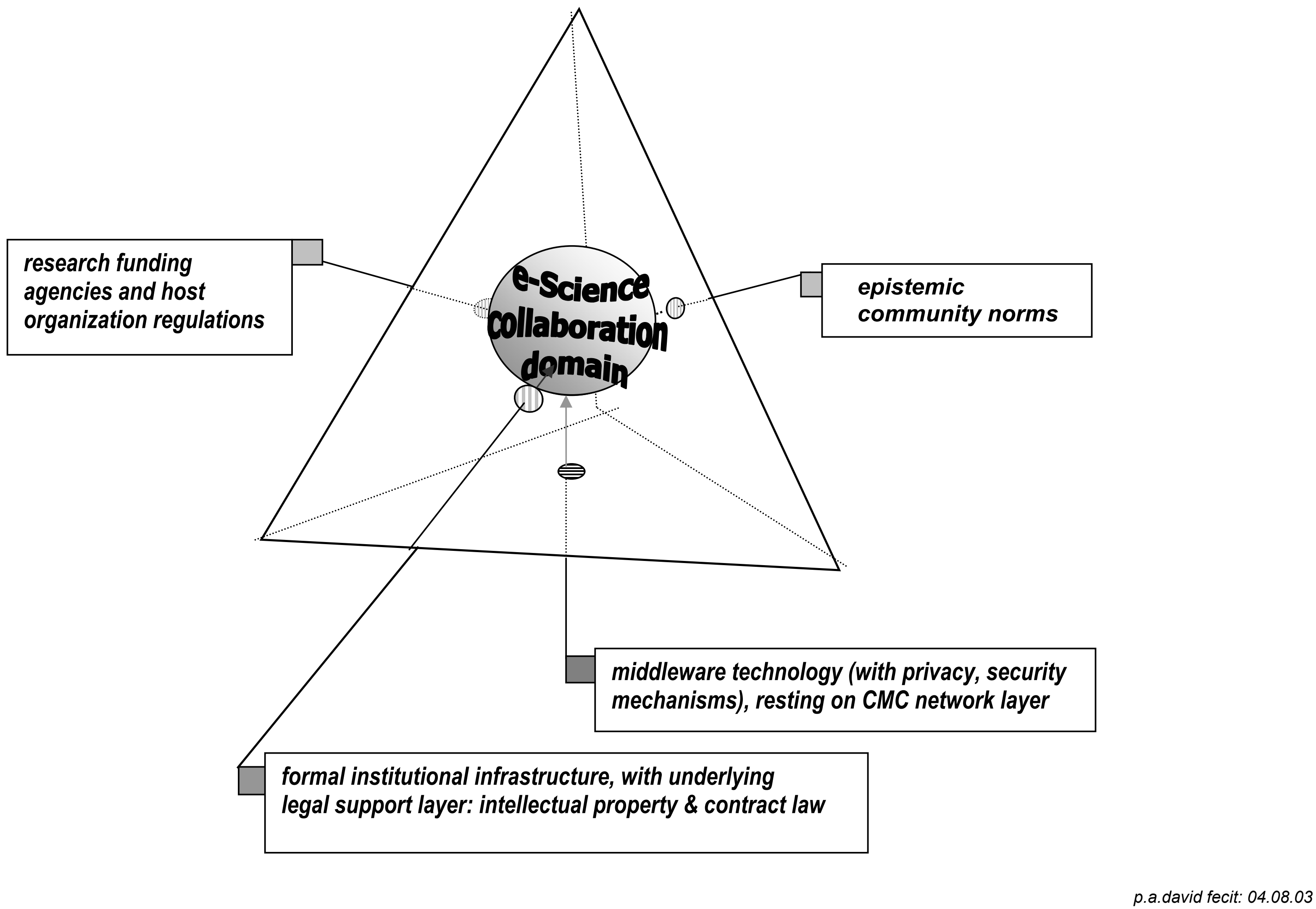
regulated tightly by social norms and institutional rules, and where the dominant ethos is that of co-operation in non-commercially oriented activities of inquiry and information dissemination. The influence exerted over the course of three decades of development work (1964-1994) by the social parameters of that essentially stable background is reflected in the technical specifications of the Internet's end-to-end architecture and the TCP/IP protocol stack. Rather unexpectedly, however, the rapid privatisation and commercialisation of the new, 'connection-less' communications facility that took place during the mid-1990s had the effect of transferring this technological artefact into a social environment that was very different from the one in which it had been designed. The consequences have not been entirely unproblematic, to say the least.

Although the Internet has now begun to be used extensively for commercial purposes, this new context of use is one for which the network of networks has been revealed to be less-than-optimally suited, in more than one respect. This has given rise to an important challenges in areas such as: finding appropriate business models for the Internet's open architecture and culture; adjusting the intellectual property rights (IPR) regime to the new structure of information reproduction and transmission costs; filtering unwanted (spam!) messages; providing security and protection from malicious, or simply non-co-operative, actions by other system users; and designing quality of service (QoS) suitable for network services that was never contemplated in the original system design.

Early experience with the Internet and collaboration-supporting technologies suggests that data and other resource sharing across the institutional boundaries of the resulting 'collaboratories' – which are at the heart of e-Science – already is confronting legal and political administrative constraints, particularly those arising around intellectual property rights issues. Experience in other institutional settings, such as e-government, reinforces these observations.[8] Finding even reasonably satisfactory solutions to such academic and commercial challenges encountered by the first generations of Internet users undoubtedly has created many opportunities for ingenuity to display itself, including some quite profitable new lines of business. Yet, all this adaptive effort, whether successful or not, has entailed considerable unanticipated costs. The open question is whether we are now in a position to make better preparations to utilise the new e-Science tools that it appears to be within our power to fashion.

---

[8] For example, e-Government research underway at the University of Southern California's Information Systems Institute (ISI) has focused on the development of middleware to support information sharing among government agencies. Technical advances have been demonstrable, but entrenched political administrative traditions in U.S. federal agencies, which are quite different from academia, have limited the ISI's ability to work with data of central importance, such as the U.S. Census.


research funding
agencies and host
organization regulations
e-Science
collaboration
domain
epistemic
community norms
middleware technology (with privacy, security
mechanisms), resting on CMC network layer
formal institutional infrastructure, with underlying
legal support layer: intellectual property & contract law
p.a.david fecit: 04.08.03

Motivated by the foregoing contemporary developments, and informed by the historical experience of the Internet's origins in the work of university-based scientists and engineers, this Report has a three-fold purpose:

> First, to articulate the nature and significance of the non-technological issues that will bear on the practical effectiveness of the hardware and software infrastructures that are being created to enable collaborations in *e-Science*;
>
> Second, to succinctly characterise the fundamental sources of the organisational and institutional challenges that need to be addressed in regard to defining terms, rights and responsibilities of the collaborating parties, and to illustrate these by reference to the limited experience gained to date in regard to intellectual property, liability, privacy, and security and competition policy issues affecting scientific research organisations;
>
> Third, to propose approaches for arriving at institutional mechanisms whose establishment would generate workable, specific arrangements facilitating collaboration in e-Science; and, that also might serve to meet similar needs in other spheres such as e-Learning, e-Government, e-Commerce, e-Healthcare.

1.2 ORGANISATION AND THRUST OF THE REPORT: AN OVERVIEW

The main body of the Report is organised in four main parts, and is followed by a collection of appendices containing technical background and illustrative case material. The following sections of Part 1 address the first of the three principal tasks that have just been described, beginning with an examination of the technological and institutional contexts of e-Science and their interrelated dynamic evolution (in section 1.3), and then reviewing the different ways in which these two infrastructure components affect the costs of collaboration (in section 1.4).

The second principal task is the subject of Part 2, which opens by considering the social conventions and norms that may be said to govern collaboration within scientific collaborative workgroups, distinguishing these from the agreements governing the contractual relations among the institutions in which the members of those work-group are employed. Section 2.1 takes notice of the additional elements complexities that are presented by recent changes in academic communities' professional norms and in the career conditions affecting university researchers. Still further complications in the relationship between informal and formal governance mechanisms are seen to arise as a result of the ways in which uncodified normative structures governing professional conduct among collaborating individuals may be affected by the introduction of explicit legal contracts with their employing institutions, as well as with their academic peers.

Section 2.2 takes two necessary analytical steps towards understanding the character of the challenges involved in finding or designing anew concrete institutional arrangements that facilitate the formation and conduct of socially productive research collaborations in e-Science. Those challenges are rendered more formidable by the need to provide for multi-institutional collaborations including international partners, as well as to accommodate trans-disciplinary projects involving distinct research units (departments, laboratories, institutes or centres) within the same Public Research Organisation (PRO).

The first step is to delineate (in sub-section 2.2.1) the variety of distinct benefits offered by co-operative organisation of research, and the multiple sources of conflict that are likely to exist among the interests of the potential collaborators, as well as among the administrative entities and employing institutions. The second step (in sub-section 2.2.2) draws upon recent economic analysis to suggest reasons why the social efficiency of publicly funded research investments is likely to be greater under a regime of more liberal contractual arrangements among the individual collaborating researchers; and, particularly, under rules that provide 'weak' rather than 'strong' protection for the commercial exploitation of intellectual property rights held individually by the participating institutions. From this analysis there follows an important meta-principle: the appropriate approaches to the institutional design problem for publicly funded collaborative e-Science are those that would be especially responsive at the margin to satisfying the collaborating agents' organisational requirements for conducting the research in question. Correspondingly, they will be less disposed to accommodate other incompatible corporate interests and goals that their respective administrative units and employing institutions may seek to attain through participation in the proposed collaboration.

Section 2.3 sets out three principal aspects of the complicated 'legal relationship' issues that can arise of among collaborating universities and other corporate bodies that accept contractual responsibilities to one another, as well as to external funding agencies for the performance of scientific and technical research. It then reviews in some detail the many theoretical and practical difficulties that are encountered when one turns to existing legal doctrine to provide a basis for common rules that could govern those relationships. Three specific classes of issues are examined *ad seriatim* (by sub-sections 2.3.1 through 2.3.3, respectively):

(1) issues in regard to material contributed to collaborations (including those involving exposure of legal risks, violations of competition law, and individual contributors' claims to the fruits of the collaboration;
(2) issues surrounding material arising from collaborations, which are primarily those concerned with assignment, ownership, and licensing of intellectual property – particularly the statutory provisions and judicial interpretations of the U.K. Copyright and Database Regulations, and some salient of difference in European and U.S. intellectual property law;
(3) issues of liability arising from collaborations – primarily those of the assignment of responsibilities for injury to third parties and losses of professional/institutional reputation due to incompetent or unethical conduct in performance of the research.

The third of the Report's major tasks is taken up in Part 3. An initial assessment is made (in section 3.1) of some oft-recommended approaches to simplifying institutional mechanism design problems in relation to the law, notably by introducing standard form agreements, and by harmonising disparate and potentially conflicting legal doctrines, and statutes. Finding many practical deficiencies in the latter approaches, the argument (in section 3.2) favours developing an alternative, more flexible, modular process to generate contextually appropriate contractual arrangements for collaborative research projects. The proposed process calls for the principal public funding bodies engaged in building the technological components of the e-Science infrastructure also to lead their authority to the work of a new "public actor" in the shape of an independent advisory body on institutional infrastructures. The role of that body, referred to as the Advisory Board on Collaboration Agreements (ABCA) in e-Science, would be to guide and co-ordinate the formulation and dissemination of an array of specific contractual clauses that could be used to construct a variety of legal agreements governing scientific and technological research collaborations among universities and other corporate partners. The Board also could develop sufficient expertise to provide guidance for research groups seeking effective informal arrangements to deal with various internal governance issues, thereby facilitating the more spontaneous, "bottom up" formation of projects enabled by the emerging e-Science collaboration tools.

The remaining sections of Part 3 set out a number of requirements that should be met by the constitution of such an advisory body, in terms of the private expertise and public agency experience and interests upon which it would need to be able to draw (section 3.3), and the development of an evolving informational base about the actual collaboration arrangements and their efficacy (section 3.4).

In the fourth and concluding Part of the Report two different questions are treated. They are related, however, inasmuch as each has a concrete bearing upon the practical implementation of the general approach, and the specific recommendations advanced by Part 3. The first of the pair concerns the nature of the broad principles that the proposed ABCA would embrace and seek to embody in a menu of contractual clauses; whereas the second addresses the need to find an expedient "way forward" that would provide near-term guidance for issues of governance arising from the e-Science projects that are presently underway.

Part 3's procedural recommendations for 'contractually constructing" arrangements to support a variety of e-Science collaborations (intentionally) are formulated in a way that remains neutral with respect to the general thrust of the guiding principles that the ABCA would embrace. Part 4, by contrast, takes notice of the growing number of calls for modifications in the intellectual property rights policies of governments – particularly in order to assure more protection of the public domain in scientific and technical data and information. Policy statements in that vein have emerged recently from a number of influential bodies in the United Kingdom, the European Union and the United States. It would be remarkable were the Advisory Board not to give weight to these concerns in delineating the principles against which it should assess proposed model contractual clauses for use in e-Science collaboration agreements.

Indeed, it is a positive advantage of the approach based upon contractual agreements that it would allow a direct and positive response to the worries expressed about excessive restraints being imposed upon open scientific collaboration by excessive recourse to intellectual property rights (IPR) protections. In other words, there is a case (developed in section 4.1) for using the establishment of the ABCA as a means to avoid having to wait for statutory reforms in the IPR regime. As an illustrative case in point, it is shown how public licensing of intellectual property under a standard form of “free and open source software”(F/OSS) license–such as the GNU General Public License–may be used in conjunction with private contractual provisions governing commercial exploitation of the “open source” code. This “dual licensing” approach can accomplish two purposes that often are presented as inimical from a practical standpoint. On the one hand, public funding authorities may regard it as efficient to maintain the essential functional features of a ‘knowledge commons’ in the software tools and products whose creation they have sponsored, but, at the same time, it may be desirable to leave some scope for market incentives to mobilize complementary private sector investment directed toward further development of basic software innovations released under the terms of F/OSS licenses.

The solution suggested is to allow “dual licensing” of some categories of publicly funded software (and middleware), combining GNU GPL licenses with the option of constructing contractual arrangements (built from standard clauses) whereby third parties obtain the copyright holder’s permission to develop modifications and extensions for private commercial distribution. Under this approach, there would be a clearly identifiable need for the services of an advisory board-like entity to develop appropriate contractual clauses that would work in conjunction with copyright licenses that embody the so-called “copyleft’ principle. To initiate effective implementation of the recommended contractual customisation approach, and its application in the context of ‘dual licensing’ of F/OSS products, will require a strongly supportive stance on the part of the national and international public agencies and private foundations that provide major funding for the e-Science programs and projects conducted in PROs.

The discussion of practical measures closes (in section 4.2) by outlining an interim course of action for the e-Science Core Program to follow, in order to furnish itself with expert advice and counsel for the decisions that must be made about non-technological governance issues affecting the use of the software systems whose creation it has sponsored, and expects to sponsor in the foreseeable future. This “way forward” could have a potent impact, not only in shaping the near-term institutional environment for e-Science in the U.K., but by initiating the first steps on a transition path towards the eventual institutionalisation of an independent ABCA along the lines envisaged here. By moving quickly to establish an interim Working Party on institutional infrastructures for e-Science, and having that body actively engage with representatives from international counterpart programs, the U.K.’s Core Program soon could begin exerting significant international influence. It would thus move closer to fulfilling the promise of e-Science to accelerate advancement of knowledge and material well-being on a global scale.

1.3 TECHNOLOGICAL AND INSTITUTIONAL CONTEXTS OF e-SCIENCE

e-Science is a term used increasingly widely as a generic label for all scientific and technical research activities conducted on the Internet.[9] But it is employed more specifically here, in referring to scientific activities supported by high bandwidth computer-mediated telecommunications networks, and particularly to encompass the variety of such digital information-processing applications that are expected to be enabled by the *Grid.* The latter may be viewed as the general purpose network technology which will serve to facilitate new, computationally intensive forms of scientific inquiry: desktop supercomputing, distributed supercomputing (a marriage of parallel and distributed computing), extensive exploration of linked distributed dynamic databases by high-speed search engines, and collaborative environments (collaboratories or virtual laboratories) including smart instruments for data capture and analysis that are coupled to supercomputing resources, and so on.

Collaborative e-Science is the aspect of the vision of 21st century science that holds out the most exciting new possibilities, and which also poses the most demanding challenges at the technical, social and legal levels. Technological and social changes are inter-twinned, and in many respects their interactions and mutual adaptations are difficult to disentangle. It is undoubtedly the case that technological advances that have placed new, more productive and more costly facilities and instruments at the disposal of researchers are prominent among the forces driving the widely observed trend towards collaborative organisation of scientific inquiries. But, beyond the increasing scale of projects utilising "lumpy" capital-intensive facilities in fields such as physics and astronomy, the sheer increase in the amount of pertinent information, and the progressively more specialised knowledge and expertise that must be brought to bear in order to conduce fruitful research programs in most branches of science, have contributed to the growth in the size of teams and the numbers of co-authors on scientific, technical and scholarly publications.[10] Although the continuous pressure toward specialisation and division of labour has pushed researchers into the forms of cooperative knowledge transactions entailed in collaborative inquiry, more recently researchers located at widely dispersed institutions have been drawn into informal and formal collaborations by the dramatic advances that have been achieved in computer-mediated telecommunications.[11]

---

[9] See Appendix 1 for a glossary of descriptive terms in the text, including: e-Science, the Grid, Collaboratories, Virtual Laboratories, Cyberinfrastructure, and their relationship to one another.

[10] Although the emergence of research collaboration in 'Big Science' fields was viewed from the 1960s onwards as a significant novelty reflecting underlying tendencies in the organisational structure of modern science, the increasing generality of collaborative organization is now attracting fresh interest as the most recent phase in a broader, longer and more continuous development. See, e.g., J. S. Katz, "Geographical proximity and scientific collaboration," *Scientometrics*, 31(1), 1994 : pp. 31-43 ; D. Hicks and S. J. Katz, "Science policy for a highly collaborative science system," *Science and Public Policy*, 23(1), 1996 : pp. 39-44; J. S. Katz and B. R. Martin , "What is research collaboration?," *Research Policy*, 26(1). 1997: pp.1-18) ; H. Etzkowitz and C. Kemelgor, "The Role of Research Centres in the Collectivization of Academic Science," *Minerva*, 36, 1998: pp.271-288.

[11] See Appendix 1: Figs. 1-3. This is reflected in the rising frequency of inter-institutional collaborative publications among U.S. university researchers in scientific domains where average team size is comparatively small (e.g., mathematics, and economics); and also by the observation that the growth of inter-institutional collaborative publications involving U.S. academic researchers in international teams has outpaced that of purely domestic inter-institutional collaborative publications, as well as the rate of growth in average team sizes. See James D. Adams *et al*., "Patterns of Research Collaboration in U.S. Universities, 1981-1999," Economics Department Working Paper, University of Florida, Gainesville FLA, March 2002.

However difficult it may be to empirically identify the separate influences of the technological from those of the other, social factors affecting collaborative research, it is not only possible but also important to draw this conceptual distinction. This report goes further, however, by directing special attention to the legal-institutional contexts of the array of collaborative research activities that are expected to be greatly facilitated by improvements of the technological components of the e-Science infrastructure. As a background for the discussion, Appendix 1 of the report presents a taxonomic framework for e-Science collaborations that is meant to highlight the various classes of interactions among collaborating parties that these technical facilities can support. This framework classifies collaborations on the basis of their major purpose, rather by reference to the particular digital information tools and services they might employ. Our taxonomy distinguishes among virtual laboratory activities conducted via the (enhanced) Internet that are pre-dominantly:

1) "community-centric" – aiming to bring researchers together either for synchronous or asynchronous information exchanges;
2) "data-centric" – providing accessible stores of data captured or extracted from remote sources, and creating new information by editing and annotating them;
3) "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;
4) "interaction-centric – enabling applications that involve real-time interactions among two or more participants, for decision-making, visualisation or continuous control of instruments.

When this scheme is applied to classify the array of Pilot Projects that have been funded under the e-Science Core Programme in the United Kingdom, the data-centric branch of the taxonomic tree emerges as far and away the most densely populated.[12] The situation contrasts with the more uniform distribution that emerges from a comparable classification of much small number of pioneer collaboratory projects that were organised under public funding programs in the U.S. during the late 1980's and early 1990's (see Appendix 1: Fig. 4). That difference reflects in part the focus of the e-Science program on the creation of middle-ware platforms and tools, and in part the greater centrality of the roles that digital databases have more recent come to occupy in the work of science and engineering communities. Yet, a suspicion remains that some influence on the profile of these the Pilot Project sample also has been exerted by consideration of the greater administrative complexities that would have to be overcome to organise more thoroughly interactive modes of collaboration among research groups situated at various institutions within the U.K.

The institutional infrastructure for e-Science collaborations might be viewed by some to be the soft part of what the report of the recent NSF Blue Ribbon Advisory Panel (February 2003) refers to as the "Cyberinfrastructure" that promises to revolutionise science

[12] Another use of the taxonomy has been to assist in identifying a subset among the Pilot Projects that contained representatives of each of the 'collaboration purposes', and whose activities could be studied more closely in order to understand the variety of e-Science research contexts for which supportive institutional arrangements would need to be constructed.

and engineering.[13] But, in truth, its design, construction, maintenance and updating pose many challenges that are at once more delicate and harder than the technical engineering feats required for reliable and secure Grid-enabled computing. Institutions simultaneously are run by and govern human agents, and, for that reason as well as others, they are considerably less plastic than most machine organisations – i.e., systems composed of technological artefacts. Often, when they function well, institutions and the behavioural norms they reinforce become unobtrusive and tend to disappear into the background, so that the question of whether they will require modification to continue functioning smoothly in new environments is often deferred until after those environments have materialised.

Scientific teams engaged in hardware and software engineering in order to forge the tools needed to support their own work are, as a rule, more than fully tasked. They seldom are able to focus concurrently on the issues of how social and technological mechanisms can best be combined to address the array of complex problems that other users of those tools eventually would need to solve before the potentialities e-Science can be fully realised. Nor should working scientists be expected to possess the necessary expertise to consider the problems of developing procedural norms and formal contractual arrangements governing collaborative contributions of research resources.[14]

1.4 TOWARDS ENVISAGING THE CYBERINFRASTRUCTURE WITH COLLABORATION COSTS

The functional domain of institutional arrangements supporting scientific collaboration is 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 organisations all of which are

[13] See Atkins et al., *Revolutionizing science and engineering through cyberinfrastructure*: (2003). [Available at: http://www.communitytechnology.org/nsf_ci_report/ .]

[14] Lest there be any doubt on this score, it should be emphasised that the limited attention accorded to institutional design by scientists and technologists is a consequence of specialisation, and therefore the comment here is not meant as a reproach. As Bertrand Russell said, "We forgive specialists, because they do good work."

not situated within and hence under the governance of a single legal jurisdiction and political authority.

It is 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: the 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. Yet, 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.

The lattermost of these requirements may be seen to be present in the very idea of an e-Science ‘infrastructure’ as that is now coming to be conceptualised. The recent report of the NSF Blue Ribbon Advisory Panel on Cyberinfrastructure (2003:pp. A-1-2) describes the latter concept in expansive terms, in which the activities of human agents and organizations also are subsumed under the heading of infrastructure. According to the Panel (2003: p. A-3), whereas, historically “infrastructure” has been viewed by people in the computer and telecommunication engineering sciences “largely as raw resources like compute cycles or communication bandwidth,” now it is critical to think of [cyber-] “infrastructure” as having three rather different basic components:

- “**Technological artefacts**. These human-constructed artifacts include facilities (computers, mass storage, networks, etc.) and software. These artifacts sometimes provide services, and sometimes they are simply available to be ‘designed into’ applications.
- “**Technological services.** Various capabilities are provided as services available over the network rather than as software artifacts to be deployed and operated locally to the end-user.
- “**Services from people and organizations**. These include everybody who is providing a shared pool of expertise leveraged by the entire scientific and engineering research community to develop and operate the technological artifacts and provide advice and assistance to end-users making use of them.”

Given the inclusion of the lattermost among these, it is rather remarkable that nothing in the report of this NSF Panel addresses the nature of the institutional settings, the incentive mechanisms, and the organizational culture of those who are “providing a shared pool of expertise leveraged by the entire scientific and engineering research community." Remarkable as that omission is when viewed from a systems design perspective, perhaps it is readily understandable as a rhetorical strategy: to focus upon the difficult and all too familiar questions posed by the human organisation components of the system undoubtedly takes something away from the construction of a persuasively enthusiastic case for devoting a very substantial amount of funding to its technological elements. The principal problem with this, however, is that to say "Well, we can always jump off that bridge when we come to it," and then to hasten onwards, is more often than not a self-fulfilling strategy.

2. THE INSTITUTIONAL CONTEXT OF SCIENTIFIC COLLABORATION

This report makes recommendations about how appropriate institutional arrangements, legal contractual arrangements in particular, might be established for collaborative e-Science. But, in constructing and seeking to implement such arrangements it is important to understand both the informal institutional conditions under which collaborative scientific research projects are organised and conducted, and the specific character of the legal issues that will arise in the organisation of e-Science projects. A few, key aspects of the institutional settings in which arrangements for collaborative e-Science projects are required can have a major effect in determining the success of such undertakings. As these have strong implications for the approach that underlies the recommendations put forward in Part 3, it is important to lay a basis for the latter by reviewing these features of the current "institutional environment."

2.1 COMPLEXITIES OF THE CURRENT INSTITUTIONAL ENVIRONMENT

A first and quite important point to notice about academic communities today is that they have been undergoing rapid changes that have left the norms of professional behaviour far from uniform, and in a state of flux. In particular, the scientific communities traditionally had similar (albeit differentiated) norms for the attribution of credit and responsibility for collaborative research. These traditional norms fostered the dissemination of scientific information because the primary incentive for individual researchers (status and recognition within the scientific community), constituted incentive for them to disseminate widely and to accept responsibility for research results.[15] A relatively secure career structure also meant that researchers had little incentive to distort or to falsify results because the risk to reputation outweighed potential short-run gains. Yet, the increasing uncertainty of scientific careers has led to more disputes about the attribution of credit and responsibility for research findings. Major journals such as the *Journal of the American Medical Association* and

[15] For an account of the classic sociological treatment of the norms of academic science communities, and the modern economic analysis in information-theoretic terms of the functioning of the resulting behavioural incentives and constraints, see P. Dasgupta and P. A. David, "Towards a New Economics of Science," *Research Policy*, 23 (1-2), 1994: pp. 487-521; also, 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: A Primer," forthcoming in National Research Council, *The Role of the Public Domain in Scientific Data and Information*, Washington, D.C.: National Academy Press, 2003. [SIEPR Policy paper No. 02-030, Stanford University, March 2003, available at: http://siepr.stanford.edu/papers/pdf/02-30.html .]

organisations such as the *International Committee of Medical Journal Editors* now have explicit policies regarding the attribution of authorship and responsibility in contexts of collaborative work.

Additionally, universities and other public research organisations (PROs) have responded to government policies urging them to cooperate with business firms sponsoring university-based R&D projects, as well as to make efforts of their own to capture value from academic research results. The resulting pursuit of intellectual property rights, and their exploitation through licensing or the creation of university-owned 'start-up' enterprises, has worked to undercut the traditional incentives to rapidly and fully disseminate research findings and methods.[16] This weakening of traditional "open science" incentives to claim priority of discovery (rather than securing property rights) has been accompanied in some quarters by the erosion of older normative structures. There is considerable heterogeneity of belief among some communities, most evident among the life sciences, as to whether or not the prime obligation of academic scientists remains co-operation for the advancement of knowledge, or the pursuit of research geared toward profitable commercial innovations – including those from which they can expect to benefit personally.[17]

These new trends have not, however, been advancing with uniform strength across all areas of scientific endeavour, institutions or geographical regions. They are, for example, quite evidently far more pronounced in the life sciences, and particularly so among departments engaged in molecular biology and genetics than among departments of theoretical and experimental particle physics. The uncertainty this creates would appear to imply that formal legal rules allocating responsibility for, and the outcomes of, collaborative projects are more important now than ever. Certainly academic researchers are increasingly aware that the law has the power to impinge upon their work, and many are sensitive to the existence of a disjuncture between the norms upon which the law operates and those that have traditionally governed collaborative science.

---

[16] Typically, this has involved a focus on licensing the use of trade-marks and logos, and university-owned patents, copyrights (and more lately database rights), as well as arrangements assigning patents to start-up ventures in which the university takes an equity position. But, more recently British universities are being encouraged by Government funding and policy initiatives to develop a wider array of so-called "Third Stream" activities – i.e., those involving the commercial provision of knowledge products and services (other than teaching and research within their individual institutional purview). Moreover, in some Government circles it is viewed as not only appropriate but imperative that universities have long-term strategies for developing and managing their "knowledge assets" so as be better able to engage in "Third Stream" revenue generating activities. See Jordi Molas-Galant et al*., Measuring Third Stream Activities: A Report to the Russell Group of Universities*, SPRU, University of Sussex, April 2002.

[17] See, e.g., Jason Owen-Smith and Walter W. Powell, "Careers and Contradictions: Faculty Responses to the Transformation of Knowledge and its Uses in the Life Sciences," in *Research in the Sociology of Work* (Special Issue on *The Transformation of Work*, edited by Steven Vallas),10, 2001: pp. 109-140; Rebecca S. Eisenberg, "Bargaining over the transfer of proprietary research tools: Is this market failing or emerging," Ch. 9 in *Expanding the Boundaries of Intellectual Property*, Eds., R. Dreyfuss, D. L. Zimmerman and H. First. New York: Oxford University Press, 2001; J. P. Walsh, A. Arora and W. M. Cohen, 2002. "Research Tool Patenting and Licensing and Biomedical Innovation, December. in *The Operation and Effects of the Patent System,* [Report of the STEP Board of the National Research Council, National Academy of Sciences], forthcoming from National Academy Press in 2003.

2.1.1 The Balance between Informal and Formal Governance Mechanisms

Individual scientists, however, have varying degrees of commitment to the traditional norms and in any case lack an obvious forum in which to express dissatisfaction with the law. Moreover, they differ in their interest and talent for handling the administrative aspects of scientific projects. Such differences notwithstanding, most working scientists tend to express impatience with, if not disdain for the effort to formal rules, norms and standards of individual conduct among researchers – even if they acknowledge that this may be necessary to create an appropriate institutional context for the conduct of a new collaborative undertaking.

These attitudes reflect in some part the shared expectation that relationships among scientific peers and co-workers can be governed by the incentive compatibility of co-operative consultative processes (where the important games are strongly 'positive sum,' and there will be a potential for significant damage to individual's reputation if they are seen to have defected from cooperative play). Impatience with efforts to articulate norms also may stem, in some part, from the supposition that the asymmetry of power relationships within scientific workgroups is well recognised by all the participants; and is understood by them to provide a reasonable enforcement mechanism to resolve the normal run of internal disagreements – for example, by appeal to the authority of the project director. Nevertheless, the introduction of new actors (either in the form of another, collaborating workgroups, or administrative representatives of host institutions) readily can de-stabilise those internal governance mechanisms.

Proposals for new collaborative arrangements, however, introduce the possibility of new incentive structures, which also may be inherently de-stabilising. Excessive interest in the details of the administrative arrangements might not only be considered a diversion from the core scientific activity, but a threat to the trust upon which scientific collaboration depends. It is important for the success of a collaborative project that the scientists involved understand the broad terms upon which it is to be conducted. At the same time, exposure to a lawyer's attempt to anticipate all the conceivable potential situations leading to collaborative failure or other unwanted outcomes, and to provide for remedies and mitigating procedures *ex-ante*, may negatively impact the collaborative spirit of the research partners.

Thus, while it surely is salutary to clarify ambiguous social norms, and to reinforce certain professional standards of conduct in situations where these are found to have undergone unwanted erosion, it could be quite counterproductive to attempt to devise complete and detailed contract-like regulations for the internal governance of relations among academic researchers. Such 'codes' may have the perverse effect of inducing researchers to think narrowly in terms of legal rights and obligations, and to resort to the often cumbersome machinery of the law in order to resolve disputes among colleagues, or conflicts with university administrators. For example, in the face of unforeseen contingencies, participants in a collaborative project may stand by the letter of the contract rather than co-operate towards averting escalation of the conflict into an organisational crisis. Equally, conscious of having the option of recourse to legal means of protecting their interests, they may be less concerned to look for ways to remove the source of the problem –

say, by rearranging features of the project's work programme. Explicit preparations to deal with 'collaboration failure' by mobilising external (legal) resources, can in this way 'crowd out' individuals' voluntary actions that would render the collective effort more successful. Researchers who feel that no or little trust is placed in them may be more likely to behave in ways that are inimical to the success of the collaboration.[18] Evidence from experimental economics and field studies shows that the introduction of explicit contractual incentives can weaken or entirely vitiate the effects whatever intrinsic impulses or social motivations might otherwise be sufficient to elicit cooperative behaviour on the part of the actors.[19]

The same principle also has a bearing upon the approach to inter-institutional contracting where legal agreements are required. In designing an institutional framework for e-Science collaboratories, a light touch approach may be required to prevent all remnants of the 'open science' ethos from being 'crowded out' even from the transactions among academic institutions.[20] Certainly, when considering the respective roles played by formal contractual agreements and informal norms and understandings regulating the interactions among members of research communities, one should not suppose that these are perfectly complementary, or even strictly "additive." At the margin, each may provide a substitute for the other. The problem is that one rarely can know, *a priori*, how deep that margin is, and whether the introduction of a requirement to enter into formal legal contracts may in effect displace, or degrade the effectiveness of informal governance mechanisms. This is an observation that can be formulated in rather general terms, and, indeed, economic analysts frequently made the point that external regulatory provisions mandated by government may 'crowd out' the provision of less formal governance arrangements among the agents involved.[21]

[18] See, e.g., M. Bacharach, G. Guerra and D.J. Zizzo, 2001. "Is Trust Self-Fulfilling? An Experimental Study", Oxford University Department of Economics Discussion Paper n. 76. [Available at: http://www.econ.ox.ac.uk/Research/WP/PDF/paper076.pdf.], and Guerra, G. and D.J. Zizzo, 2003. "Trust Responsiveness and Beliefs", *Journal of Economic Behavior and Organization*. Discussion paper version available at http://www.econ.ox.ac.uk/Research/WP/PDF/paper099.pdf (in press, forthcoming in 2003).

[19] See, e.g., S. S. Frey and R. Jegen, 2001. "Motivational Interactions: Effects on Behavior', *Annales d'Economie et de Statistique* 63-64, pp. 131-153); U. Gneezy and A. Rustichini, 2000. "Pay Enough or Don't Pay at All", *Quarterly Journal of Economics*, 115, pp. 791-810; .and Deci, E. L., R. Koestner and R. M. Ryan, 1999. "A Meta-Analytic Review of Experiments Examining the Effects of Extrinsic Rewards on Intrinsic Motivation", *Psychological Bulletin* 125, pp. 627-668. For directing U.S. to this pertinent literature, the writers are indebted to Daniel Zizzo, of Christ Church College, Oxford.

[20] Further research undoubtedly would be required to assess how light this 'light touch' would have to be in various situation, in order to minimise the displacement of informal understandings in contexts where those are more supportive of fruitful on-going inter-institutional co-operation.

[21] A very simple illustration of this very general point arises in discussions of the "moral hazard" problem that is alleged to have been created by government programs that insured depositors in mutual savings banks in the U.S.; lulled to a sense of security that they would be protected from losses due to the bank's inability to pay, the depositor-shareholders paid little attention, and so left unchecked the unsound loans and fraudulent transactions that were made by the executives who managed the affairs of many of those federally chartered institutions. A rather more subtle point to notice is that the introduction of formal prescriptive arrangements governing behaviour – such as contracts – in circumstances where there is uncertainty, and consequently less than complete information, means that such contracts necessary will be incomplete. At best, they can specify *ex ante* features of the process, or procedure that the parties are to follow in coping with unanticipated events affecting their enterprise.

The implications of the foregoing are quite straight-forward. Firstly, the parties seeking to establish a mutually beneficial collaborative research project have both incentives and capabilities to start the process on their own, in a "bottom up" fashion. Beginning without anything like a complete and explicit set of governance arrangements provided by legal contracts (which would carry external "third party," enforcement provisions), they undoubtedly will quickly enter into some informal discussions on key issues: the division of research responsibilities, the arrangements for access to data-streams while the research is in progress, and afterwards; also high on the agenda for discussion will be the project's publication plans, and the general 'collective policy' vis-à-vis intellectual property rights claims to such results as may be anticipated.[22]

Secondly, there is reason to expect that there will be latent or manifest grounds for members of such groups to devote time and effort to activities that are likely to reinforce cooperative attitudes and behaviours among the participants. Those efforts include an array of 'natural' social contacts (dyadic transactions as well as collective assemblies) which facilitate monitoring of the personal dispositions and social attitudes of colleagues, and can contribute to raising levels of trust and trustworthiness. Engaging in non-committal speculative discussions that explore the possibilities of successive projects of potential mutual benefit, the formation of which would be contingent on the successful outcome of the immediate prospect, would have a similar function. They serve to embed what otherwise might be construed as a 'one-time transaction' in a 'super-game' that features repeated play. The pay-off structure of the latter form of game tends to induce (rational) participants to defer defections from co-operation, even if it remains unwarranted to assume that acts of self-interested opportunism at the expense of the rest of the group will therefore have been foresworn by every one of the players.

Therefore, it is not unreasonable for scientists to be disposed to avoid, to the utmost extent possible, both efforts to codify administrative rules for research management, and the framing of legal contracts for governance of collaborative projects. In addition to the analytical considerations already reviewed, there is empirical experience in the field of contract law that even in business affairs many parties are reluctant to use the law in the planning of their relationships, because they fear that it will harm the collaborative nature of purely commercial relationships.

The upshot is that the desire of scientists not to become embroiled in such administrative and legal matters ought to be respected in determining appropriate governance

[22] See, for example, S. Hilgartner,. and S. I. Brandt-Rauf, "Data Access, Ownership and Control: Toward Empirical Studies of Access Practices," Knowledge, 15, 1994.: pp. 355-372; S. Hilgartner, "Access to Data and Intellectual Property: Scientific Exchange in Genome Research," *in Intellectual Property Rights and the Dissemination of Research Tools in Molecular Biology.* Summary of a Workshop Held at the National Academy of Science, February 15-16, 1996. Washington, D.C.: National Research Council, NAS, 1997: pp. 28-39; S. Hilgartner, "Data Access Policy in Genome Research," in *Private Science*, A. Thakray, ed., Oxford: Oxford University Press, 1998. on the arrangements for access and control of data-streams that emerged among the teams participating in the Human Genome Project. These internal data-sharing agreements may be contrasted with the observations of the weakening in recent years of the ethos of general data-sharing in the biomedical sciences (noted above, in section 2.1).

structures for collaborative projects, even when these require the provision of legal contracts. From this position it follows that what scientists would find most helpful in pursuing research by means of multi-party collaborations is a menu from which to select ‘ready-made’ solutions to the more commonly occurring specific problems in such enterprises; and the option to reconfigure the elements of those solutions to fit the idiosyncratic requirements of their particular circumstances.

The indicated solutions can be of two sorts. For intra-project relationships among employees of a given administrative unit (which includes an entire hierarchically administered institution such as a research institute or a university), the menu should emphasize reliance upon informal, peer-enforced norms of conduct, and alternative procedures for dispute resolution. But, the governance arrangements pertaining to research relationships that involve collaborations across institutional boundaries, whether with other PROs or with business firms, will require legal contracts; the menu of alternative contractual clauses in that case must feature the array of provisions from which a comprehensive agreement can be constructed.

2.1.2 Conflicting Interests in Institutional Collaborations and Partnerships

The principal *legal* actors in the establishment of a collaborative research project generally are not the researchers involved, but, as has been noted, the institutions by which they are employed. This is both appropriate and presents real dangers. On the one hand, it is appropriate that scientists should be relieved of the burden of establishing and maintaining the infrastructures of collaborative work. On the other hand, there is a danger that the farther the creation and maintenance of such structures is placed from the activities of collaborative science itself, the greater is the likelihood that those structures will reflect the interests of actors other than the collaborating scientists, and worse, actors whose goals may be inimical to the effective conduct of the project.

Universities do have multiple and conflicting institutional interests in relation to collaborative scientific research. They may find it in their interest to foster collaborative science as a means of supplementing their intellectual capabilities, enhancing the institutional research reputation of their schools and departments, and attracting ‘star’ researchers and research funding – both in the near term and in the longer run. In some areas, such as the European Union Framework Programmes, contractual participation in multi-institutional networks is the *sine qua non* for obtaining external funding, and, increasingly in recent years public research agencies and private foundations have encouraged the organising of collaborative projects of that kind.[23]

---

[23] For discussion and analysis of endogenous coalition formation in response to such programs, see P. A. David and L. C. Keely, “The Economics of Scientific Research Coalitions: Collaborative Network Formation in the Presence of Multiple Funding Agencies” in *Science and Innovation: Rethinking the Rationales for Funding and Governance*, A. Geuna, A.Salter and W. E. Steinmueller, eds., Cheltenham, Eng.: E. Elgar, 2003.

At the same time, however, universities also have acquired a stake in the promotion of some government policies that may well turn out to impede collaborative research in science and technology. Universities in the United Kingdom, for example, have adopted increasingly comprehensive policies of asserting ownership of any intellectual property in material produced by their research and teaching staffs. In some instances it is true that clear policies regarding the ownership of intellectual property may facilitate the resolution of disputes between collaborators. In other circumstances, however, the potential conflicts of the IPR claims asserted by the different institutions that employ the would-be collaborators (i.e., the research scientists) can create impediments to the formation of scientifically promising projects. Indeed, the prospective transactions costs incurred in trying to resolve what are in essence distributive conflicts among the prospective participants may even frustrate formation of research undertakings that are likely to yield high rates of return to the coalition as a whole.

Universities are particularly complex organisations, however, and their key administrative leaders typically are well aware of the multiplicity of distinct missions that society expects them to pursue concurrently, so that the balance of priorities among those missions often is less than clearly defined within each institution.[24] In some respects this degree of ambiguity is a source of flexibility and accommodation to special circumstances, but it also may permit formal contractual arrangements for co-operative research among them to becoming 'snagged' on points of importance to one or another of the participants – even when those matters have little to do with distributive conflicts, and consequently are difficult to resolve by arranging inter-institutional 'side-payments'.

Quite understandably, legal departments operating within universities must maintain a strong professional commitment to protecting the institution by limiting its exposure to the particular risks of collaboration, as well to the losses that may ensue from the conduct of research by their own employees and contractors. Sometimes, therefore, foreseeable and uninsurable risks the individual institutions would have to bear will appear too large in comparison to the uncertain benefits they might derive, and so it will turn out to be far easier to negotiate 'safer' collaborative projects, or ones with more readily calculable future income-streams. Furthermore, in response to the thrust of recent government policies

---

[24] In Britain, the Russell Group of Universities has recognised this in a recent expression of concern that Government core funding for universities might be associated with easy-to-measure features of entrepreneurship, technology transfer, commercial knowledge service provision, and still other so-called "Third Stream" activities – at the expense of the many other forms of interaction between universities and the economy. "This concern is particularly acute "among leading traditional universities that value the close integration of teaching and research, that operate across all the disciplines, and that engage with society in very many diverse ways that include, but are not limited to economic transactions. Some are undertaken to achieve a directly financial outcome, while most are promoted for their wider, out-reach and often long-term benefits. Furthermore, there is a broad spread of missions within the British University sector." See: http://www.clo.cam.ac.uk/3rd_arm_metrics.htm] It is a sign of the times that the statement of the Russell Group, rather than focusing concern upon the possible sacrifice of the universities' creditability and capacity in a wider set of non-commercial social interactions, focuses on the need for methodologies to allocation government support for the diverse array of "Third Stream" activities in which they typically engage, and therefore to protect there institutions position in the face of potential competition from more specialised claimants for "Third Stream" support.

promoting the search by universities for income from the commercialisation of their so-called "knowledge assets", the efforts of technology transfer offices and intellectual property management organisations are directed towards fixing the terms of collaborative research projects so as to augment the flow of income to their respective institutions. Of course, this pits each against the similar interests of the other collaborating institutions, and likewise against the business companies whose participation (and sponsorship) is predicated on obtaining a satisfactorily large share of the prospective economic returns.

One should not underestimate the seriousness of the difficulties that have thus been created for university officers who are given responsibilities for negotiating true inter-university agreements that have to resolve conflicts over the division of prospective proceeds from the commercial exploitation of research findings. The same conditions also may give rise to tensions between groups of academic researchers who are keen to participate in a particular collaborative project and their respective university legal counsels, who are equally intent upon limiting (as far as is possible) their own institution's exposure to liability and other legal risks, as well as asserting its claim to the largest possible flow of material benefits that the project might yield. The immediately perceived interests of the university as a legal corporate entity may involve asserting provisions that do not necessarily advance the interests of the researchers, and indeed would be enforced at their expense. A more subtle effect that, as has been noticed, can have a corrosive effect on the trust necessary for successful research collaboration, is the well-intentioned effort of a diligent lawyer to render explicit *ex ante* all the things that conceivably could go wrong, and all the forms that betrayal of trust that the collaboration might sustain.

Decision-making becomes particularly complex when the interests of the parties diverge (let alone conflict) and no one set of specialised actors understands all the issues and is in a position to balance the potential risks of going forward with the collaboration as it has been designed by the researchers, against the risks of losing the prospective benefits by imposing a different set of arrangements that would frustrate the research itself, or weaken the incentives of the participants to behave co-operatively. Numerous examples might be cited in this regard, but a familiar comparatively benign illustration of the general problem may be seen in the situation of the specialised service offices that most research universities have found it necessary to establish in order to facilitate the transactions with external sponsors of research.

Typically, the "sponsored research office" in a U.S. academic institution (and its counterpart, the "university research services" office in Britain), develops familiarity and expertise in regard to the panoply of regulations and requirements that funding bodies impose upon applicants, and recipients of awards; and equally specialised skills in anticipating the issues that will arise in negotiations with corporate research partners. The performance of theses organisations is likely to be gauged primarily in terms of the volume of funding that their respective institutions receive from public (including charitable foundation) grants and contracts, and also from research partnership agreements concluded with business firms – not the satisfaction of the researchers, or the scientific and scholarly productivity of their sponsored research projects. University research services officers often have qualifications and practical professional experience in the law, but their role is quite distinct and in a sense more demanding than that of university legal counsel *per se*. They are thus able to provide intermediation services that greatly reduce the burden of ‘negotiation and administration’ upon the researchers. But, with expertise and the quest for efficiency comes a tendency on the part of such offices to promote compliance with standardised contractual formulae, to avoid undertaking to provide novel or highly ‘customised solutions’ that may better fit the needs of a particular research project. The main institutional problem posed by such solutions is that they are unfamiliar, and hence all too likely to occasion time- and attention-consuming special negotiations with other departments within their own university, as well as with the external agencies.[25]

As a specialised intermediary the sponsored projects office is thus pre-disposed by its own incentive structure to implement, rather than to question the need for contractual provisions that are asked for by diligent university solicitors – functioning in their specialised domain to protect the institution from the entire array of harms to which it may be exposed by a proposed research agreement. Thus, it is the extraordinary ‘university research services office’ that can be expected to take upon itself the role of serving university researchers as an ‘agent’ serves a ‘principal,’ and a professional firm of lawyers serves it clients. At the same time, although it occupies the intermediary position upon which converge the varieties of diverse and possibly divergent interests within the university in regard to particular proposals for collaborative research, these service organisations have not been given explicit discretionary authority to strike a balance among those contenting interests.

For the individual researchers who have initiated a collaborative proposal, matters are made more complicated and potentially more difficult by the fact that the principal contracting parties, legally speaking, will not be their other scientific colleagues but, instead, those colleagues’ respective institutional employers. This really is a two-sided problem, because the success of most collaborative projects will depend upon the work of individual researchers and research administrators with scientific expertise. The peculiar employment structure of academic institutions, with their tradition of academic freedom, and the increasing mobility of scientific researchers, means that the contracting university often has little control over its ability to deliver the services that it promises in the contract to establish a research project. It may well be that if the incentives of universities to enter collaborations are further increased in the future, there will be corresponding efforts to alter the contractual relations with their employees so as to more closely emulate business corporations – particularly in respect to the power of the latter to direct and control the participation of employed researchers in designated co-operative projects involving other organisational entities.

While there will be obvious practical difficulties in pursuing such a restructuring of university employment relations, it is not evident today that the idea of would be rejected quickly, either as inappropriate or as difficult to reconcile with the ethos of scholarly autonomy. Increasingly, universities in the United Kingdom and elsewhere are coming to

[25] In the latter context, it is quite possible that the presentation of unusual (anomalous) contract proposals may be read by their counterparts outside the university as a symptomatic of deficient professional competence, or lack of authority within the institution they serve; whereas, just the opposite interpretation would be more forthcoming were they perceived to be acting on behalf of a legal client – in this case the researchers, or their university.

regard their members (faculty and student members alike) as 'knowledge assets'. The very term suggests that those responsible for the disposition of university assets should be able to exercise more complete and exclusive control over faculty members' activities, so as to better deploy them in pursuit of greater revenues from the intellectual property and commercially valuable 'knowledge services' they are able to generate. In most cases there is little likelihood that the incremental revenue streams thereby captured from the intellectual property arising through the assigned work of employees and students will materially alter the institution's financial situation.[26] Yet, as economists are quick to point out, resource allocation decisions are determined "at the margin." Consequently, it is quite frequently observed that the prospect of the university achieving a comparatively small financial benefit will elicit the expenditure of significant administrative effort in altering long-established policies and operating rules. Likewise, the prospect of the university being exposed to a low-probability and low-cost risk, or of having relinquished some small gain to a "partner" institution in a particular project, is all too likely result in efforts by the institution's solicitors to forestall such outcomes by imposing blanket *ex ante* restraints upon the actions in which individual employees participating in a collaborative project are allowed to engage.

The three dynamics the have been reviewed in this section are at present only partially understood. This is a problem, because the viability of various possible institutional arrangements to support e-Science ventures depends to some extent upon knowledge of the situation in which many collaborators in a designated research area are likely to find themselves. More information, therefore is needed not only about the direction of future institutional changes affecting PROs, but about current individual experiences with informal arrangements, and with the ways that formal legal rules are presently being utilised by co-operating institutions. As information of this sort would be gathered more or less automatically as a part of the process recommended in the Report's fourth and concluding part, further discussion need not be pursued at this point.

2.2 THE INSTITUTIONAL DESIGN CHALLENGE: ANALYTICS

Seen in properly broad perspective, the global e-Science system design challenge is one of finding the set of technical *and* social and mechanisms that will provide the collaboration facilities, incentives and controls needed for human-machine research organisations to emerge and function efficiently in cyberspace. Inevitably, this challenge will present itself in many different and very specific contexts in the course of the co-evolving interdependent adaptation of the system's institutional and technological components.

---

[26] See David R. Mowery, et al., "The Growth of Patenting and Licensing by U. S. Universities: An Assessment of the Effects of the Bayh-Dole Act of 1980," *Research Policy*, 30 (2001); Richard Nelson "Research and Technological Progress in Industry-An analysis of the American Experience*" International Symposium on Economic Development through Commercialization of Science and Technology,* Hong Kong, 2002.

It follows that close attention should be devoted to the requirements of both sets of components, and to exploring the possibilities of applying ingenuity and resources to achieve innovative solutions in the institutional as well as the technological domains. This will be a critical strategy not only for the long-term success of e-Science programmes, but for securing the potential benefits of the contribution they can make to enhance the effectiveness of collaborative work in many other areas of human endeavour.

2.2.1 'COLLABORATION GAMES': PLAYERS, INTERESTS AND RULES

To make headway towards that worthy objective, it is important to try to more clearly delineate the fundamental sources of the organisational and institutional problems that need to be addressed by PROs when defining the terms, rights, responsibilities and powers of the legal parties to research collaborations. 'Collaborating parties' is of course a blanket term that covers at least four categories of entities/actors. For the purposes of analysing the sources of such conflicts, the following should be distinguished:

- the research scientists as individuals;
- research units that have formal governance arrangements (departments, laboratories, and other consortia bound by agreements among the participating researchers);
- host institutions with whom the researchers are connected by contracts, and through which they may receive financial or other material support (universities, public institutes, foundations and trusts, private partnerships and corporations);
- public and charitable funding bodies, and private business organisations that furnish material support to PROs for the conduct of research (and related training) activities.

These are presided over, of course, by national and international entities that may exercise primary or derived regulatory jurisdiction over both the individual researchers qua citizens (in the case of national governments) and their host institutions.

For the purposes of a general discussion, and also in some specific contexts, it is helpful to further simplify matters by consolidating the foregoing list into two categories, the first pair of parties being lumped under the heading of 'research collaborators' and the second aggregated into 'institutional partners/hosts.' Seen from that highly stylised perspective, there are two sets of core difficulties of designing supporting institutional arrangements for collaborative projects. These may then be succinctly characterised as arising in the first place from the imperfect alignment of the interests of the research collaborators, on the one hand, and the institutional partners, on the other hand; and in the second place, from the existence of mal-alignments or outright clashes of interests among the institutional partners. In some significant degree, the second class of difficulties also may contribute to the tension between the researchers whose goals impel them towards collaboration, and their respective institutional hosts for whom the terms of such coalitions may be problematic. Only in what economists term 'perfect teams' is it appropriate to ignore the consequences of conflicting goals and interests among the parties, but in the world of human agents the 'perfect team' is a conceptual device, not a reality.

Incentives for competition and collaboration are important for both types of actors in e-Science collaborations. But, it must be recognised that the point where those two forces would be 'naturally' balanced is not the same in the typical case of collaborating researchers as it for the institutional entities (i.e. public and private corporations and institutes and universities) that enter into a research partnership or consortium. The situation within scientific work-groups is usefully distinguished from that which typically holds among participating institutions, and can be examined first.

Within scientific work-groups situated in academic milieus, it generally holds true that the 'open science' ethos and traditions of scientific co-operation among researchers forms at least the point of departure (or default position) for the 'bottom up' organization of collaborative activities. This remains the predominant orientation, even though norms of co-operative behaviour are strained by rivalries for scientific recognition and reward; and it continues to be useful as a first-order idealisation of complex situations, the existence of considerable variations among the local *mores* characterising 'open science' communities in different fields of inquiry notwithstanding.

What the future holds in this regard remains unclear. A more elaborate and nuanced account of the current situation in academic science would emphasize the respects in which norms of professional behaviour and institutional policies are in flux. It is uncertain that the traditionally prevalent disposition in favour of scientific co-operation will be able to withstand the pressures from the newly ascendant spirit of 'academic entrepreneurship.' Similarly, at the institutional level it is far from obvious that commitment to the collective goal of 'the pursuit of knowledge' will continue to guide the policies embraced by a majority of leading research universities. In many places it already has ceased to prevail in the face of the instrumental emphasis placed by public funding agencies upon the wealth-creation function of knowledge, and the growing legitimisation of the pursuit by researchers of personal wealth through ownership of intellectual property.[27] Quite obviously, these are important issues not only for the scientific communities involved, but also for society at large. Moreover, they are issues whose ultimate outcome can be powerfully shaped by the effects of myriad, seemingly small decisions about the technological and institutional infrastructures of e-Science.

Among the institutional partners of a collaboration, by contrast with the scientists carrying out the research (and for whom publicly funded research universities may be said to serve as hosts), the predominant natural orientation lies more towards competition than co-operate. This generalisation may seem paradoxical, but at a fundamental level it follows from a simple contrast in motivations. The primary objective of collaborating with other researchers is to gain access to the immediate data and informational fruits of each other's knowledge. But, data and information are public goods that may be exploited by all the collaborators in their respective research endeavours – without becoming exhausted or in any respect diminished. By contrast, the driving motivation for the corporate entities to enter a research partnership or coalition (qua university or *qua* business corporation) is to gain access to material benefits that do not possess "public goods" properties, so that the existence of opposing interests among them over the division of the pie is ineluctable. When university/hosts are moved to become partners in a multi-institutional project, each is likely

[27] See, e.g., Owen-Smith and. Powell, "Careers and Contradictions" (2001).

in some degree to be responding to the influence of derived motives for co-operation – whether it is to accommodate the scientific work of academic collaborators, or thereby to gain the overhead funding, or the possible payoffs in prestige and command over material resources that a successful project might bring. But such co-operative and accommodating motivations tend to be tempered, and sometimes over-ridden by the attention that the institutions accord to protecting and promoting their respective individual, and essentially competing interests.

Where universities are involved as the proximate corporate parties, the primary impetus towards facilitating co-operation derives largely from the interests of the researchers themselves. This impetus also is likely to be reinforced by the terms on which funding may be obtained from the public sponsors of those research projects. In addition to the prospective division of whatever 'pie' will thus be made available to the coalition, the negotiated terms of each institution of higher education's (HEI) immediate relationships with its institutional partners also must be shaped by its entirely understandable concern to manage the uncertainties surrounding the conflict-laden aspects of these partnerships. Moreover, in the United Kingdom and other countries, the policies of governments now encourage higher education institutions and other non-profit research organisations (such as research hospitals) to try to exploit – whether for themselves or for their own private sector partners – the intellectual property arising from the contractually specified activities engaged in by their employees, and even their students. It is especially relevant for e-Science (as it is also for e-Learning) that these legal and administrative arrangements for institutional appropriation of the benefits of new knowledge have been focused particularly upon works created in the form of digital information.

Furthermore, the United Kingdom's government (along with other states in the European Union) appears to be pursuing regional development strategies based upon the formation of business 'clusters' in the environs of publicly supported research institutions, and the promotion of industrial enterprises founded by the licensing of intellectual property generated within the university, or in university-industry research partnerships. University leaders are understandably responsive to the prospects of direct and indirect benefits they may derive by successfully fulfilling this development role, and are thereby induced to aggressively seek to expand and exploit their intellectual property portfolios. It might be noticed, however, that presently there is very little in the incentive structure that enjoins United Kingdom institutions to consider the burdens that their policies in the latter regard can impose upon the work of their own research and teaching activities; and, *a fortiori*, upon the parallel activities carried on by colleagues at other institutions in the UK. Indeed, nothing in the present incentive structure requires a non-profit, publicly subsidised HEI to consider whether its intellectual property licensing strategies, and its promotion of university "start-ups", will be likely to impose burdens upon the innovation activities of private commercial firms.

2.2.2 COOPERATION v. COMPETITION WHEN ASSETS ARE COMPLEMENTS: SOME GUIDANCE FROM ECONOMIC ANALYSIS

For the purposes of this Report, the foregoing conditions may be taken to characterise the prevailing and prospective state of affairs. They thus describe the pertinent environment within which practical institutional arrangements for the conduct of collaborative e-Science will need to be achieved. It may be argued that societal interests would be better served by promoting more active co-operation among the public sector entities, whereas the recent thrust of government policy (in the United Kingdom particularly) has been to encourage higher education institutions to form co-operative relations with business companies while competing ever more vigorously with each another for command over material resources. But, in the likely absence of a radical policy reversal, the task of creating an appropriate institutional infrastructure for e-Science must be one of devising mechanisms that are better able to strike the socially most efficient attainable balance between the proximate goals of collaborating researchers and the immediate objectives of their host institutions.

The rationale for choosing this particular 'second best strategy' is not that scientific collaborators should be accommodated because they are intrinsically good, more selfless, and more worthy of trust than other members of society. Quite the contrary. Precisely that reality must be taken into account by the internal governance structures of the institutions that employ human agents and support their activities, whether scientific or other. The stance adopted in here derives from a different consideration altogether: namely, a recognition of the larger beneficial consequences for society that derive from the strong imperatives for co-operative behaviours in the field of scientific inquiry (and equally in other complex forms of cultural production). This rationale follows from some fundamental propositions in the economics of knowledge and information, and their application to the analysis of the role of implicit and explicit collaborations scientific and technological research.

Modern economic analysis of the production and distribution of reliable knowledge proceeds from the widely accepted proposition that the advancement of scientific knowledge and the technological progress are intertwined cumulative processes. Both are synergised by complementarities among the data and informational inputs that enter into systematic research activities and which, in turn, are the primary outputs of those activities.[28] Co-operative sharing of knowledge resources is well known to be the most efficient allocational scheme for the production of goods when different agents hold complementary inputs. This holds for normal economic goods, and is true for data and information – true, *a fortiori*, as these possess special properties that render them akin to 'pure public goods.' The latter typically are integral and hence indivisible; yet they are infinitely expansible, being useable repeatedly and concurrently by many agents without becoming depleted or otherwise exhausted. Furthermore, to utilise such goods while denying others access to them is typically costly to arrange, even when it is technically feasible.

[28] For a more extensive presentation see, e.g., 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: A Primer," forthcoming in National Research Council, *The Role of the Public Domain in Scientific and Technical Data and Information: A Symposium*, National Academy Press, 2003. [Available at: http://siepr.stanford.edu/papers/pdf/02-30.html].

Formal analysis based upon the foregoing propositions has established an important result that also (satisfyingly) receives confirmation in behavioural experiments.[29] Where the complementary elements required to obtain a valued outcome are controlled by multiple agents, each of whom has the power to exclude others from use of at least one input, the 'prices' that these private goal-seeking agents independently will place upon the resources under their respective control will be too high. That is to say, they will fail to consider the upward-cascading effects of the charges they set for access to their respective resource-holdings, with the result that the level of level of production in the system as a whole is sub-optimally low.

Moreover, if there are as well strong complementarities in the use of the outputs for final purposes (such as commercial innovations based upon scientific discoveries or research tools), the dispersal of exclusive ownership (and hence access control) among many parties tends to yield a less-than-socially optimal degree of utilisation of the available information-inputs, and a correspondingly sub-optimal level of consumption of the final goods. Overall, the anticipated results of the envisaged 'anti-commons' equilibrium are less socially efficient than those that would obtain were the knowledge production activity to be organised under the terms of an intellectual property licensing "pool" that pursued a monopoly pricing policy.

The latter is a very strong finding, indeed. It is widely accepted that regimes characterised by competitive rivalry generate incentives for cost minimisation and efforts to satisfy the needs of final consumers, whereas monopoly imposes significant inefficiencies, particularly where the product that is monopolised possesses the properties of a pure public good" – which is the case for information-goods. Hence, finding principles for formally organising e-Science collaborations, and doing so under contractual terms that manage to avoid outcomes that will be "worse than monopoly" is a challenge well worth trying to meet.

2.3 THE FORMAL LEGAL CONTEXT FOR COLLABORATIVE E-SCIENCE

It is notoriously difficult to describe exhaustively the legal issues that could arise from the many different types of collaborative e-Science project. Nevertheless, common patterns of legal problem are likely to arise. These concern the legal relationship between the parties to

[29] See James Buchanan and Yong J. Yoon, "Symmetric Tragedies: Commons and Anticommons," *Journal of Law and Economics*, 43(1), April 2000, for a theoretical analysis that makes use of Cournot's theory of oligopoly behavior in markets for complementary products. On the latter, see Carl Shapiro, "Theories of Oligopoly Behavior," in *Handbook of Industrial Organization*, R. Schmalensee and R. Willig, eds., Amsterdam:Elsevier Science Publishers, 1989: pp. 330-414. Carl Shapiro "Navigating the Patent Thicket: Cross Licenses, Patent Pools, and Standard-Setting," in *Innovation Policy and the Economy*, Vol.1, eds., A. Jaffee, J. Lerner and S. Stern, eds., Cambridge: MIT Press, 2003) develops the argument that where intellectual property rights in complements are distributed among many agents, compulsory "pooling" of rights and cartel pricing yields a socially more efficient allocation than absolute individual monopoly rights of the sort granted to patent-holders. For experimental validations of propositions about the symmetry and commons and anti-commons forms of market failures, see Charles. F. Mason and Owen R. Phillips, "Mitigating the Tragedy of the Commons through Cooperation: An Experimental Evaluation*," Journal of Evironmental Economics and Management,* 34, 1997: 148-172; Steven Steward and David J. Bjornstad, "An Experimental Investigation of Predictions and Symmetries in the Tragedies of the Commons and Anticommons," *Joint Institute for Energy & Environment Report*, JIEE 2002-07 (August),2002. .

an e-Science collaboration, the material that the parties bring to an e-Science collaboration, the material to which such a collaboration gives rise and the liability of the parties for harms arising from the project. Each of these issues merits separate consideration. That consideration must remain fairly general as the identification of legal issues is very fact-sensitive. But, even general consideration of the likely issues suffices to reveal the complexity of the legal context in which collaborative e-Science will operate. In light of that complexity, it is clear that these are issues that cannot be navigated by individual scientists, and that institutional arrangements similar to those outlined in the final section of this paper will be essential.

2.3.1 RELATIONSHIPS AMONG THE COLLABORATING PARTIES

The legal rules that govern disputes between the parties to an on-line collaboration are determined by the nature of the legal relationship between those parties. Three possibilities suggest themselves. First, the parties may be in no particular legal relationship and the general law will determine issues such as the allocation of the fruits of their joint activities. This situation is the most unlikely in all but the smallest and most informal of collaborations. Second, the parties may be in a contractual relationship. If they are, then the terms of that contract will in most circumstances determine the conduct of their relationship and the allocation of its outcomes. Third, the parties to collaboration may establish some type of institutional vehicle for their collaboration such as a joint venture company. This is the most formal way of establishing a collaborative relationship. It has the advantage that it can facilitate the structuring of complex relationships. It also means that individuals and institutions can avoid liability in situations in which a collaborative project may give rise to harm to third parties. However, it entails the maintenance of a system of legal formalities that is likely to make it unattractive to all but the very largest of collaborations.

Of these three possibilities, the most likely one is that the relationship between the parties will be governed by either an express or implied contract. The question of whether the parties are in a contractual relationship is not simply that of whether they have entered into a written agreement. Even in situations in which the parties cannot point to any such written agreement, the courts may well determine that they are bound by contract either because of an express verbal agreement between them or because the court is prepared to imply a contractual relationship. The freedom that parties have to enter into contractual relations can give them considerable control over the conduct of their relationship and the allocation of its outcomes. But, difficulties may arise; both when collaborators find themselves party to contracts the terms of which they did not expressly agree and, alternatively, when apparent agreements between collaborators are ineffective as contracts.

A first difficulty that arises in relation to the contractual organisation of the relationship between collaborators relates to the issue of when, and on what terms, their relationship is formed. This problem will not arise in circumstances in which the collaboration is based on a written signed agreement which is remade by all the parties to the collaboration each time its terms are amended or the identity of the parties changes. Many collaborations, however, will be organised on a more informal basis. A part of the attraction of e-Science is that it can involve many different parties with different contributions to make who become involved at different stages in a project. Two particular problems are likely to emerge. First, the parties to a collaborative project will often amend their agreement informally during its life. This may cause a problem if one of the parties is promising to do more than she undertook under the original contract, while the other is not. Under English law, the problem of giving effect to such amendments is a textbook difficulty in the law of contract. Other legal systems have less difficulty with such situations and the English courts have been working to find a solution to the problem. Nevertheless, it is one that may well arise in on-going scientific collaborations.

Second, latecomers to a project may be included in its work without entering into an express agreement with all the existing parties.[30] In relation to the rules of unincorporated members' clubs, the law has developed methods for dealing with the problem of how contracts governing on-going relationships can be extended to include new-comers. In such contexts it is generally assumed that the members of a club are in contractual relations with one another; that the terms of that contractual relationship are set out in club rules, including procedures for the admission of new members; and that those rules can only be altered either in the manner prescribed in the rules themselves or by the agreement of every member of the club. There is no reason why such an approach might not also be adopted as a means of organising the entrance of participants who joined the collaboration at various points after its inception. This may not be possible, however, if the newcomer were to enter the relationship on a basis that has not been anticipated at the outset of the project. Moreover, even in situations in which the agreement between the collaborators allowed for the admission of new-comers, the terms of the contract among the original members of a collaboration would need to be made clear to the late-comer. The difficulty with contractual analyses of these types is that the outcome to a dispute between the parties, particularly one involving parties to a collaborative relationship who have joined at different points in time, may not reflect the expectations of all the parties. Indeed, such expectations are unlikely to coincide.

A second difficulty will concern the appropriate parties to any written contract that forms the basis of the collaboration. For example, a written agreement may be arrived at among institutions rather than individuals, but a collaboration will require agreements (whether formal or information) among particular individuals who possess with particular expertise. Difficulties may arise where an institution enters into an arrangement for a particular collaboration, the individual who is central to the collaboration moves institutions and the original institution seeks to replace them and to continue the project with someone with whom the other collaborators are not content. A similar difficulty might arise in the context in which the parties to collaboration assume that they are entitled to the fruits of that collaboration but, in fact, the general law grants first ownership of those fruits to another party. An example of this might be the situation in which individual collaborators assume that they are able to make arrangements concerning the fruits of their collaboration, but at least one party's share falls under the general law to the institution by which she is employed. There are various ways in which such problems can be overcome, but it is essential that the issue of the most appropriate contracting parties be addressed before the project begins.

---

[30] Academic lawyers puzzle over cases such as *Clarke v Dunraven* [(1895) P 248] and how rules of contract law designed to reflect the paradigm of two or more parties in an express agreement on terms that govern their relationship throughout its life can be adapted to the context in which parties attempt to join a single agreement at different points in time.

This highlights a third potential difficulty with the contractual organisation of a collaborative research project. Where the fruits of a collaborative arrangement constitute the subject matter of one of the statutory intellectual property regimes, the statutory code itself allocates its first ownership. As outlined below, the parties to collaboration are free to arrange between themselves that the normal rules for the allocation of property in such circumstances should be altered. Yet, the process of doing so involves some possible pitfalls. In particular, informal agreements may be difficult to enforce. Take, for example, the situation in which one of the parties to collaboration writes the software upon which the project is based and copyright law allocates first ownership of that software to her. The author of the software may, by agreement, assign her interest in the software to her collaborators or license them to use it. She may even, as a part of the original collaboration agreement, have assigned her interest in any prospective works, including the software that she might write as a part of the project. In the United Kingdom, this is provided for by section 91 of the Copyright, Designs and Patents Act 1988. However, to do so she must comply with the legal requirements of such an assignment and under section 90 of the Copyright, Designs and Patents Act 1988, these include a requirement that assignments be in writing signed by the assignor. Any agreement that may be entered into as to the allocation of the fruits of collaborative work must not only be clear and made between the appropriate parties, but also meet the formalities required by any applicable law.

A fourth set of difficulties relates to the often-unequal bargaining power of the parties to collaboration. The law is able only in the most extreme circumstances to correct the effects of unequal bargaining power. This is one reason why it is vital that common understandings of the appropriate response to particular issues that arise in planning scientific collaborations should be developed. Scientists may come under considerable pressure from their commercial partners to agree to terms that do not further the goals of collaborative research generally, and to which they would prefer not to agree. To illustrate this point, one may consider the way that the effects of such asymmetries shaped the outcome of an unusual bargaining situation involving intellectual property, a recent case where no revenues from IP licensing were at stake. The case in question involved negotiations between the publicly funded Globus Project at the Argonne National Laboratories (Chicago) over the form of the "open source" license for the Globus Toolkit software under which major business firms – including IBM and Oracle to be specific – would undertake the distribution of that software package.[31]

Compared with those companies, the Globus Project had very limited capabilities to engage on its own in a widespread "free and open source" distribution of the Globus Toolkit software. It might have done just that under the familiar form of GNU General Public Licence (GNU GPL).[32] But IBM and Oracle were unwilling to expose themselves to ranges

---

[31] See Appendix 4, below for further details and references.

[32] The question of whether or not a computer program qualifies as "free and open source software" (F/OSS) can be approached as a matter of legal definition. (See David McGowan, "Legal Implications of Open Source Software," *University of Illinois Law Review*, 241,2001.) Copyright law gives developers who write programs the exclusive right to reproduce the code, distribute it, and make works derived from their original work. Copyright holders can grant other parties permission to do such things through licenses, and "free"

of risks they perceived would ensue were they to distribute code under the 'copyleft' type of license familiar in "free and open source software" (F/OSS) products released under GNU GPL. Recognition of its limited leverage, and a desire to rapidly establish the GT protocols as a Grid standard, appears to have been sufficient grounds for the Globus Project to accede to its industrial partners' demands by working with the companies' attorneys to devise a customised software license.[33]

The effects of asymmetry of bargaining power are further revealed by considering a parallel case of negotiation, in which a different and less restrictive form of software license was adopted. Sun Microcomputer's Project JXTA has created a set of open, generalized peer-to-peer protocols that allow any connected device (cell phone, to PDA, PC to server) on the network to communicate and collaborate. The source code for Project JXTA has been released under a variant of the Apache software License: *The Sun Project JXTA Software License* is functionally equivalent to the *Apache Software License (Version 1.1)*, with minor changes to reflect the Project JXTA name and Sun Microsystems as the original contributor. In addition, by contrast with the Globus Toolkit's exacting "contributor's agreement," a developer seeking to contribute to Project JXTA – either through patches or by becoming a "Committer and/or Project Owner" – can sign an agreement similar to the one required by the Apache Software Foundation. The contributors simply attest that to the best of their knowledge the code submitted is their own development work, and that they possess the authority to provide it and any related intellectual property to Project JXTA. In this case, the software development project had access to distribution capabilities, giving Sun the freedom to make use of the less restrictive and more familiar terms in the Apache Foundation's agreements.[34]

---

software or "open source" software refers to software distributed under licenses with particular sorts of terms. The "Open Source Definition" (maintained by the Open Source Initiative, see *The Open Source Definition* version 1.9 [Available at www.opensource.org/osd.html.]) provides a convenient and widely accepted reference guide to such licenses. It sets out several conditions a license must satisfy if code subject to the license is to qualify as "open source software." Several well-known licenses satisfy the Open Source Definition, the most widely used and still more widely discussed among them being the GNU General Public License (GPL). Programs distributed under a F/OSS license "must include source code, and must allow distribution in source code as well as compiled form; it also must allow modifications and derived (GPL'd) works, and "permit them to be distributed under the same terms as the license of the original software."Such a license does not restrict any party "from selling or giving away the software as a component of an aggregate software distribution containing programs from several different sources," but the licensee cannot "require a royalty or other fee for such sale." Much of the attention given to F/OSS development focuses on the GPL's requirement that authors who copy and distribute programs based on GPL'd code (derivative works) must distribute those programs under the GPL. This requirement is specified in Section 2(b) of the GPL and is refered to as the "copyleft" term.

[33] For reasons that similarly may have been rooted in the risk averse stance of IBM and Oracle, these negotiations also resulted in the creation of a novel form of agreement governing future contributions of code to the Globus Project. Contributing developers must grant Globus a perpetual, world-wide royalty-free license to use the submitted code, and automatically would lose their right to use Globus Toolkit were they to file suit for infringement of their intellectual property rights in their contribution(s). See Appendix 4.

[34] It may be noted that Sun sells a range of complementary JXTA products (software applications) that are available under proprietary software licenses, as well as commercial software systems services. On the emergence of business models that build commercial offerings around free and open source software, see the survey in S. Arora and P. A. David, "Commercialization of Open-source: Symbiotic or Parasitical?," SIEPR-Project NOSTRA Working Paper, Stanford University. May 2003. [Papers from this project are available at: http://siepr.stanford.edu/programs/OpenSoftware_David/NSFOSF_Publications.html.]

A fifth set of complexities may arise because the parties to collaboration, or their institutional employers, are in several different legal jurisdictions. In such a circumstance, one or more of the relevant parties may assert that issues concerning the existence, interpretation and enforcement of a purported contractual relationship should be determined according to the law of their own jurisdiction or by their own courts. They may also assert that issues concerning the first ownership of the fruits of collaboration are matters for their own law and their own courts. This becomes a matter for the complex rules as to the conflict of laws.

Questions concerning the law applicable to disputes concerning collaboration contracts will be governed in the English courts by the terms of the 1980 Rome Convention on the law applicable to contractual obligations. Under the terms of this Convention, the parties can choose the law that will govern disputes as to their contractual relationship, including its validity, by carefully worked out choice of law clauses included in express agreements. Yet, the ability of the parties to choose the law applicable to contractual disputes concerning intellectual property sometimes will be limited under the Convention by the application of a "mandatory" rule as to the applicable law. In situations where no such choice of law clause is included, complex rules will determine which law ought to govern the contract. Under Article 4 of the Rome Convention, this will be the law of the country which has the closest connection with which the contract or, in cases concerning the existence of a contractual relationship, with the purported contract. In ordinary commercial situations, however, a number of more or less clearly formulated presumptions govern the issue of what constitutes connection with a particular country. But, it would be very difficult to predict what might be taken to be 'connected with a particular country' in the context of a collaboration agreement that pertained to activities carried on in several different jurisdictions, none of which could be said to be the "home" of the project. Legal doctrine concerned with 'conflict of laws' has yet to come to terms with activities that take place in no particular territory. Once again, it should be emphasized that issues concerning contracts that involve intellectual property rights pose special legal difficulties.

Questions as to the jurisdiction of the English courts in disputes concerning collaboration contracts are likely to be governed by a regulation of the European Union, Council Regulation (EC) 44/2001 of 22 December 2000 on jurisdiction and the recognition and enforcement of judgments in civil and commercial matters. Article 23 of that Regulation permits parties to make agreements as to the courts that will have jurisdiction over their disputes. This provision applies when one of the parties is domiciled in a Member State of the European Union, but a similar possibility exists at common law. In the absence of such an agreement, extremely complex rules will determine jurisdiction, these rules starting from the presumption that a defendant ought to be sued in his or her own jurisdiction. The rules governing jurisdiction in the absence of express agreement will again vary depending upon whether the defendant is, or is not, domiciled in a Member State of the European Union. Moreover, disputes involving multiple parties from different countries add a further layer of complexity to the rules determining jurisdiction. It is essential that the arrangements underpinning international collaborations include clear and effective choices about both the laws applicable to the collaboration and the courts in which disputes concerning it will be heard.

A sixth set of problems arises because of the expertise required to interpret contracts for scientific collaboration. It may be that the parties to a scientific collaboration cannot agree on the courts in which they would like their disputes to be resolved simply because they do not believe that any court in any jurisdiction has the expertise to interpret the contract. This may either be because the contract touches upon technical issues, or because they want the contract to be interpreted by someone who understands the cultural norms of the scientific community in which they are operating. In such circumstances, the parties may enter into an agreement that their disputes will be subject to arbitration or mediation rather than to the jurisdiction of a court. In most jurisdictions there is provision, as there is in the United Kingdom under the Arbitration Act 1996, to render the decisions of arbitrators binding in most cases and the parties may well choose to do so in situations in which they have chosen an arbitrator for her scientific expertise. In such circumstances, the parties may also agree to the legal or other standards by which their dispute is to be resolved. Arbitration can be relatively less expensive than court proceedings and the choice of an arbitrator to resolve disputes may solve some of the difficulties associated with international collaborations.

From these doctrinal considerations, then, it is clear that close attention must be paid to the legal basis of a collaborative relationship right at the outset of the project. But the potentially high cost to the project of negotiating its every possible outcome is extremely high. The third part of this paper is devoted to developing a mechanism for minimising that particular problem. To anticipate, it seems that there is a need to develop an array of standard contractual clauses, covering particular issues that are likely to arise in e-Science collaborations and a set of principles for the development of such agreements. This approach is not to advocate the introduction of a set of standard e-Science collaboration agreements. The formulation of such agreements would be an almost impossible task given the likely diversity of e-Science projects and the fact that the uses of the technology supporting on-line collaboration are currently largely unknown. Standard form agreements may petrify norms at an early stage in the development of the social practice of e-Science. Moreover, standard form agreements are fraught with legal risk. The development of standard clauses, rather than standard form contracts, and principles for the use of those clauses is likely to yield a far more reliable mechanism for producing robust collaboration agreements at a minimum cost to the relationship of those involved.

2.3.2 ISSUES REGARDING MATERIAL CONTRIBUTED TO COLLABORATIONS

Three questions arise in relation to the material that is contributed to on-line collaborative projects. The first relates to the legal risk to which a party to such collaboration might be exposed by the use of material contributed to it by another participating party. The second relates to the extent to which the pooling of resources protected by intellectual property rights might constitute a breach of the rules of competition law. A third relates to the extent to which the contribution of particular resources gives a party a claim to the product of a particular collaborative research activity.

2.3.2.1 Legal risk

Take first the question of legal risk. An important feature of establishing trust between the partners to on-line collaborative research will be the extent to which they can confidently use the resources that each brings to the collaboration. Each party will need to know that her involvement with the project has not exposed his or her to unanticipated legal risk to third parties. Assume that two parties to a collaboration bring material together for a database in the health care sector. The material that they bring may well be the subject of data protection rights, rights against breach of confidence, or intellectual property rights such as copyright and database right, which are violated by its use as part of the project. A party to a collaboration who receives and uses such material may be in breach of her obligations under data protection or intellectual property laws – even though ignorant of the fact that the data in question contains material for whose use adequate consent was not obtained , and thus constituted material infringing an intellectual property right. In general, information received in breach of an obligation of confidence will be protected only if there is notice at some time before the information is used of the circumstances in which the information has been obtained. (The requisite notice, however, may be constructive rather than actual.)

At least in relation to obligations as to data protection, copyright and the database right, this risk may be less great than it at first appears. Under the terms of the European law of data protection – enacted in the United Kingdom as the Data Protection Act 1988 – consent will not be necessary for the secondary use if the data does not relate to identifiable individuals. It will, in any case, be available to scientific research without additional consent, as long as it is not being used to support decisions with respect to particular individuals, or used in a way that is likely to cause substantial distress to an individual data subject. Similarly, an exception to infringement of copyright and database rights exists as long as the use of the protected material is for research the use is not “commercial”.[35] Importantly, however, no broad research exemption exists in relation to patent rights. Although patent lawyers often talk of a research exemption, in practice it is extremely limited.[36] The risk that legal consequences attach to the use of material brought by a partner to collaboration is a real one. This risk is exacerbated in the context of the use of the potentially unlawful material in international collaborations, and particularly those that take place over the Internet. But even in the straightforward situation involving the law of only one jurisdiction, the problem of legal risk remains an important consideration in the establishment of collaborations.

The lawyer’s answer to the foregoing problem is that the parties to collaboration ought to take cross-indemnities from one another as regards liability arising from the research material that each expects to provide. There are two difficulties with such a solution. The first, though somewhat theoretical, is that it will not protect a party from criminal responsibility in the unusual situation in which use of the material constitutes an offence in some jurisdiction whose courts find its law applicable. The second, and more real, difficulty

---

[35] The reference here is to the Copyright and Rights in Databases Regulations, 1997 section 20 and the Copyright, Designs and Patents Act, 1988 section 29. The latter is soon to be amended under the terms of Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 – on the harmonisation of certain aspects of copyright and related rights in the information society.

[36] For the U.K. see *Monsanto Co. v. Stauffer Chemical Co.* [1985] RPC 515 and *Smith Kline & French Laboratories Limited v. Evans Medical Limited* 1989] FSR 513. For the US, see *Madey v. Duke University* 307 F.3d 1351 (2002), currently on appeal to the Supreme Court.

is that the negotiation of an indemnity clause may operate to undermine, rather than to build, the trust that is necessary to establish an effective working relationship. This effect may be exacerbated when, as is usual, it is an institution that is entering a collaboration agreement and not individual collaborators. In most circumstances it will be an employing institution that will be vicariously liable for the wrongful acts committed by those in its employment and lawyers for the institution may well seek the strongest protection that can be negotiated. This is another situation in which the development of standard form clauses and principles for their implementation may effectively avoid the potential costs to a collaborative relationship of the need to anticipate the problems that a project might entail before it has begun. Once such indemnities became established in a field of scientific collaboration, the inclusion of them in a contract governing a particular relationship betokens less suspicion of a particular collaborator.

Of course, another and more positive way of addressing these questions about the legality of the materials used by a collaborative project is to see it not as a problem of legal risk, but as a matter of good information management. The United Kingdom e-Science 'Pilot' project known as CLEF is working to build a system that connects databases in the healthcare sector.[37] The framework will integrate clinical histories, radiology and pathology reports, annotations on genomic and image databases, technical literature and web-based resources to serve the needs of patients, their families and carers, clinical professionals and biomedical scientists, healthcare enterprises and the public at large. CLEF is acutely aware of the importance of ethical and legal issues regarding, in particular confidentiality and intellectual property. One of the stated goals of the project is to devise agreed policies on information governance and technical measures for their enforcement. CLEF follows a two-pronged strategy to ensure that it has permission for the use of data protected by confidentiality and intellectual property rules. The first strategy is to enlist the intellectual property owners and confidentiality watch-dogs as collaborators or supporters. The second is to foster consensus on information governance policies among these partners. The emergence of standard form contracts in the formation of collaborative databases could help to spread and to entrench these agreed standards of good information management in similar applications contexts.

2.3.2.2 Competition Law

The second issue that arises in relation to the material contributed to an e-Science collaboration concerns the rules of competition law. The parties to a contract governing a collaborative project need to be careful that their agreements are not subject to control by the relevant competition law authorities. This will occur in situations in which collaborative agreements are either collusive or constitute an abuse of a dominant position. In the European Union Articles 81 and 82 of the EC Treaty render such behaviour unlawful. The danger of falling foul of Article 81 arises particularly when parties to a collaborative agreement pool resources such as intellectual property in a way that has the effect of excluding competition. Imagine, for example, that a number of research groups own database rights in a series of databases which, if pooled, would effectively cover the field of a particular area of research: in some circumstances the very creation of such a pool may be anti-competitive under the

[37] See Appendixes 2 and 3 for further information about the CLEF project and its position among the 24 pilot projects of the e-Science Core Programme.

principles of European competition law. Alternatively, the creation of the pool may be allowed either under the terms of Article 81(3), which exempts from the operation of Article 81 certain agreements relating to research and development and the licensing of its results; it also could be found permissible under the terms of the technology transfer block exemption.[38]

In some other circumstances the resources to which a collaborative project gives rise may also create competition law problems under the principles expressed in cases such as C-*76/89R, C- 77/89R and C-91/89R RTE and Others v. Commission* [1989] ECR 1141. Such a situation would occur when the fruit of a collaborative research project becomes established as a dominant industry standard in a particular area of technology. The Globus Toolkit example [presented in Appendix 4] offers an instance of just such a suite of technical specifications which may emerge as the *de facto* Grid standard. But there the issues of potential abuse of monopoly rights over an "essential facility" have, in a sense been fully anticipated by the Globus Project, and therefore are unlikely to arise; the Globus Project and its industrial partners have agreed that GT is to be distributed as "open source" software package on a royalty free basis, under the terms of a (newly designed) form of public license.[39]

In this connection it may be remarked that the use of "free and open source software" (F/OSS) licenses such as the GNU GPL obviously offers an attractive way of invoking copyright law to deal with many of the difficult legal complications that otherwise arise from the ownership of intellectual property arising from "horizontal" collaborations (i.e., collaborations at the same stage of production and distribution). But, F/OSS licenses equally can facilitate the resolution of issues arising among the parties to "vertical" collaboration agreements – as has been demonstrated in the case of the Globus Toolkit. Where public funds support the development of infrastructure technologies of this sort, especially network technologies for which interoperability standardisation is critical, such solutions might be fully justified on economic efficiency grounds. Yet, at present the United Kingdom and other governments have held back from embracing this approach as a general principle. Indeed, as has been noted, the recent direction of government policies vis-à-vis PRO's has been to encourage the generation of revenue through the use of proprietary software licenses.[40] In this environment the issues of claims to the fruits of "essential" research tools created by publicly supported collaborative projects will have to be thought though very carefully with regard to the implications of competition law as well as intellectual property law.

---

[38] Commission Regulation (EC) No 240/96 of 31 January 1996 on the application of Article 85(3) of the Treaty to certain categories of technology transfer agreements

[39] See Appendix A, sect. 1, for details of the Globus Toolkit Public License (GTPL).

[40] Copyright law has been invoked by commercial software vendors in distributing programs physically and electronically under a variety of "shrink-wrap", "click-wrap" and "browse-wrap" licenses. In Appendix 4, section 2.3, notice is taken of an objection to the royalty-free terms of the Globus Toolkit Public License made by an evaluator of the GT protocol suite: the commentator, from a British university pointed out that the U.K. government, and the researchers themselves, expected to see their middleware (platforms) and applications generate revenues from licensing agreements.

The whole issue of the potential competition law problems involved in the establishment of e-Science collaborations and the exploitation of their fruits has been thoroughly addressed in the recent European Research Area Expert Group *Report on Strategic Use and Adaptation of intellectual Property Rights Systems in Information and Communications Technologies-based Research*. The conclusion of that *Report* is that the current application of the law to collaborative research remains unclear, and that:

> "… some clearer guidance from the Commission would be welcome on how it sees competition law applying in situations when industry standards require the use of a technology that is IP-protected and when access to research tools that are IP-protected is denied, or granted only at unreasonable rates."[41]

2.3.2.3 Individual contributor's claims to the fruits of collaboration

The third question to be considered regarding materials contributed to research collaboration is whether one party's contribution can give rise to a claim to the fruits of the whole project. This is a very complex legal matter. The nature of the answer turns, in large part, upon the nature of the contribution. Whether or not the latter will give rise to such a claim turns depends largely upon the rules as to the ownership of the fruits of a collaboration – a subject to be considered in the section following this. This is an important matter that deserves emphasis because when the parties to a collaborative project consider the allocation of the fruits of the collective efforts, they quite naturally tend to focus on the resources that each has brought to the project. But, as will be evident from the discussion in the following section, the significance that the parties attach to the nature (and scientific or economic value) of their respective contributions is not reflected in the legal rules concerning the allocation of first ownership of the fruits of a (scientific) collaboration.

2.3.3 ISSUES SURROUNDING MATERIAL ARISING FROM COLLABORATION

A scientific collaboration may give rise to 'information goods' and 'knowledge resources' of many different kinds. These 'outputs' may be protected as confidential information as long as they are kept secret. In some jurisdictions they may also, though very rarely, be protected by the law of tort under the general action against misappropriation. This tort prevents one party from "reaping without sowing" by exploiting a valuable intangible that another has created and, in effect, amounts to an uncodified system of intellectual property protection. However, this tort is unknown in many jurisdictions, including England, and is very limited in its application in those in which it is recognised. In most circumstances, the fruits of collaborative research will be protected, if at all, only by the statutory intellectual property codes. Thus the fruits of a collaboration will be able to be captured if it constitutes, for example, an invention that can be patented; a work (including a computer programme) that is the subject of copyright protection; a database over which the database right can be asserted or a plant variety which can be protected by plant variety rights.

---

[41] European Research Area Expert Group Report on Strategic Use and adaptation of Intellectual Property Rights Systems in Information and Communications Technologies-based Research (2003) p.26

A significant level of international harmonisation of the intellectual property codes has taken place in the last 150 years. This occurred first under a series of multi-lateral international treaties that emerged in the 19th century and has more recently come about under the Agreement on Trade Related Aspects of Intellectual property Rights including Trade in Counterfeited Goods (the so-called TRIPS Agreement).[42] Nevertheless, the rules concerning the ownership of intellectual property are significantly different in different jurisdictions. The discussion here will call attention to some interesting and important differences between the relevant bodies of law in several national jurisdictions, but the determination of which nation's law is to govern, and what court will hear disputes when the latter involve international collaborators, again is a very complex matter that must remains beyond the bounds of this review.

In the case of intellectual property disputes, the applicable law usually will be the law under which protection is sought, although the application of this principle is far from straightforward. In relation to the forum in which such a dispute should be heard, Article 22(4) of Council Regulation (EC) 44/2001 of 22 December 2000 on jurisdiction and the recognition and enforcement of judgements in civil and commercial matters and equivalent common law rules, provide that questions concerning the registration or validity of an intellectual property right will be heard in the courts of the jurisdiction in which the intellectual property right is claimed. So, for example, questions as to the validity or registration of a German patent will need to be heard in the Bundespatentgericht. By contrast, however, the conflict of law rules of many countries, including those of the United Kingdom, provide that infringement actions can be heard in the courts of other jurisdictions, often those in which the defendant is domiciled.[43] The context in which an alleged infringement takes place over the Internet raises even more uncertain questions of jurisdiction, particularly in light of the growing tendency of the courts, evinced in cases such as *Menashe Business Mercantile Ltd v. Julian Menashe* (2002) to claim jurisdiction over the infringement of intellectual property rights on the Internet involving activities that might be seen as taking place in another country.[44]

Again, the question of conflict of laws presents a minefield for potential international collaborators who may find themselves either defending suits, or bringing suits to protect their intellectual property, in foreign jurisdictions and under foreign law. This cannot be entirely avoided by carefully drafted contractual provisions as to the choice of law and choice of courts, but such provisions can significantly reduce the risks to a successful collaboration presented by the arcane rules as to the conflict of laws.

The first issue to be addressed in determining whether the outcome of a collaborative project gives rise to intellectual property that can be owned is whether it falls within the subject matter requirements of the statutory regimes. The most likely outcomes of such a

---

[42] The TRIPS Agreement is an appendix to the Marrakesh Agreement establishing the World Trade Organization (1994).

[43] See *Pearce v. Ove Arup Partnership Ltd and Others* [1997] 2 WLR 779.

[44] See *Mercantile Ltd v. Julian Menashe* [Court of Appeal, 28 November 2002, unreported].

project will be a patentable invention, a copyright computer programme or a database which might be the subject of either copyright or database right. A consideration of the issue of the copyright and database protection of scientific databases will reveal something of the intricacy of this class of issues.

Scientific and technical databases will be protected by copyright if they amount to original literary works. Under section 3A(2) of the Copyright, Designs and Patents Act 1988, a database will constitute such a work only where "by reason of the selection or arrangement of the contents of the database the database constitutes the author's own intellectual creation." It is therefore the manner in which the database is organised that is protected by copyright rather than its contents. There is academic debate about whether the standard of originality that must be met for copyright in a database is higher than that applied to literary works generally in relation to which is usually held that "original" simply means "not-copied" (*University of London Press v. University Tutorial Press* [1916] 2 Ch. 601). In the implementation of the Council Directive (EC) 91/250 of 14 May 1991on the legal protection of computer programs, the United Kingdom government seems to have assumed that it did not, as it was not thought necessary to amend the law of the United Kingdom so as to meet a similar provision defining originality for computer programmes. By the time of the implementation of the Council Directive (EC) 96/9 of 11 March 1996 on the legal protection of databases the government may have changed its position, as the Copyright, Designs and Patents Act 1988 was amended to include the same definition of originality for databases.

In the United States, the arrangement of a database may also be protected by copyright, but again the standard of originality required for protection is uncertain. The U.S. Supreme Court's ruling in the leading case, *Feist Publications v. Rural Telephone* (111 S.Ct. 1282 (1991)), made it clear that originality must consist in more than mere effort in the compilation of information. But inasmuch as the arrangement of the information in that particular case was alphabetical, it is difficult to determine just how much originality was required and whether the standard was significantly higher than that required in the United Kingdom. The situation is clearer in the European Union, where a database might also attract the protection of a database right under the terms of the Council Directive (EC) 96/9 of 11 March 1996 on the legal protection of databases, which was implemented in the United Kingdom as the Copyright and Rights in Databases Regulations, 1997.

This protection prevents not merely the reproduction of the selection or arrangement of a database, but the extraction or re-utilisation of a substantial part of the contents of the database. In order for a database to qualify for this protection there must have been substantial investment in obtaining, verifying or presenting the contents of the database. Copyright protection in the United Kingdom will be available to nationals of most countries. By contrast, the Database right will be available only to nationals of the European Economic Area, although it will also be available to someone not from the European Economic Area who jointly makes the database with a person (or company) who is resident there. Despite the inducement for reciprocal introduction outside the EU, which was seen to be the motivation for this departure from the "national treatment" provisions familiar under the Berne Convention on copyright protection, attempts to introduce similar legislation in the United States have so far not met with success.

A problem arises under the database right because of the content of many databases, and, indeed, most of the particularly valuable scientific and engineering databases are continually changing. The U.K. Copyright and Databases Regulations 1997 provide that:

> "17(3) Any substantial change to the contents of a database, including a substantial change resulting from the accumulation of successive additions, deletions or alterations, which would result in the database being considered to be a substantial new investment shall qualify the database resulting from that investment for its own term of protection."

This paragraph currently is the subject of a reference to the European Court of Justice from the English Court of Appeal in the case *C 203/02 British Horseracing Board Limited v. William Hill Organisation Limited*. In the court of first instance, Laddie J. held that the Regulations were intended to protect dynamic databases that are constantly being updated. He held that as a database is updated it is subject to a new term of protection on an on-going basis but that an unauthorised user who takes older data "only faces a database right which runs from the date when all of that older data was present in the database at the same time." Moreover, if someone "takes an existing database and adds significantly to it, he obtains protection for the database incorporating his addition."[45]

It is clear that this legislation will be difficult to apply in many circumstances involving on-line scientific collaboration. A database may be being constantly updated. It may therefore be very difficult to determine either the point at which it has changed identity and its term of protection thereby extended or the point at which new rights in the database have been acquired by someone adding to it. Good examples of such a database might be: (i) the Comb-e-Chem project in which results from an on-line test-bed in combinatorial chemistry are stored as a part of a database that is constantly being updated by each new user, or (ii) the GENIE project which creates a constantly updated database of results arising from the use of a Grid enabled integrated earth system model.

The determination of the ownership of the knowledge resources arising from on-line collaborations is often far from simple. Assuming that the outcomes of on-line scientific collaboration include subject matter that can be protected under the various statutory intellectual property codes, it then becomes a matter of determining who has ownership over that subject matter, either under the statutory regime itself or some legally effective agreement. Most of those involved in on-line science will be working for a university or other research institution. They may also be working in partnership with a private firm or with funding from a public or private body. Importantly, they may well change employers during the life of a particular collaborative research project. Disputes will inevitably arise as to the ownership of valuable knowledge resources between these different parties.

The allocation of the first ownership of intellectual property varies slightly from jurisdiction to jurisdiction and from statutory regime to statutory regime. In general terms, however, the law of the United Kingdom allocates first ownership to the creator of the

[45] *British Horseracing Board Limited v. William Hill Organisation Limited* [2001] R.P.C. 31.

particular resource or to her employer. For these purposes, an employer is not simply someone who provides the funding for, or commissions, the research. Rather, it is someone with whom the creator is in a "master-servant" relationship. In order for first ownership of the resource to be allocated to the employer, the material in question must be created in the course of employment. In the field of inventions, the allocation of first ownership to the employer is somewhat offset by section 40 of the Patents Act 1977, which provides that an employee inventor can be remunerated in circumstances in which his or her invention proves to be of "outstanding benefit to her employer." But claims under this provision are rare.

The application of all this law to the position of academics in the United Kingdom has been much debated. For some while it was the received wisdom that copyright in works created during university employment first vested in the academic author, whereas rights to patentable inventions first vested in the university.[46] In the wake of the decision in *Greater Glasgow Health Board's Application* [1996] RPC 207, however, it is now widely believed that rights to patentable inventions also will first vest in an academic inventor. The continental European and American intellectual property regimes are even more complicated in the allocation of first ownership than are the rules in the United Kingdom; and some countries, such as Germany and Sweden have special rules that apply to university employees.

In all jurisdictions, these rules as to first ownership become even more complicated in situations of collaborative creation. This is particularly difficult for international collaborations because the rules as to joint creation vary more from jurisdiction to jurisdiction than do some other types of intellectual property rules. Take, for example, the rules as to copyright. In the United Kingdom, copyright will be jointly owned if the contribution of one author is not distinct from that of the others under section 10(1) of the Copyright, Designs and Patents Act 1988. In Germany, however, there is a requirement of joint creation and a requirement that each of the parts is incapable of independent exploitation, but it is acceptable that the different parts of the jointly created programme have been made at different times. (*Buchaltungsprogramm* BGH I ZR 47/91 (1995)) In the United States, U.S. Copyright Act §101 provides that the intention of the authors as to whether their work will be merged into inseparable or interdependent parts of a unitary whole will be determinative. In practice some jurisdictions, such as the United States, have a tendency to categorise works as jointly created, and therefore have very fully articulated rules as to the way in which different authors must deal with one another and with strangers; whereas other jurisdictions, such as the United Kingdom, have far less developed law as to joint authorship because they are less likely to categorise works as jointly authored. Thus a song is likely to be treated as a single copyright work in the U.S., even though one person has written its words and another its music; whereas in the U.K. the same material would be treated as independent literary and musical works. Particular problems of joint ownership arise when a party brings to a collaboration material to which either copyrights or database rights may already apply – especially when the ownership status is not clearly known *ex ante*.

---

[46] See Cornish, W. R., "Right in University Innovations: The Herchel Smith Lecture for 1991" *EIPR* 13, (1992) : pp.15-16.

At least in the United Kingdom, it is difficult to predict whether in such circumstances the courts would ever be prepared to treat the whole resulting outcome as the subject of a new proprietary right that is jointly owned, or simply regard the resulting product to be a composite of different items of protected subject matter. In most circumstances it is likely that the latter result would prevail, but this may not always be the case. For example, under section 17(3) of the Copyright and Databases Regulations 1997 it is plausible that a new, jointly owned database can be made out of two existing databases. The distinction between these two interpretations might become important when a researcher changes institutions. If the first employer has rights to the database that she creates in one place and she takes that database and adds to it in the new institution, the different analyses suggested here will determine whether the original university has any claim to rights in the whole of the resultant resource or simply in those parts of it which the researcher took with her when she left her first employment.[47]

While the rules as to the first vesting of intellectual property rights in jointly created resources are very intricate, it is important also to note that express agreement, or the assignment of rights can in most jurisdictions displace the effect of all these rules as to first ownership. The issue upon which to focus then becomes that of the terms of the contractual arrangements between the individual members of a collaborative project and their employer institutions and funding bodies. The Association of University Teachers in the United Kingdom has recently conducted a survey of policies concerning the first ownership of intellectual property amongst university employers in this country and has found significantly divergent practices.[48] This policy diverges even more widely amongst universities internationally, although the trend in recent years, not surprisingly, has been for universities to claim more and more. When outside parties get involved they too may have their own requirements as to the assignment of intellectual property rights. Because the funding arrangements will vary depending upon the type of project, it is difficult to generalise about arrangements for the allocation of rights.

It seems clear, however, that distributive norms need to be established in relation to the "proper" allocation of rights among researchers for projects of different kinds. These norms also should be flexible enough to respond to situations in which the identities of the parties involved in a collaboration, or their employers, have altered. This is again a situation in which widely shared principles and standard contractual clauses, interpreted by competent *fora*, could be used to reduce the uncertainties surrounding the establishment of collaborative projects. The difficult in addressing this need is that without the assistance of institutional mechanism for the efficient resolution of these issues at the outset of the collaborative undertaking, subsequently emerging disputes over the 'knowledge assets' that have been created may make it difficult for the latter to be used effectively, thereby defeating the purposes of e-Science.

---

[47] For a cautionary tale regarding the institutionally mobile author of a scientific database, see Appendix 5

[48] This is available in the Members Only section of the web site http://www.aut.org.uk

2.3.4 ISSUES OF LIABILITY ARISING FROM COLLABORATION

Most thinking about on-line collaboration has focused upon questions of ownership, either of the inputs to a collaboration or of the knowledge resources to which it gives rise. An equally important issue, however, is that of professional responsibility for the conduct of a project and liability for any harm that it might cause. These are linked, but not identical, questions.

The question of professional responsibility is that of who will bear the reputational or other loss associated with the discovery that research has been conducted either incompetently or unethically. The scientific community has traditional norms controlling the damage caused by incompetent work or unethical behaviour, such as the falsification of research results. More recently those norms have been given effect formally through the policies of universities, funding bodies and journals, but they also are enforced informally through the shared understandings about the subsequent treatment of malefactors by colleagues, especially those involved in the same branch of scientific enquiry. In general terms, norms of professional responsibility traditionally have reflected, and should reflect, the structure of norms for the assignment of credit for the project's scientific achievements. That is, the senior scientist with responsibility for directing the work of the project (or the facility whose resources it uses) gets most credit when it is a success, and corresponding also ought to be the person whose reputation suffers most harm if the research is found to have been poorly conducted, or the findings have been misreported.

Parallel to this question of professional responsibility, there exist legal rules that determine legal responsibility for harm that a project may cause. These legal rules may well attribute responsibility in a way that is different to that in which the traditional norms of the scientific community would do. Moreover, questions of liability might arise even for parties who do not engage in e-Science collaboration themselves, but merely provide the platform for others to do so, such as DiscoveryNet or myGrid. In England, the law of tort often will determine the question of whether, and how, harm caused by a collaborative project ought to be compensated. This will be the case whether the relevant harm has been suffered by one of the parties to a project or by a third party. In broad terms, the question will frequently become one of whether the behaviour which has given rise to a particular harm was negligent in that the harm was reasonably foreseeable and could have been prevented. For example, the rapid spread of a new virus through a computer system (causing losses of data, or damage to instruments controlled by that system) may be deemed to have been impossible to guard against, whereas the failure to update available anti-virus filters in a system firewall could be more readily construed as negligence. Negligence, and indeed the law of tort, is far from the only potential sources of liability to which the participants in collaborations may be exposed. For example, parties also might be liable because the project uses or creates material that is in breach of some intellectual property right, or because confidential data is accidentally disclosed, violating the privacy of outside parties.

Collaborative projects give rise to a host of potential questions of liability. As is the case with issues of responsibility for the legality of inputs to a research project, the issue of responsibility for the conduct of research and its outcome needs to be addressed by the parties to a collaborative project before the it is launched. The lawyer's solution is likely to be a combination of indemnities and insurance for the most obvious risks, but this again raises the various costs of negotiating an e-Science collaboration. Once more the discussion has moved into an area of concern where it would be most helpful to promote the articulation of common principles for collaboration which reflected the traditional norms for the allocation of responsibility in the scientific community, and where the development of standard form contractual clauses consistent with those principles clearly would be recognised as desirable.

The purpose of the foregoing review has not been to offer an exhaustive list of the legal issues that will arise in the context of e-Science collaborations. Rather, it is intended to point to the range and complexity of the issues that are likely to arise. It is clear that these issues cannot be addressed in a single programme of 'law reform'. Equipping the scientific community to deal with the complexities of the institutional environments in which they work demands a more subtle and responsive institutional design of the kind that is outlined by the next Part.

3. A PROPOSED INSTITUTIONAL DEVELOPMENT PROCESS FOR E-SCIENCE

To devise one or several approaches for arriving at institutional mechanisms whose establishment would generate workable specific arrangements that facilitated collaboration in e-Science is a considerable challenge. The multiple parties and jurisdictions involved in e-Science collaborations, and the need to balance conflicting interests among them, make the design of effective governance arrangements an extremely complicated and thorny problem. The costs of getting those arrangements even slightly wrong can be very high [see Appendix 6]. Furthermore, most research scientists have socially more productive things to do than spend their time thinking about how to arrive at a good set of governance mechanisms, even though it would be impossible for others to achieve that goal without thoroughly involving the affected research communities in the process. More difficult still will be the task of obtaining quickly negotiated contractual arrangements that also could be of use to facilitate collaborative activities in other spheres where Grid infrastructure is likely to be extended – including e-Learning, e-Government, e-Commerce, e-Healthcare.

It is clear that these appropriate institutional mechanisms cannot simply be put into place by legislation and that the problems created by the international nature of collaborative e-Science cannot be solved by the international harmonisation of formal legal rules. Legislation and the harmonisation of legal rules have a potentially stultifying effect on the development of appropriate institutional mechanisms in this area. When legislation is enacted and international conventions are agreed, they tend to have the effect of petrifying the norms regulating a given area of behaviour for a long period. This poses the risks that the norms which have been set may have been set at a particular point in the development of a social practice, such as on-line scientific collaboration, and may rapidly become inappropriate. The history of the international agreements on the protection of semiconductor topographies provides a good illustration of the need to proceed in a way that can avoid this potential danger.

3.1 CREATING AN APPROPRIATE INSTITUTIONAL MECHANISM: BASIC CONSIDERATIONS

Any standardisation of the norms that govern e-Science must be sufficiently flexible to undergo non-disruptive evolution with the development of mode of organising and conducting research that is, after all, still in its infancy. Moreover, the international harmonisation of legal rules is unlikely to be effective. The international harmonisation of law is a slow and frustrating process as the TRIPS negotiators have found. Harmonisation would be a particularly daunting task given the range of legal issues that might impact upon the conduct of collaborative on-line research. Moreover, the harmonisation of legal norms is only ever partially effective in achieving the goal that disputes determined under the same norms will find the same result in different courts. The history, for example, of the European Patent Convention shows that the same norms can lead to different outcomes in different courts with different interpretative traditions. Formal law reform and the harmonisation of laws do not seem to be the answer to establishing norms that can facilitate collaborative e-Science.[49]

Consequently, the nub of this Report's recommended approach to constructing appropriate institutional infrastructures for e-Science is the creation of a co-ordinating and facilitating mechanism, in the shape of a novel public agency. We envisage the establishment of an independent body that could be formally designated as 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 specific 'Grid-enabled' collaborations in science and engineering research. This advisory board would, of necessity, play a leading role in enunciating a set of fundamental principles to guide the formulation of those contractual clauses and ensure that the effects of the agreements into which they are introduced will not be inconsistent with the intent underlying those principles.

In view of the importance of finding some suitable response to the needs that have been described by the preceding discussion, some new institutional departures along these lines appears very much in order. That appropriate institutional mechanisms can make a critical contribution to the success of the United Kingdom's investments in e-Science seems beyond serious dispute. There is no point in investing in new technologies to facilitate and empower complex collaborations if at the same time we are imposing rules and regulations (including legal and institutional administrative arrangements) that excessively raise the costs

[49] This legal approach is neutral with regard to the balance struck in such agreements between providing for commercial exploitation of intellectual property rights and protecting the public domain in scientific and technical data and information. One should note, however, that it can accommodate the thrust of recent proposals that address the latter concern, most notably, J.H. Reichman and P.F. Uhlir, "A Contractually Reconstructed Research Commons for Scientific Data in a Highly Protectionist Intellectual Property Environment," forthcoming *in Law and Contemporary Problems*, 66 (Winter/Spring, 2003): pp. 315-462. Explicit recommendations favouring such an approach are advanced below, in Part 4.

of actually carrying out those sorts of collaborations. So we need non-technological governance mechanisms (of which institutional regulations and legally sanctioned contractual forms constitute important examples, but not the only ones) whose effects tend to reinforce, rather than to counteract, those of the technological infrastructure of e-Science. To initiate a deliberate movement towards this goal, this section of the report proposes the launching of an exploratory process of institutional learning. It identifies the types of expertise that should be engaged, and the forms in which it could be mobilised by the United Kingdom agencies that would fund the suggested programme of consultative research and institutional experimentation.

3.2 AN INDEPENDENT 'PUBLIC ACTOR'– THE PROPOSED MECHANISM AND SOME EXISTING MODELS

What is needed is the establishment of a new "public actor", a separate agency 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 these clauses would address a particular problem among the myriad legal issues that have been seen to arise from the formation of research collaborations, and variant solutions would be provided by the clauses developed under each topical heading. Much of this detailed work could be entrusted to specialised task force-like committees – possibly resembling the many "study committees" set up under one or another of the U.S. National Research Council Boards.

The activities of the study committees organised under the auspices of what we shall here call "the Board" would focus upon framing appropriate standard contractual clauses that could be readily assembled into a variety of alternative collaboration agreements, much in the same way that software sub-routines and modules can be assembled into functionally more comprehensive software systems that are suited for particular applied tasks. As part of its supervisory and co-ordinating role, the Board would have not only to think about the underlying principles that will be implemented through the contractual clauses of those agreements. It would also need to determine the best ways of organising the accumulation and dissemination of information and analyses concerning the actual formulations and manner of implementation of contractual agreements. These principles for the establishment of e-Science collaborations and model contractual clauses could then be put into effect in individual cases by universities and research bodies.

The part of the Board's activities which involved promoting principles of best practice in establishing institutional arrangements for e-Science collaborations might find a parallel in the work of the Basel Committee on Banking Supervision.[50] Under the Basel Accord the Governors of the Central Banks of thirteen countries develop common principles for banking supervision. The Basel Committee on Banking Supervision does not make laws of any kind, rather it builds consensus between important actors in the international banking

[50] The Basel Committee is briefly considered as a model in the context of e-Commerce regulation in Casey, D. and Magenau, J, "A Hybrid model of Self-Regulation and Governmental Regulation of Electronic Commerce" (2002) 19 *Santa Clara Computer & High Technology Law Journal* 1 at 27

community and these standards are given appropriate effect by relevant actors in local contexts. The experience of the banking community has been that this approach has the advantage of flexibility and that principles develop slowly rather than being imposed at what might be an inappropriate stage in their articulation. In time, it might be hoped that an international body for the development of collaborative research principles might be established, similar to the Basel Committee. However, the establishment of a national body with such a task would be an important step in establishing socially desirable rules for the organisation of e-Science at a national level. Indeed, it is likely that a United Kingdom body charged with the functions of the Board could set a lead in the organisation of e-Science collaborations and that its principles and contractual clauses would be widely adopted even by those not a part, either direct or indirect, of the United Kingdom e-Science network.

The part of the Board's work that involved the development of standard contractual clauses might also find a parallel in the work of the Basel Committee and, with one important reservation, in that of the Grain and Feed Association ("GAFTA"). The Basel Committee makes recommendations for contractual clauses in certain areas of banking practice as a means of ensuring that the principles it articulates are given appropriate and certain effect. Central amongst the clauses that the Board might be expected to suggest would be those reflecting a consensus as to the appropriate *fora* for the resolution of disputes under e-Science collaboration agreements, particularly those involving parties from different jurisdictions. A model of how such choice of forum clauses might operate can be found in the standard form agreements established and maintained by GAFTA. GAFTA has 80 different standard contracts under which more than 80 million tonnes of the world's trade in cereals and 70% of trade in animal feeds moves annually. These contracts contain arbitration clauses that allow parties to make use of the GAFTA Dispute Resolution Service. This service provides for the speedy and final resolution of disputes in an expert forum that is cheaper and quicker than traditional legal systems; further, it has the advantage of being outside the legal system of any of the parties to the international collaboration. The process involves a possible appeal to a GAFTA appeal board. Awards given under the arbitration system are enforced either informally through publication of the fact that a party has failed to comply with an agreement or, as a last resource, through the court system of the jurisdiction in which the party against whom the award has been made under the New York Convention on the Recognition and Enforcement of Foreign Arbitral Awards 1958. In the event that arbitration fails for some reason, disputes in international commodities contracts of this type are by agreement usually referred to the courts of a small number of jurisdictions, in particular those of England and New York. The establishment of standard arrangements for the resolution of disputes in transnational e-Science collaborations would greatly reduce the uncertainty surrounding such projects.[51] It is important at this point to emphasise why the approach recommended is one of standard contractual clauses and principles for the development of collaboration agreements rather than standard form agreements of the type established and maintained by GAFTA. This is for three reasons. Firstly, the potential contexts of e-Science collaborations are very various and the formulation of standard form agreements would be an

[51] Yet another parallel might be found in the highly effective matrix of formal and informal norms that regulate the international trade in cotton under the rules of the Liverpool Cotton Association, see Bernstein, L., "Private Commercial Law in the Cotton Industry: creating Cooperation Through Rules, Norms and Institutions", (2001) 99 *Michigan Law Review* 1724

almost impossible task. Even GAFTA must maintain 80 agreements in relation to the relatively routine transactions that make up the international trade in grain and animal feeds. The contexts in which standard form contracts tend to be most successful are those such as domestic conveyancing in which only a very limited range of issues is likely to arise.

Secondly, the introduction of standard form agreements, like legislation and the harmonisation of laws may have the danger of petrifying norms at an inappropriate stage in their articulation. We are at the beginning of a new era of scientific collaboration, based upon high bandwidth telecommunications and grid-enabled computing, and to put standard form agreements in place at this moment may entail the danger of ossifying the development not only of appropriate norms for e-science, but also of inhibiting flexibility in the elaboration of the enabling technological infrastructure. The absence of extensive experience (and hence of the weight of precedents) concerning arrangements in the new environment creates an opportunity to exercise greater leverage over the future evolution by setting standards early and firmly.[52] Yet, the very same conditions make it difficult to gauge what the new standards should be.

The fact that principles, more obviously than standard form agreements, must be developed over time will serve to emphasise to the community of lawyers and administrators in academic institutions and funding bodies that the agreements they produce need to reflect changing scientific practice.

Thirdly, the blanket use of standard form agreements is fraught with legal risk. This is because such agreements can come to be used without appropriate legal advice and institutional support. Indeed, this is more likely in the context of e-Science than it is in context of international trade. Three particular legal problems tend to emerge when standard form agreements are used without appropriate professional advice. In the first place, it is obvious that a standard form agreement may not be appropriately adapted to the project which it is intended to govern. Second, in situations in which the parties simply adopt a standard form without addressing their minds to the appropriateness of its terms, the courts and legislators in most jurisdictions have come to regard them as inherently suspect. This is because contractual liability is generally, though not universally, regarded as essentially consensual. In the case of a standard form agreement, it is sometimes difficult to see how at least one of the parties can be taken to have consented to all its terms. The issue of how consent can be given to standard form agreements on-line has been thought to give rise to particular problems that have attracted both legislative and academic attention around the world. Third, when different standard forms are used by the parties entering a particular

[52] This is the non-technical form of the classic problem of "anticipatory standard-setting" in regard to network interoperability standards, sometimes referred to as "the Blind Giant's problem." See P. A. David, "Some New Standards for the Economics of Standardization in the Information Age," ch. 8 in P. Dasgupta and P. L. Stoneman, eds., *Economics and Technological Performance*, London: Cambridge University Press, 1987, pp. 206-39; P. A. David, "Standardization Policies for Network Technologies: The Flux Between Freedom and Order Revisited," ch. 3 in *Standards, Innovation and Competitiveness: The Political Economy of Standards in Natural and Technological Environments*, (R. Hawkins, R. Mansell and J. Skea, eds.), London: E. Elgar, 1995, pp.15-35; P. A. David and M. Shurmer, "Formal Standards-Setting for Global Telecommunication and Information Services," *Telecommunications Policy*, vol. 20 (10), December 1996, pp. 789-815.

transaction, each can purport to rely upon her own usual form; it then may be difficult to determine which, if indeed either of the two had ultimately been adopted as the basis of the agreement. This is sometimes referred to as the "battle of forms" and will become a problem as "form shopping" encourages a proliferation of standard form agreements.

It is therefore suggested that the approach of developing standard contractual clauses covering most of the issues likely to arise in establishing and conducting an e-Science project, together with principles for their use to be observed by lawyers and administrators in academic institutions and funding bodies is a better way of developing this area of law than the development of even a stable of alternative standard form contracts. What is called for, in effect, is the institutionalisation of an iterative, adaptive procedure for developing 'meta-agreements' – contractual analogues of 'meta-standards' in the technological domain. This way of navigating the legal thicket can be likened to a technological meta-standards approach based upon setting "performance standards" rather than technical specifications, which has long been advocated as most appropriate for anticipatory technical standards development purposes. (In the latter, almost as a matter of definition, the relevant underlying technology and the markets are evolving rapidly, which means that very substantial uncertainty surrounds the capabilities of new systems and the attributes that their users will most value)?

Setting performance standards, however, implies that criteria of acceptability must be defined in the relevant dimensions, and procedures for compliance testing and certification need to be established. Although those sometimes are held to be impracticably costly requirements to impose in the technological sphere, an analogous process would appear to be considerably less problematic (indeed, quite natural) when applied in the domain of contractual clauses. The apparatus already exists for the effects of the latter to be examined by reference to expert opinion about pertinent legal doctrine, the empirical experience of contractual negotiations, and the law courts' rulings in a variety of jurisdictions.

In summary, therefore, three key conditions seem necessary for the Board to be effective in its undertakings. First it is necessary that it delineate readily intelligible transcendent principles for agreements governing collaborative e-Science projects – across as wide a variety of academic research domains as is possible. Second, to be workable and substantially self-enforcing, these guiding principles also must reflect the essential values of the scientific communities whose collaborative work is to be facilitated by institutional instruments (and the technological infrastructures of Grid-enabled computing) are meant to facilitate. Third, in order to yield contractual agreements that are flexible enough to accommodate distinctive features of the relevant research community norms, as well as of the requirements of the particular science that is being planned in each case, the Board should eschew trying to write "model contract" or "standard form" agreements, and focus instead upon the development of standard contractual clauses.

3.3 ORGANISATIONAL FORM AND COMPOSITION OF "THE BOARD"

The precise form of the responsible independent body thus envisaged can best be left for discussion and subsequent determination, once there is substantial agreement about its goals and operating procedures. One approach deserving serious consideration would be to have the proposed advisory board on collaboration agreements would independently from, but in close liaison with bodies representing the pertinent administrative agencies – including the office of the Director of the Research Councils, the Higher Education Funding Council for England (HEFCE), the Scottish Funding Councils for Further and Higher Education, Universities U.K. (UUK) and the Standing Committee of Principals (SCOP).

Alternatively, the establishment of the proposed body (the ABCA) could be approached more gradually, by organising a continuing "Working Party on Agreements" which reported regularly to the Joint Information Systems Committee (JISC-WPA). Once a body of contractual clauses and information about the circumstances in which they proved most suitable had been developed, the Working Party might be dissolved and replaced by the Advisory Board on Collaborations envisaged above. Eventually, as its recommended contractual clauses came to be more wide used for publicly funded collaborative projects in e-Science, the ABCA might eventually be transformed into a completely free-standing and permanent official government "Commission on Institutional Infrastructures for Collaborations in e-Science" (CIIC:e-Science). The colon in the latter looks ahead to the formation of a succession of such commissions, each having analogous responsibilities for assisting the formation of collaborative agreements in a different sphere, e.g., e-Commerce.

The composition of the Board's appointed study committees, like the membership of that body itself, would have to be multidisciplinary. The working groups particularly need to draw upon technical expertise regarding the hardware and software infrastructures supporting collaborative e-Science, and the complex systems of resource allocation of which scientific and technological research and teaching is a part. Moreover, through the guidelines provided by the Board as part of the committee's remit, it must be attentive to the larger societal goals and values that HEIs and public research institutes need to serve. The e-Science Board's members also need to stay informed and especially foresighted regarding potential opportunities and challenges that are likely emerge when the technological infrastructures created for e-Science are opened for the development of applications supporting e-Learning, e-Commerce, e-Government, and so on.

In embarking on what must be an evolutionary programme of research, practical experimentation, assessment and redesign, it will be vital to enlist the assistance of lawyers (both academic and practising) who are well versed in the array of legal issues identified by this report – and with their treatment not only in common law, but under other legal systems. No less important for this work will be the recruiting of a core cadre of social scientists who have been involved in science and technology policy studies in the United Kingdom and other leading research countries. They should have particular expertise in the social and economic organization of academic research communities, as well as with corporate research management practices and policies regarding intellectual property rights. They should be complemented by rotating groups of experienced practitioners drawn from two constituencies: senior scientists familiar with the variety of research communities that are likely to take the lead in moving their activities onto e-Science platforms, and representatives of university administrations who are engaged in solving the practical problems posed by research collaborations.

With those needs in mind, and equally with the thought that the analysis and proposals presented here should be submitted for review by some of the same experts that eventually would be recruited to launch this initiative, Appendix 7 of the report presents a preliminary roster of scholars and practitioners whose opinions it would be very useful to solicit.

3.4 REQUIREMENTS FOR INFORMATION ON THE EXPERIENCES OF e-SCIENCE COLLABORATIONS

The development of a clearer picture of the institutional context of collaborative e-Science therefore can be viewed as one of the derivative implications of the recommendations advanced by this Report. Evidently, this entire proposed programme of legal mechanism design will need to be informed by systematic data collection about the informal ways in which disputes among collaborating researchers, and among institutional partners too, actually may be resolved before the parties enter litigation. Corresponding research is necessary about the circumstances in which scientific and technological projects are most frequently delayed, or irremediably disrupted by conflicts involving contractual maters. A multidisciplinary inquiry into the role played by institutional infrastructure factors in the experience of successful (and unsuccessful) e-Science collaborations should be initiated in conjunction with the assessment work that the Board should plan to carry out in regard to the effects of its own work.

The discussion (in section 2.1, above) of the current institutional context of scientific collaboration brought out the comparative paucity of concrete empirical knowledge concerning the specifics of current individual researchers' experiences with informal governance arrangements for scientific research collaborations, and a parallel lack of systematic information about the ways that formal legal rules are being utilised by co-operating institutions. As this kind of information that could be gathered as a part of the process recommended by this report it is appropriate to take fuller note of this requirement; and equally of the opportunities that would thereby be opened for social science and legal studies of the changes occurring in the social and economic organisation of contemporary science and engineering research.

First, more information is needed about current practices in planning the institutional structures of e-Science collaborations. In particular, it is important to know the extent to which the working scientists themselves have an input into planning the institutional aspects of those projects, rather than simply specifying the technical requirements. If they only, or even principally, address the technical questions, there is a risk that planning will be on the basis of a "perfect team" assumption – with issues of imperfect team behaviour therefore being left unaddressed. Further, if the actual collaborating agents have little or no role in planning the institutional arrangements, there is the real danger that those arrangements will reflect choices that are at odds with the culture of collaborative research and inimical to its success.

A further question regarding the planning phase of collaborative work is the issue of how, and from whom, the parties to collaborations, particularly those in publicly funded institutions, receive legal advice. Work needs to be done to discover the extent to which this advice is enabling, and the ways in which it creates possibilities for the establishment of individual collaborative projects, rather than simply imposes costs upon them. Empirical studies should be directed to document the extent to which existing collaborations can and do support their institutional arrangements by implementing technical controls upon access, modification and reproduction of data and information. It is as yet quite unclear the extent to which parties to scientific collaborations in academic institutions are using technical measures to monitor and enforce compliance of member with the rules of participation in a project, although employment of such devices by commercial database providers is well documented. Finally, under this head, it is important to learn whether the rules for the administration of a particular collaboration are simply agreed rules of behaviour, or whether they are built into the very way in which the project is structured.

Second, more needs to be known about the way in which the participants in e-Science collaborations currently use the legal rules that touch upon their projects. This involves a number of issues. For example, when a dispute between collaborators arises, in what way do collaborators invoke the traditional informal norms of their particular scientific community and in what ways do they invoke formal legal rules? Evidence from contractual disputes in the commercial world suggests that formal rules are invoked only rarely, and when they are, it occurs at a point where the relationship among the parties already has become very strained.[53] It is important to know whether this pattern also holds true in the world of collaborative e-Science. Similarly, there is a question regarding the extent to which collaborators allow their relationships to develop as they proceed and the formality or informality with which they vary the agreements supporting their dealings with one another. The experience of commercial lawyers is that even parties with access to sophisticated legal advice can allow a significant divergence to emerge between the formal legal basis of their relationship and the rules upon which it actually operates. This can have unfortunate legal consequences, but it would be surprising if it were not the case in collaborative science.

More generally, there is a need to learn much more than presently is known about the interplay among the technical, social and institutional constraints on e-Science. For one thing, such studies would be of great help in validating a number of the assumptions underpinning the analysis and recommendations of this Report, and could be expected to contribute significantly to the future design of institutional arrangements that would more effectively promote 'e-collaborations.' The sort of advisory body whose creation has been recommended here could be charged with the responsibility not only to propose effective legal devices and organisational procedures to facilitate collaborative e-Science, but also to commission on-going social science research that should underlie its work.

---

[53] And quite often, not even then. In many instances formal legal action is initiated only when one of the parties reaches a 'threat point' imposed by the potential for third party action (e.g. shareholder lawsuits or bankruptcy proceedings).

4. Conclusion: IPR Policy Reforms and the Way Forward

In setting out the scope of the challenge of providing workable institutional conditions for productive e-Science, and in proposing a particular approach towards that goal, the preceding parts of this report have avoided explicit discussion of the underlying policy positions that might be reflected in the 'working principles' adopted the proposed Advisory Board. It has been noted, of course, that these would have to achieve some balance between, on the one side, the purposes of the scientists and engineers engaged in the research, and, on the other side, the corporate concerns of the institutions in which they were working. That the public and private non-profit funding agencies also will bring a further array of policy goals to bear – both directly and indirectly – on the determination of the Board's operating principles is to be expected. In that context, it is only realistic to acknowledge that policy questions about intellectual property rights protections are likely to emerge among those that prove to be most problematic for the participants in research collaborations that receive substantial non-commercial (public sector and charitable foundation) support. This Part therefore offers a some concluding observations on the issues raised in that connection, and the way in which they might best be resolved through the agency of the Advisory Board.

By emphasising the need to devise a new, flexible process for the 'bottom up' construction of institutional arrangements that will promote and support collaborative e-Science research, priority has already been accorded here to the public sector goal of rapidly and efficiently advancing scientific and technological knowledge. This position may be contrasted with according priority to the goal of capturing 'private' economic benefits from possession of new additions to the body of knowledge. The approach accordingly is to be favoured over efforts to codify existing institutional agreements for publicly funded research in standard form contractual agreements -- especially those which would simply carry over into the academic institutional sphere intellectual property rights provisions modelled on the legal agreements governing commercial R&D partnerships, research joint ventures and similar consortium arrangements. The arguments for the latter position thus goes well beyond the point that standard form agreements may or may not strike the right balance between access and incentive in certain types of scientific endeavour. A still much more serious problem lies in the present imbalance that has appeared within the intellectual property regime -- between the extent and strength of the protection being accorded to holders of private (monopoly) rights, and societal protection of the public domain or a protected 'research commons' in scientific and technical data and information.

4.1 PRESERVING THE EFFECTIVENESS OF 'OPEN SCIENCE'

There is today a growing consensus among informed observers that the dominant trend of the past two decades towards broadening and strengthening of legal protections for intellectual property rights, and the privatising of the sources of scientific and technical data and information, has now gone too far. This assessment pertains to the situation existing among the handful of rich, economically advanced countries that do most of the world's organised science and engineering research. It is by no means confined to concerns that also have been expressed about the adverse impacts of the global IPR regime upon the developing economies' access to new scientific knowledge and knowledge-intensive goods and services.[54]

In Britain, the European Union and the U.S., several influential organisations have issued statements calling for a re-consideration of the place of intellectual property rights in contemporary science and technology. They point to a number of unintended, yet nonetheless undesirable impacts of current intellectual property policies upon the effective conduct publicly funded, academic research collaborations in science and engineering.[55] Especially notable in this regard are the Royal Society's report entitled '*Keeping Science Open': The Effects of Intellectual Property Policy on the Conduct of Science* (April 2003), and the Report of the EC Research Directorate General's European Research Area Expert Group on *Strategic Use and Adaptation of Intellectual Property Rights Systems in Information and Communications Technologies-based Research* (March 2003). These same concerns also permeate the U.S. National Academy of Sciences recent publication: *The Role of the Public Domain in Scientific and Technical Data and Information: Proceedings of a Symposium* (September 2003).

None of these assessments are one-sided; all acknowledge that the protection of intellectual property rights can stimulate useful discoveries and inventions by protecting creative work and investments in costly research and development efforts. Further, they recognise that the prospective award of monopoly rights in the exploitation of new ideas can elicit the disclosure of discoveries that might otherwise be kept secret, and on that account may contribute to stimulating further advances in useful knowledge. Nevertheless, they concur in expressing serious concerns about the potential adverse impacts on the culture and practice of academic open science of the legal innovations, and the changes of institutional policy among the PROs in response to the emphasis that Western governments during the past two decades have placed upon near-term goals of 'wealth creation' through research. They deem it necessary to emphasise that "high quality research is the gateway both to advances in knowledge and the wealth creation based on science;" that the competitive pursuit of patent rights creates incentives for secrecy that generally will be inimical to the rapid advancement of knowledge; that intellectual property rights are a basis for the imposition of costs, and the threat of costs which "can hinder the free flow of ideas and information upon which science thrives." [56]

---

[54] On the latter, however, see *Integrating Intellectual Property Rights and Development Policy*, The Report of the Commission on Intellectual Property Rights (2nd Edition), London, 2002, esp. Chs. 1, 5-7. [Available at: http://www.iprcommission.org/papers/word/final_report/reportwordfinal.doc.]

[55] The same theme is emerging more strongly in the recent writings of 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; Arti K. Rais and Rebecca S. Eisenberg, "Bayh-Dole Reform and the Progress of Biomedicine," Law and Contemporary Problems, 66(1), 2003 [available at: http://ssrn.com/abstract – id=348343]; P. A. David, "Can 'Open Science' be Protected from the Evolving Regime of Intellectual Property Protections," *Journal of Institutional and Theoretical Economics*, forthcoming in Fall 2003. (Stanford Institute for Economic Policy Research, Discussion Paper 02-29. Stanford University, (July), 2003 [available at: http://siepr.stanford.edu/papers/papersauth_D-H.html].)

[56] 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]: p. v.

Consequently, the recent report of the Royal Society recommends, *inter alia*, the clarification and harmonising of the presently ambiguous exemptions from infringement of the patent laws permitted to scientific work under the headings "private and non-commercial" and "experimental" use.[57] The same document explicitly calls for reversal of the recent introduction in U.K. law – specifically in the statutes implementing the EU Database Directive (1996) and the Copyright Directive (2001) -- of narrow and ill-defined limitations on the 'fair dealing' exceptions provided for research; in a departure from traditional practice, these exceptions are confined to research that has "non-commercial purposes."[58] Specifically, the Royal Society's report notes that in the law enacting the EU Database Directive in the U.K. (on 1 January 1998), the fair dealing exception for research (and education) permits only extraction and not re-utilisation of the protected contents. Given the failure of the exiting statute to accommodate the needs of the scientific community in regard to digital databases – a 'tool' that has acquired increasing importance in numerous research contexts, the Royal Society's report recommends that the laws be revised, so that even research that might be regarded as having some commercial value would be exempted from infringement of the database right and copyright.

Salutary as we believe these and related recommendations to be, their immediate practical force is mitigated by the fact that effecting significant legal reform is more often than not a complex, highly politicised and uncertain undertaking. This is likely to be true particularly in the area of intellectual property rights, where determined opposition must be expected from entrenched business firms whose strategies are predicated upon perpetuation of existing legal arrangements, and where the international repercussions and concerns for programmes of "harmonisation" – not to mention the conflicts between the interests of the industrially advanced and the developing nations of the world -- are likely to frustrate rapid progress.

The usefulness of the recommendations contained in those reports is further circumscribed by their own recognition that the effects of the intellectual property rights regime on the progress of particular fields of research, and on specific projects, may be quite different. Revisions of provisions in the intellectual property rights statutes cannot readily accommodate the effects of contextual variations without become inordinately complicated. Moreover, if differential rules are introduced that are perceived to create advantages for some research areas, or types of collaborations, this may induce efforts to reconfigure projects – or at least to configure the outward appearance in order to exploit such advantages. Lastly, it is relevant to bear in mind that the foregoing proposals for legal reform at best would address the entire problem; intellectual property law represents only one among the regulatory thickets on the institutional obstacle course and through which researchers attempting to advance a particular collaborative project need to find a feasible path.

---

[57] See, The Royal Society, *Keeping Science Open,* April 2003, esp., sects. 3.21 (patents); 4.11-4.20 copyright; 5.5 databases. .

[58] In U.S. legal parlance, such exceptions (to copyright law) are referred to under the heading "fair use".

4.2 TIMELY ACTION THROUGH CONTRACTUALLY CONSTRUCTED COLLABORATION AGREEMENTS

Consequently, it is our view that those who seek to advance collaborative e-Science in the here and now will be more effective if their attention and efforts focus, not on the codified details of intellectual property law, but upon the specific institutional structures that can be created contractually to facilitate particular collaborative structures that are most suitable to for the work of specific scientific projects. This approach, based upon the development of standard contractual clauses, is congruent with the conclusion of legal scholars in the U.S. who advocate "contractually reconstructing the research commons" as a way to mitigate the adverse effects of "a highly projectionist intellectual property environment."[59] It also bears some kinship with recent initiatives to encourage the creative reuse of copyright protected material by providing a variety of readily implemented contractual alternatives to the full set of rights available to copyright owners under prevailing statutes.[60] In the present context, contractual construction of an e-Science research commons does not require that collaboration agreements must be created *de novo* for every occasion; the principles of modularization and component standardisation may be applied as usefully in this sphere as they are in the art of software engineering itself.

A concrete illustration may be given of the way that contractual clauses can be used in conjunction with the licensing terms of copyright protected software to solve difficult problems involving the balancing of different public policy interests. In the immediate context of the U.K. e-Science Programme it is germane to point out how contracts can work with existing law protecting intellectual property rights to accommodate seeming conflicting objectives of public funding agencies (and PROs), specifically in regard to the licensing of middleware and software applications tools that are being developed under the EPSRC e-Science Core Programme.

Much of the near-term rationale underlying investment in the development of that part of the technical 'cyberinfrastructure' rests on the contention that if access to data, information and facilities can be made easier, and less costly, there will be very substantial efficiency gains from the collaborative search for scientific knowledge. The economic case for reinforcing the 'open science' mode of collaboration is especially compelling in this area, given the modularity of well-engineered software, and the possibilities of generating recombinant novelty through re-use of already developed sub-routines. This purpose would be served by mandating the distribution of publicly funded code as free and open source software, making use of the terms of the already widely used GNU General Public License

---

[59] See in particular, J. H. Reichman and P. F. Uhlir, "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.).

[60] A related contractual approach, utilising a menu of machine-readable copyright licenses, has been implemented by *Creative Commons*, a non-profit organisation developed on the intiative of Professor Lawrence Lessig (Stanford University, which describes itself as "dedicated to the creative reuse of copyrighted material". For description of collaborative projects and organisations that are cooperating with Creative Commons, see http://creativecommons.org/learn/collaborators.

(GNU GPL).

On the other hand, both in government policy circles, among university administrators, and individual members of the academic research community, it is held to be highly desirable that the knowledge and information-goods generated by publicly funded research be available as a basis for private sector investment in its further elaboration and commercial distribution. To attain the latter goal by means of permitting PROs and their employees to exercise proprietary rights in software developed with substantial public support, however, would conflict with the rationale for public funding of this kind of R&D: the grant of copyright monopolies (and of software patents in some jurisdictions) if proved effective, would raise the economic costs of utilising the information-goods in question. This means it would raise the their cost both as final goods, and as inputs for the production of further software innovations, including those by the private sector.

A pragmatic solution to this policy dilemma may be available. The proposal has been advanced to allow both goals to the goals to be served concurrently in some degree, by means of private contracts permitting modifications and further developments based upon publicly funded code released as "free and open source software."[61] This ingenious use of features of the GNU Public License would call for public funding agencies in the first instance to mandate that all software created by their research projects be released under GNU General Public License (requiring distribution of the source codes if along with the machine code, among other terms of this standard license); and second, to allow the original copyright holder of such programs assign the copyright to some non-profit foundation or other entity that would oversee the granting of private contracts allowing modifications and elaboration upon the GPL'd code for ultimate commercial distribution.

The details and interpretation of the legal aspects of the GNU GPL that would permit this form of "dual licensing" are interesting, but they need not be entered into here. This proposal has been injected into the present discussion primarily to illustrate the point that the development of contractual clauses (in this instance for commercial exploitation of the original extensions based upon GPL'd code) can be a useful device in the hands of collaborative public science communities, permitting collaborative projects between PRO's and private sector firms. In addition, this concrete illustration serves to underscore the observation that the practical implementation of a novel proposal of this kind nonetheless would need to address many of the same institutional issues that have been examined by the preceding parts of this Report. Consider just the following two sets of questions about the non-trivial practical details of implementing a program of copyleft and dual licensing for publicly funded software:

a) Should a special foundation be created to hold the IP rights on publicly funded software, and should each funding agency have their own such foundation, or they designate an independent national or international entity to which the original

---

[61] See R. A. Ghosh, "Copyleft and dual licensing for publicly funded software development," Draft version (1.0), MERIT/Institute of Infonomics, University of Maastricht, July 2003.

copyright author (and his/her host institution) would be required ultimately to assign the ownership rights? Is an international or transnational foundation politically feasible? What about charities, such as the Wellcome Trust, or the Rockefeller Institutes -- would they too have to form special foundations to fulfill this function, or would they be expected to voluntarily require assignment of rights to some pre-existing foundation(s)? How would such foundations be funded -- by retaining a small proportion of the royalties garnered under from the private contracts that they issued? Would that create an institutional motivation to "market" such permission on revenue-maximising terms, and if it did would the consequences be desirable?

b) Are the foundation(s) to which publicly funded software would be assigned also to be made responsible for negotiating the equivalent of "cross-licensing arrangements" affecting these private contractual permissions? What would be the mechanism for resolving negotiating conflicts among individual foundations attached to different funding agencies and countries? What would be the sources of such conflicts, and could those be suppressed by centralising at least the national assignment of GNU GPL licenses to a single entity? To the extent that the "dual licenses" for private exploitation of GPL code are really contracts, how would one deal with issues of harmonisation among the jurisdictions in which such contract can be enforced?[62] How can those issues be prevented from obstructing contracting and cross-licensing agreements among those contract holders?

Hence, it may be seen that one of the functions that the assignee foundations could perform would be that of establishing uniform contractual formulae, including jurisdiction-setting rules for the private contractual agreements. Alternatively, the ABCA or the pre-cursor Working Party (both of which were proposed in Part 3 of this Report) could be tasked to provide suitable standard contractual clauses for this purpose. Further task of some importance would be to achieve some degree of harmonisation of the rules imposed by the funding agencies upon the initial holders of copyrights. Would one want the same search algorithm, or encryption programme to be differently 'regulated' in regard to its commercial exploitation, simply because it had been developed under a bioinformatics project supported by the Wellcome Trust, rather than by an EPSRC middleware development project? Probably not, but, as has already been noted, a policy of "one size must fit all" would run the risk of removing flexibility and accommodation to the different realities of commercial exploitation opportunities, as well as the characteristics of different classes of software. This vexed issue of the appropriate degree of standardization of the contractual rules for 'dual licensing' is one that certainly could be referred to the proposed Advisory Board, as one more problem to be addressed in the course of its work -- along with others of a kindred nature.

---

[62] This is not only a matter of differences among national legal jurisdictions, for, in the U.S, even though there is a uniform federal commercial code contract law is a matter for the State courts. Copyright, and hence 'copyleft' licenses avoid this problem because IPR is protected under Federal statutes.

From the foregoing it may be concluded that the services of an expert advisory body will be needed to deal with the many interrelated issues that arise in just this one area connection, even were it to be pre-determined that middleware developed by the e-Science Core Program's projects would all be released under the GNU General Public License. The precise form of the novel institutional that has been envisaged here as an Advisory Board on Collaboration Agreements is not what really matters, although features that assured its independence would be essential. What is required to meet the challenges of adaptive design of an appropriate institutional infrastructure, above all, is a guiding, architectural vision, and sufficient resources to mobilise and maintain the necessary technical expertise: first, to select and standardise the contractual components, and then to assess the performance of the various collaboration agreement that they have been used to construct. An entity able to sustain and assure continuity to those two, inter-twinning tasks ultimately could exert a powerful influence towards realising the global promise of advancing knowledge and improving human welfare through e-Science.

ADDENDA

Contents

Note: The material presented in Appendices 1-4 was prepared by Matthijs den Besten (OII Research Officer for this project) and Paul A. David.

Appendix 1

Computer-mediated telecommunication network supports for collaborative research activities: concepts and definitions

1.1 Concepts and terms: a glossary for the uninitiated

1.1.1 The GRID

The Internet as we know it is far from perfect – connections are unreliable, bandwidth is scarce, and computers are vulnerable to intrusion, to name a few. The vision of the GRID is to develop a computer infrastructure that does not suffer from these flaws. Akin to the electricity GRID, users of the GRID should be able to plug in their devices anywhere and make use of the GRID's computing power, shared data, and shared instruments without being forced to worry about the underlying architecture.

The GRID is not just another application, however. 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.

The vision of seamless access to ubiquitous computing sketched here is a near utopia. 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, etc. It is these issues with which the field of GRID computing is concerned.

1.1.2 Web Services and the GRID

Web services is the container term for efforts by 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 conventions spelled out by the Globus projects at the Argonne National Laboratory in Illinois. The interfaces and conventions should make it possible for software to discover which services are provided and/or compose them on the fly.

1.1.3 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 organised in a decentralised fashion (as opposed to the prevalent client-server model). Famous examples of peer-to-peer applications are Napster, Seti@home, and ... the Internet itself, for it is a non-hierarchical, connectionless telecommunications system. While it is likely that the GRID will follow a peer-to-peer architecture, it is not impossible and perhaps easier to implement the GRID as a central server that keeps track of what its clients are doing.

1.1.4 Collaboratories

The Internet may not be perfect, yet it still is a very powerful tool to provide access to and promote collaboration among laboratories scattered across the world. In the early 1990s NSF funded a number of collaboratory projects to explore this idea. The objectives of the projects ranged from the provision of remote access to expensive experimental equipment to collaborative analysis of data and community building.

With the exception of a pair of collaboratory projects designed to permit students in high schools to have remote access to sophisticated observational facilities (e.g, Bugscope, and Chickscope), support for the real time colloborative interactions among individuals forming distributed teams and using distributed facilities, was the hall-mark of these pioneering implementations of the collaboratory.

GRID research and collaboratory research complement each other in that the first will eventually provide the architecture on which the latter can run its applications. Yet, all GRID applications are not collaboratories, and *vice versa*.

1.1.5 Cyberinfrastructure

According to the recently released report of the NSF Blue Ribbon Advisory Panel on Cyberinfrastructure, a new age has dawned in scientific and engineering research. That is, if NSF can provide $1 billion of funding per year for a large-scale, interagency, and internationally co-ordinated Advanced Cyberinfrastructure Program. Ideas and argumentation brought forward in the Advisory Panel's report are reminiscent of GRID and Collaboratory. Whether there is a conceptual content in this term that extends beyond the latter concepts and their implementations is a matter that remains to be seen.

1.1.6 e-Science

e-Science is a relatively recent concept that has come into use chiefly in the United Kingdom. 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.

1.2 A taxonomy of collaborative research activities on the Net

Traditional laboratories are located at a specific place and are only accessible to people who happen to be there. Virtual laboratories extend traditional laboratories and make them accessible by moving parts of the labs into cyberspace. An obvious way to differentiate among the forms of the “virtual laboratory” is to ask the question: What is this organisation/facility meant to do, i.e., what is the primary purpose of the activities are thus enabled? [See, e.g., Allen, Gabrielle, Goodale, Tom, Russell, Michael, Seidel, Edward & Shalf, John, 'Classifying and Enabling Grid Applications', *Concurrency – Practice and Experience*, vol. 1, no. 7 (2000).]

The taxonomy of Figure 1 attempts to do just that. It defines four prime categories and divides each of these categories up into two subcategories. The prime distinction is between community-centric, data-centric, computation-centric, and interaction-centric applications.

A **community-centric** virtual lab seeks to bring researchers together for collaboration. If the lab values synchronous interaction, it will provide chat rooms and video conferencing tools to allow the researchers to collaborate in real time. A virtual lab can also foster communities via tools such as mailing-lists and bulletin boards that do not require simultaneity. It will then be classified as asynchronous community-centric.

A **data-centric** virtual lab concerns itself with storage, management and mining of data collected from sensors, experiments, and researchers. A distinction can be made between labs whose sole aim is to make existing data accessible to researchers and labs that allow for annotation and modification of their data by accredited researchers.

A **computation-centric** virtual lab seeks to provide high performance computing. One way to do this is to ask each of the participants to share their computer's computing power (P2P). Another way is to build a super computer or cluster and make its computing power available to clients (server).

An **interaction-centric** virtual lab is an application which requires real-time user interaction where responsiveness is of great concern. Decision- making, visualisation, and control of instruments are examples of applications of this kind. The interaction can be between two participants/sites or between multiple participants/sites.

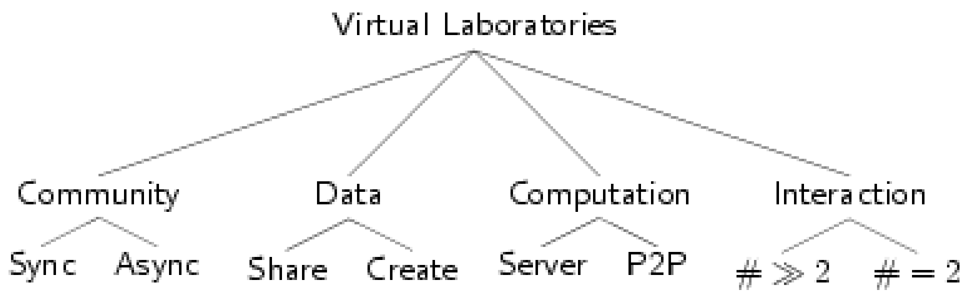


Figure 1: Taxonomy of on-line collaborative activities

1.3 Filling-in the boxes: the distribution of the United Kingdom e-Science pilot projects

1.3.1 Classifying the e-Science Core Programme's Pilot Projects

Figure 2 is a classification of e-Science pilot projects following the teleological taxonomic principles summarised by Figure 1. [See below, Appendix section 1.5 for specifics of the projects.] Although every branch of the tree is populated, the data-centric branch is by far the most populous. This may be a recognition of the importance of data management and exploitation in modern science, or it may indicate the absence of collaborative research goals beyond those reachable via linkage of databases. There is another difference between the e-Science pilots and a previous generation of U.S. "collaboratory" projects (see below, section 1.3.2) that is not immediately visible from Figure 2: many more among the collaboratory projects were concerned with access to data collection instruments: microscopes, weather balloons, and the like. E-science pilot projects, by contrast, tend to stay within the virtual world. Lastly, it should be noted that the lack of community-centric e-Science pilot projects does not imply that community building has faded altogether from the radar, but it no long appears to be so central an objective for the developers of "virtual laboratory infrastructures".

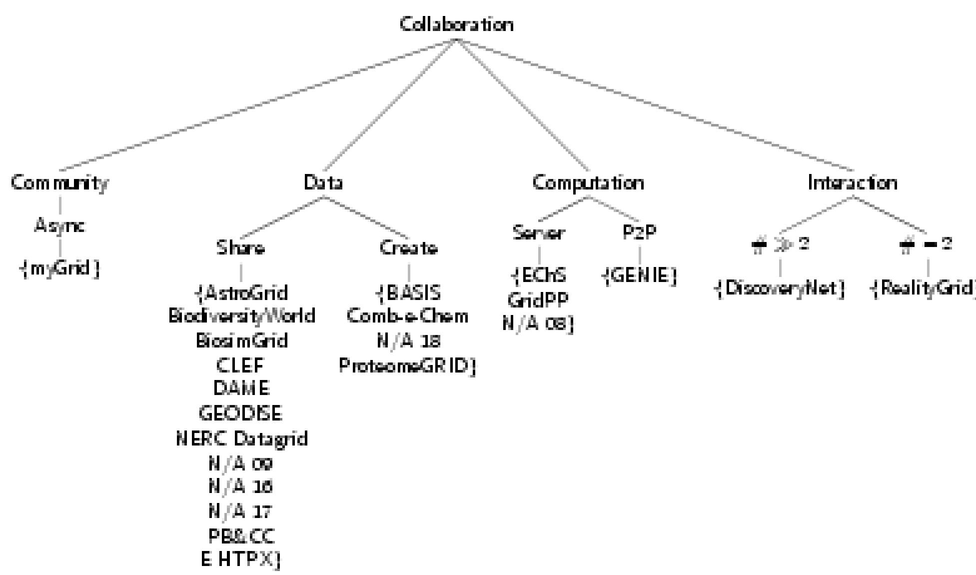


Figure 2: e-Science Pilots Classification

Because the coarse classification of Figure 2 heaps so many projects together in one class, a finer classification is in order. Such a classification is shown in Figure 2. Data-centric virtual laboratories may aim solely provide access to data (share), or they may allow for online modification of the data (create). Virtual laboratories may be conceived as a product to be sold to one or more customers or they may be conceived as a service provided to a community of interacting researchers. Finally, virtual laboratories may aim to integrate dispersed sets of data or they may content themselves with providing a common interface. In the latter case, they can choose to provide extra services (DB++) or not (DB). The branches of the tree in Figure 3 are ordered in such a way that the least ambitious endeavours end up on the left. The resulting tree is clearly out of balance.

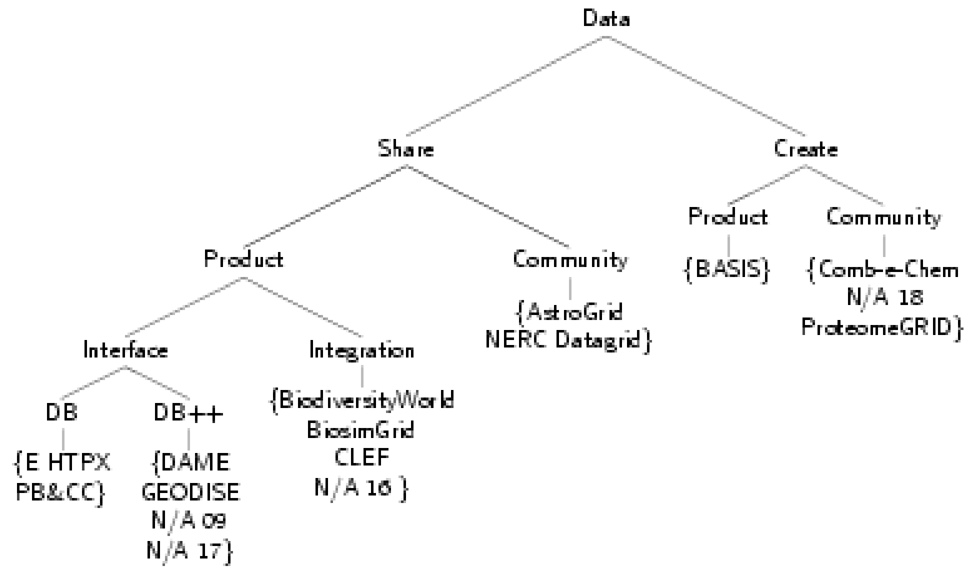


Figure 3: Classification of data-centric e-Science pilots

1.3.2 Comparison with the NSF's pioneer "Collaboratories"

E-science is the United Kingdom contribution to efforts to propel science into the information age. Ten years ago, the NSF program on collaboratories pursued a similar aim. Located within the taxonomy of online collaborations the pioneer NSF collaboratories may be compared with the United Kingdom's e-Science pilots. The general picture that emerges for NSF's "Collaboratories" in Figure 4 shows them to have been rather uniformly spread out over the taxonomic tree. [For specific project details and discussion, see: Computer Science and Telecommunications Board, *National Collaboratories: Applying Information Technology for Scientific Research* (Washington, D.C.: National Academy Press, 1993); Thomas A. Finholt, 'Collaboratories', in Blaise Cronin ed., *Annual Review of Information Science and Technology*, vol. 36 (American Society for Information Science and Technology, 2002); Thomas A. Finholt, "Collaboratories as a new form of scientific organization," *Economics of Innovation and New Technology*, 12, 5-25, 2003.].

The absence of collaboratories whose aim it is to share data does not indicate that collaboratories do not share data. Rather, it indicates that using the data to create new data or using the data to foster a community was deemed more important. The absence of peer-to-peer computing reflect the fact that Internet computing a la Seti@home had still to be invented in the early 1990's.

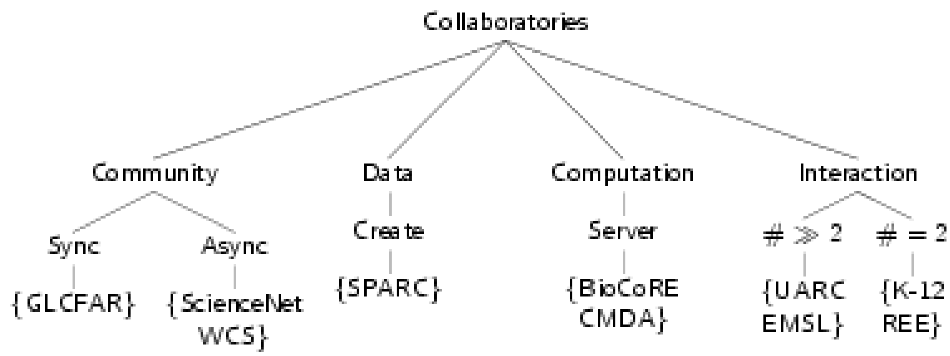


Figure 4: Classification of NSF Collaboratory Projects

1.3.3 Discussion and Conclusion

From this it would appear that, on the whole, collaboratory projects were more concerned with exploring the possibilities of co-ordinated work among researchers who were not co-located, teams, whereas the e-Science "pilots" are more concerned with management and mining of data. This contrast in some part reflects the restriction of the e-Science pilots to software development, and specifically to middleware platforms and applications. Another contrast may be seen: NSF's collaboratories tended to be build by a single research project (albeit in some cases projects organised into a number of teams) and aimed at a geographically distributed community; whereas e-Science pilots tend to be built by consortia but often are aimed at individual users or institutes. Collaboratories and e-Science pilots are very similar in the strategies they adopt for the creation of virtual laboratories: both appear to rely heavily on modularization.

1.4 Other Descriptive Dimensions

The teleological taxonomy discussed in the previous section does not capture all elements of virtual laboratories. Ownership and intended audience is such an element, the organisation of creative activities is another.

1.4.1 Ownership & Audience

A virtual laboratory can be "owned by" a single institution such as a university, or a particular school within a university, or a department of a government institute; alternatively, the vitural laboratory may be "owned by" (in the sense of being under the control of) a consortium of such institutions; or it can be freely available in public domain.

With regard to its intended "users", or "audience," a virtual laboratory similarly may be targeted at a single user or an institutionalised collectivity; it could be designed to serve a club of known individuals; or it can be open to all comers. Together these dimensions of ownership and audience form a matrix. NSF collaboratory projects from the 1990's and current e-Science projects also can be located in the cells of the resulting array. Because there were no instances of pure public domain collaborations, in this application we are can confine the discussion to the distribution of projects over the 2x2 tableau. This comparison shows that the e-Science pilots are typically "owned" by consortia, whereas the collaboratories typically were "owned by" university or other public institutions. Further, e-Science pilots tend to be aimed at single clients more often than at "clubs" of users, whereas the opposite is the case for the NSF collaboratories. Thus we could describe the contrast by saying that the United Kingdom and United States samples tended to cluster at the opposite ends of the principal diagonal of the 2x2 matrix. (As noticed, this simplification is possible because public ownership and/or public accessibility are virtually absent from both sets of projects.)

1.4.2 Organisational scale and structure

Finally, an array of statistics is available on the organisation of e-Science projects, permitting them to be tabulated by, Internet-presence, grant value, start and end date, the number of principal investigators involved, the number of collaborating organisations involved, and the number of components to be produced. These show considerable heterogeneity in every dimension. [See: e.g., NeSC-provided a summary for each of the e-Science pilots.] The fact that the number of components is often more or less the same as the number of collaborating organisations gives rise to the supposition that typically e-Science projects have been split among the collaborators so that cross-institutional collaboration has been minimised. Specifics of the projects detailed below (see Appendix 2) tend to substantiate that conjecture.

Appendix 2

The United Kingdom's e-Science Projects and the Core Programme's "Pilots"

The United Kingdom e-Science Projects

2.1.1 Pilot Projects

A Co-operative Clinical e-Science Framework(CLEF)
A distributed pipeline for structural-based proteome annotation using GRID technology(ProteomeGRID)
A GRID Database for biomolecular simulations(BiosimGRID)
A problem solving environment for global biodiversity: prototype and demonstrator(BiodiversityWorld)
An e-Science resource for High Throughput Structural Biology(E HTPX)
Artificial Neural Networks in Individual Cancer Management(N/A17)
AstroGrid(AstroGrid)
Biological Atlas of Insulin Resistance(BAIR)
Biology of ageing e-Science integration and simulation system(BASIS)
Co-ordination, integration and distribution of sequence and structural family data(N/A16)
Discovery Net: An e-Science test Bed for High Throughput Informatics(DiscoveryNet)
Distributed Aircraft Maintenance Environment(DAME)
Environment from the molecular level: an e-Science proposal for modelling the atomistic processes involved in environmental issues.(N/A08)
Exploring Chemical Structures: Fast Track to New Engineering and Biomedical Materials(Exploring Chemical Structures)
Grid Enabled Integrated Earth system model(GENIE)
Grid Enabled Optimisation & Design Search for Engineering(GEODISE)
GRID for Environmental Systems Diagnostics and Visualisation(N/A09)
Grid-enabled Modelling Tools and Databases for neuroinformatics(N/A18)
MYGRID: Directly Supporting the E-Scientist(myGrid)
Profiling for Breast and Colorectal Cancer(Profiling for Breast and Colorectal Cancer)
Structure-Property Mapping: Combination Chemistry & the Grid(Comb-e-Chem)
The Grid for U.K. Particle Physics(GridPP)
The NERC Datagrid(NERC Datagrid)
The RealityGrid - a tool for investigating condensed matter & materials(RealityGrid)

2.1.2 IRC Projects

Collaborative Advanced Knowledge Technologies in the grid.(CoAKTinG)
Dependability IRC: Dependable, Service-centric Grid Computing(Dependability IRC)
Environmental e-Science IRC: Linking Environmental e-Science in the lab and in the Field(Environmental e-Science IRC)
Grid Based Medical Devices for Everday Health(Medical Devices IRC)
Grid enabled knowledge services: collaborative problem solving environments in medical informatics(Medical Informatics IRC)
MIAS-Grid. A Medical Image and Signal Research Grid(MIAS-Grid)

2.1.3 Demonstrator Projects

3D OPT Microscopy Grid: Bringing the Grid to the Biomedical workbench(3D OPT)
3Dmouse Grid(3Dmouse Grid)
Chemical Structures(Chemical Structures)
Climateprediction.net: Distributed Computing for Global Climate Research - Demonstrator Project(Climateprediction.net Demonstrator)
Collaborative Visualisation(Collaborative Visualisation)
Dynamic Brain Atlas(Brain Atlas)
e-Star: Demonstrating the Power of Grid Computing to Solve Complex, ill defined problems (e-Star)
GODIVA: Grid for Ocean Diagnostics, Interactive Visualisation and Analysis(GODIVA)
GRAB: Biodiversity and the Grid(GRAB)
Health Care Planning with Data Driven Resource Allocation(HYDRA)
Medical Imaging on the Reactor(Reactor)
NeSC ePortal Demonstration(e-Portals)
Seamless Access to Multiple Datasets - An ESRC e-Science Demonstrator(SAMD)
Telemedicine on the Grid(Telemedicine)

2.1.4 International Projects

Distributed Information Services for the Grid(N/A04)
Effective Multi-User and Multi-Job Resource Utilisation(N/A03)
GRID Resource Scheduling(GRS)
Security Aspects of the EU DataGrid(N/A01)
Software Applications for Data Management(N/A02)

2.1.5 Open Call Projects

A Grid-Based approach to the Validation and testing of lubrication models(Grid-Based V&T of lubrication models)
A Scalable Monitoring Platform for the GRID(GridProbe)
e-Diamond: Digital Mammography(e-Diamond)
e-Science Technologies in the Simulation of Complex Materials(N/A13)
e-STORM; e-Science Technical OpeRations Meetings(e-STORM)
Performance-based Middleware for Grid Computing(N/A15)
Visualisation Middleware for e-Science(N/A14)

2. 2 e-Science Core Programme Pilot Projects, Briefly Described

The following descriptive summaries of the current "Pilot Projects" funded under the e-Science "Core Programme" has been derived from the listing: NeSC*, 'e-Science Centre Pilot Projects'*, available at: http://www: umbriel.dcs.gla.ac.uk/- NeSC/general/projects/pilots (2003). These projects are distributed across the taxonomic tree in Figures 2 and 3 of Appendix 1. Here they are listed alphabetically.
The highlighting of names in the following list indicated the 6 pilot projects that have been selected for further discussion (see Appendix 3, below) as "representative" of *collaborative* forms of e-Science.

AstroGrid Support efficient and effective exploitation of key astronomical data sets of importance to the United Kingdom community. Closed club of United Kingdom institutes where six datasets magically match six universities, from which I conclude that CLRCRAL is the sole developer of the Virtual Observatory software.

BASIS Biology of ageing e-Science integration and simulation system. Allow the integration of data and hypotheses from diverse biological sources. Similar, but simpler: University of Newcastle provides the data, CLRC does the integration. Outside ageing modellers are invited to collaborate directly with the BASIS project. Sounds like open source mode.

BiodiversityWorld A problem solving environment for global biodiversity: prototype and demonstrator. The main features of the prototype will be: linkage to an existing partial catalogue of life sufficient to demonstrate the system, a partial array of linked thematic data sources and a system for collating data from the sources and using it in existing biodiversity analytical tools. Southampton has a CS department with grid expertise; Reading does the Biodiversity analytic tools; the others provide data.

BiosimGRID A GRID Database for biomolecular simulations. Aims to build a database, and since it is funded as a grid project, why not make it a distributed database. Building the database is a collaborative effort by members of the project. They will establish a formal consortium for this purpose. In addition, one post-Doc will be paid to find out that industrial products typically come with a better, more user friendly, interface than their academic counterparts. She will then try to develop a better interface on top of the existing product in spite of software development rules of the thumb indicating that starting with a good interface might actually be a better idea.

CLEF A Co-operative Clinical e-Science Framework. CLEF will build tools to link clinical and genomic research and to make clinical trials faster, more effective, and easier to manage throughout their cycle – from informal hypothesis evaluation through patient recruitment to data collection and analysis. Royal Marsden NHS Trust and UCL(?) do cancer information systems; The universities of Manchester and Sheffield provide MyGRID infrastructure. Cambridge provides the ethico-legal perspective. Policies of information governance have to be agreed upon by all, say Marsden + UCL + Brighton, upon which technical enforcement mechanisms are developed by, say, Brighton+UCL.

Comb-e-Chem Structure-Property Mapping: Combination Chemistry & the Grid. Integrates existing structure and property data sources within a grid-based information and knowledge-sharing environment. Bristol provides one collaborator: Guy Orpen, Professor of Structural Chemistry. The rest is from Southampton, so presumably development is done there.

DAME Distributed Aircraft Maintenance Environment. Develops a Distributed Diagnostic Grid Test-bed. Rolls Royce is the customer. Three Universities have their pet Diagnostic systems. University no.4 will build a common interface to them.

DiscoveryNet Discovery Net:An e-Science test Bed for High Throughput Informatics. Continuation of existing architecture developed at IC. Supports real-time processing, interpretation, integration, visualisation and mining of massive amounts of time critical data generated by high throughput devices.

EChS Exploring Chemical Structures: Fast Track to New Engineering and Biomedical Materials. Store and analyse data on a scale that will support faster, cheaper synthesis of a whole range of new materials. Yet another Southampton project.

GENIE Grid Enabled Integrated Earth system model. Develop a fast, scalable, grid enabled, modular 3-D Earth systems model. A collaborative effort. Grid by e-Science centres; collaborative modelling by the rest.

GEODISE Grid Enabled Optimisation & Design Search for Engineering. Grid-based seamless access to an intelligent knowledge repository, a state-of-the-art collection of optimisation and search tools, industrial strength analysis codes, and distributed computing and data resources. Collaborators: Prof. Simon Cox University of Southampton GRID Technologies Prof. Mike Giles Oxford University Computational Fluid Dynamics Prof. Carole Goble Manchester University Ontology Prof. Andy Keane University of Southampton Design Optimisation Prof. Nigel Shadbolt University of Southampton Knowledge Management

GridPP The Grid for U.K. Particle Physics. Grid software (middle-ware) and hardware infrastructure to enable testing of a prototype of the Grid for the Large Hadron Collider (LHC) project at CERN. 2 PIs, both particle physicists, one from Glasgow, the other from Queen Mary, London. Every United Kingdom University Particle Physics Group is participating apart from Kings, Plymouth, and, of all places, Southampton. Presumably Glasgow and Queen Mary will do the (open) development and the rest of the project members have committed themselves to testing.

NERC-Datagrid The NERC Datagrid. The first one to cross United Kingdom boundaries: building on technology developments carried out primarily in the United States by collaborators from the Earth System Grid. Builds a grid which makes data discovery, delivery and use much easier than it is now, facilitating better use of the existing investment in the curation and maintenance of quality data archives.

N/A08 Environment from the molecular level: an e-Science proposal for modelling the atomistic processes involved in environmental issues. Simulation of processes at molecular level. To be implemented by a consortium of scientists from the Universities of Cambridge, University College London, Reading and Bath, the Daresbury Laboratory, and the Royal Institution.

N/A09 GRID for Environmental Systems Diagnostics and Visualisation. Develop oceanographic diagnostics grid with data from NERC OCCAM model and Met Office FOAM with remote visualisation and application drivers. Met Office provides the data and the usual suspects of CLRC, Reading, IC, and Southampton do the rest.

N/A16 Co-ordination, integration and distribution of sequence and structural family data. Five databases, five project members, and four collaborating organisations.

N/A17 Artificial Neural Networks in Individual Cancer Management. Develop and train a system of artificial neuronal networks (ANNs) providing information on prognosis, staging and optimal (multidisciplinary) management in-patients with (a) breast, (b) upper GI and (c) colorectal cancers. Nice project, but it has nothing to do with e-Science whatsoever.

N/A18 Grid-enabled Modelling Tools and Databases for neuroinformatics. Provides:
(i) technologies that enable the exchange of simulation, analysis and visualisation software components ; (ii) access to software tools that support discussion, development and exchange of computational models and components, the data they are based upon, and the results that they generate, including databases of models and linkages to datasets. Involves collaboration with a university in Germany and an institute in France.

PB&CC Profiling for Breast and Colorectal Cancer. Linkage of bioinformatic and clinical databases. Collaboration between Imperial College and Imperial College.

ProteomeGRID A distributed pipeline for structural-based proteome annotation using GRID technology. Establishes local databases with structural and functional annotation and disseminates the databases to the biological community via a single web-based distributed system (DAS) based at the EBI. Finally, it aims to share computing power. (At least they mastered the e-Science jargon) EBI does DAS; IC+ULC add GRID.

RealityGrid The RealityGrid – a tool for investigating condensed matter & materials.
Enable the realistic modelling of complex condensed matter systems at the molecular and mesocale levels, and for the discovery of new materials. Grid hardware and middleware are provided by e-Science centres; virtual reality by Loughborough; experimental data and modelling by Oxford (new materials), Queen Mary (condensed matter at molecular level) and Edinburgh (mesoscale level).

E-HTPX An e-Science resource for High Throughput Structural Biology. Each collaborator provides facilities and somehow they will be merged into an easy-to-use resource.

myGrid MYGRID: Directly Supporting the E-Scientist. The project will develop an e-Scientist's workbench that supports: (i) the scientific process of experimental investigation, evidence accumulation and result assimilation; (ii) the scientist's use of the community's information; and (iii) scientific collaboration, allowing dynamic groupings to tackle emergent research problems. Looks like truly open source with Twiki, Bugzilla, and BBS. For instance, Twiki documentation is modifiable by anyone.

Appendix 3

The Role of Databases in Collaborative e-Science:
Examples from Representative Pilot Projects

The six projects selected here exhibit a range of collaborative e-Science activities that are depicted in the principal branches of the taxonomic tree of Figure 1. They are the ones whose names appear highlighted in the brief descriptions provided by Appendix 2.2. The notes supplied here focus upon the role that databases play in these collaboration tools.

MyGrid MYGRID–Directly Supporting the E-Scientist £ 3,483,003

This (*community-centric*) project will develop an e-Scientist's workbench that supports: (i) the scientific process of experimental investigation, evidence accumulation and result assimilation; (ii) the scientist's use of the community's information; and (iii) scientific collaboration, allowing dynamic groupings to tackle emergent research problems.

This project is about ``information weaving": connecting large amounts of heterogeneous data. (for metadata ontology/services). The classes of information the system intends to provide include:

Who else asked this question & can I use/adapt their approach? (workflow repository)
What were the results at each stage? (dynamic data repositories)
When was X last updated? Which version of Y did I use? (provenance)
Has Z changed since I last ran this? (notification)
Personalisation (personal repository)

2. NERCDatagrid The NERC Datagrid £ 721,188

Based on technology developments carried out primarily in the United States by collaborators from the Earth System Grid, this (*data-centric/data-sharing*) project builds a grid which makes data discovery, delivery and use much easier than it is now, facilitating better use of the existing investment in the curation and maintenance of quality data archives.

Project objective: ``To provide an easy-to-use interface supporting the discovery, pre-processing and visualisation of heterogeneous data from heterogeneous sources for earth scientists."

Prior experience: NERC Metadata Gateway. Task: find out which data are available, i.e. search through multiple data archives simultaneously. NERC Datagrid aims to do essentially the same thing but better – more flexible and extensible, capable to handle differences in jargon across scientific fields, hide heterogeneity of data, etc.

Comb-e-Chem: Structure-Property Mapping for Combination Chemistry £2,314,665

Structure-Property Mapping: Grid-enabled Combination is a larger (*data-centric/data-sharing*) tool that integrates existing structure and property data sources within a grid-based information and knowledge-sharing environment.

Users using Comb-e-Chem search primary databases and fill new databases with information about their experiments. The new databases generated can be accessed, but not modified, by others. (Whether annotation of others' database entries is possible remains unclear from published descriptions.)

4. CLEF : Co-operative Clinical E-Science Framework £1,683,000

CLEF will build *(data-centric/data-generating)* tools that link clinical and genomic research to make clinical trials faster, more effective, and easier to manage throughout their cycle – from informal hypothesis evaluation through patient recruitment to data collection and analysis.
Agreed policies on information governance – confidentiality, access, authentication, and consent – and technical mechanisms for their enforcement.
Tools for information capture.
Integrated repository
Tools to access and present the information

Users of CLEF make data on their PC (such as reports, DBs, x-ray scans) available to CLEF, which will extract and integrate information and present it in an intuitive fashion to the users to be searched by them.

5. GENIE: Grid Enabled Integrated Earth system model £ 705,414

The Grid Enabled Integrated Earth system model is building a (*computation-centric, peer-to-peer)* tool: a fast, scalable, grid enabled, modular 3-D Earth systems model.
The following simplified example envisages how the GENIE system will be used: The system will be accessed via a portal, which will allow a user to compose, execute and analyse the results from an Earth system simulation. After authenticating themselves with the portal, users will have access to a library of components each of which can model different aspects of the Earth system (for example, ocean, and atmosphere) at different resolutions. The user constructs a composite application by selecting from these components; a selection is informed by meta-data provided by a component's author and made available in the library. Together with these components the user selects appropriate mesh conversion tools to enable data exchange at model boundaries, and defines an event-queue, which specifies the timing of the data exchanges and indicates what simulated data is to be archived for later inspection. The user also provides the data necessary to initialise the model. From this, an intelligent meta- scheduler determines the resource requirements and maps the processing required to a distributed Grid of computing resources using middleware such as Globus and Condor. At runtime each component produces distributed data, which can be monitored during execution and is also archived automatically as specified by the user. From the portal it is possible to browse this archive of results using post-processing visualization tools and to re-use results from the archive to seed new calculations.

6. DiscoveryNet: e-Science Test Bed for high Throughput Informatics £2,082,704

The project, a continuation of development work on of existing architecture developed at Imperial College, with (*interaction-centric*) facility is intended to support real-time processing, interpretation, integration, visualisation and mining of massive amounts of time critical data generated by high throughput devices.
Allows scientists to plan, manage, share and execute complex knowledge discovery and data analysis procedures available as remote services (accessing a database for the search)
Allows service providers to publish and make available data mining and data analysis software components as services to be used in knowledge discovery procedures (accessing a database as the directory).
Allows data owners to provide interfaces and access to scientific databases, data stores, sensors and experimental results as services, so that they can be integrated in knowledge discovery processes.

APPENDIX 4

The Globus Project's Approach to Software Licensing[63]

0. Introduction

The Globus Project and Globus Toolkit and are the hardware and software pillars, respectively, on which the technological infrastructure of e-Science rests. Many e-Science projects are implemented on top of the Globus Grid architecture and even when other workgroups will come to devise and employ alternative middleware, the descriptive language of e-Science is likely to depend heavily on concepts first conceived at Globus Project. But that the same can be said in regard to its influence upon the institutional infrastructure of e-Science rather quite uncertain.

This appendix document first outlines how the Globus Project has resolved the question of intellectual property protections for the computer programmes and protocols comprising the Globus Toolkit, and then described the motivating considerations behind the development of a novel form of software license that represents a variant form of General Public License. A third section indicates to what extent the Globus licensing approach may be more widely applicable to the tools created by e-Science projects.

1. The Globus Toolkit Public Licence

Globus Toolkit is distributed under the so-called Globus Toolkit Public Licence (GTPL). Contributions are welcomed, as long as they are explicitly licenced to the Globus Project. That is, contributors must sign a license agreement granting non-exclusive, worldwide, royalty-free perpetual copyright to the Globus Project for use of their contribution. If parts of the contribution are patentable, the patents preferably should be royalty free. Further, a contributor who sued the Globus Project for patent infringement would immediately lose any right to use the Globus Toolkit. Globus Toolkit and Globus Project are trademarks and the use of these names requires prior approval by the Globus Project.

GT is not GNU

Much of "free and open source software" is distributed under GNU General Public License (GPL), or similar licenses that incorporate the "copyleft" principle – that the user may modify the code, but must make it available to others on the same licensing terms under which it had been obtained.

GPL, however, has been pronounced unacceptable by some commercial software vendors and distributors, and since widespread adoption of the Globus Toolkit is one of the major goals of the Globus Project, a choice has been made to abandon GPL in its entirety. Instead, the project has opted for a BSD-like license. [See Bollinger (2002), slide 7 for a comparison of licenses].

The abandonment of GPL required audit of existing code and replacing GPL-ed code by new code; it further imposes upon new contributions the requirement of similarly complying with GTPL (e.g., by not using GPL code in their own contributions).

Contributors' Rights & Duties

[63] The research for this document benefited greatly from Steve Tuecke's presentation at the 2nd EPSRC e-Science Pilot Projects Workshop in Edinburgh (31 January 2003); and the interview (on 9 March 2003) that Matthijs den Besten conducted with Helen Cordell, the attorney at Argonne National Laboratory, Chicago, who was primarily responsible for designing the GTPL license.

In open collaborations like the Globus Project contributions from third parties involve a certain risk. First of, contributors may not actually own the copyright to their contribution and hence not be entitled to give it away. For instance, the University of Oxford asserts copyright to commissioned works and some companies assert copyright to any creative work done by their employees. Secondly, contributors may suddenly discover that they possess patents on parts of their contributions and try to reap benefits. An example of this is Unisys' discovery of its patent on LZW, a compression procedure used in GIFs [see http://www.BurnAllGIFs.org].

The experience of Linux shows that these are not insurmountable risks. But the Globus Project has been risk averse, and to mitigate the foregoing risks it has requires future contributors to explicitly license their contributions to the Globus Project and, in effect, promise that they will not seek to enforce their patent rights. It should be noted that contributors to Globus Toolkit would retain full rights to license their software to other parties under completely different terms, although the contributors' bargaining power will undoubtedly lessen when the contribution can be obtained freely under GTPL from the Globus Toolkit.

Patents & Trademarks

Patenting its software does not presently rank highly among the Globus Project's objectives. Nonetheless, it may see itself forced to apply for patents in the U.S. in order to fend off patent infringement claims by others. The patents obtained in this way will be royalty free (RF). Contributors' patents also should be RF, although that is not mandatory. Alternatively the Globus Project may accept contributions that are patented and licensed in a reasonable and non-discriminatory (RAND) fashion, but no final decision on this question has been made (as of 9 March 2003)..

Another issue of contention is the scope of the defensive termination clause, which is part of the Grant of License in Globus Toolkit Contributions. According to the weak variant of this clause, the act of suing for patent infringement revokes any right of the suitor to the use of the Globus Toolkit (module). In the strong variant, the suitor surrenders his rights to use any of the software created by any of the partners in the Globus Project. In the latter case, the Globus Project would bear the burden of a patent dispute between, say, IBM and SUN. In the former case, the cost may not be high enough to deter the contributor from suing.

Globus Project and Globus Toolkit are trademarked so that the Globus Project can protect the brand and keeps full control of the versions that are developed in its name.

The Globus Projects' Motivation

The IPR solution developed for the Globus Project by personnel at the Argonne National Laboratory was the outcome of a bargaining process between the Project and its industrial and academic partners.

Globus Project

The goal of the Globus Project is to set the standard for Grid computing. To further this goal, it is crucial that a large user base adopts the Globus Toolkit early on. The Globus Project can draw from considerable resources: it is a collaboration between the University of Southern California and the University of Chicago (as Operator of Argonne National Laboratory). So, although contributions from outside are welcome, they are not necessary for the success of the project. At the same time, the endorsement by major software distributors in industry is deemed to be highly important, as that has appears to be the most expedient means of promoting the use of the Globus Toolkit in the realms of e-Government, e-Business, etc.

IBM c.s.

For industry, the Globus Project is one of the many horses to bet on. If the Globus Toolkit becomes a *de facto* standard, its industrial partner will benefit from having a stake in the products and services that it can provide commercial and which are complementary to Grid services and Grid computing (the Grid represents a multi-million business generator for IBM). If not, partnership with Globus still is a useful way to stay abreast of the rapidly advancing state of knowledge in academic software engineering.

Academia

Getting credit for the contribution is crucial for most researchers at universities. They tend to be hesitant to grant a license for their contribution to the Globus Project since they cannot be certain that the Globus Project will give due credit. [Allan *et al.,* 2001] An additional problem may be posed by conflicting IPR policies from universities and funding bodies.

Globus' Scalability

The Globus Project is a collaborative effort between a limited number of institutional partners predominantly based in the United States. Undoubtedly, the Globus approach becomes more difficult to emulate as the size and complexity of the collaboration increases.

Direct Applicability

Many e-Science projects may be able to manage their intellectual property on the same terms as the Globus Project – specifically those projects that are organised as a collaboration between a limited number of institutes within the United Kingdom, and for which the code has to be developed from scratch or contains easily replaced GPL-ed code. It is however an open question whether this strategy should be actively encouraged by funding agencies, as there is a danger of premature convergence on standards whose ramifications are not fully understood. The implications of the GNU GPL by comparison, are far more thoroughly understood by those in the software industry (despite the continued emission of a certain amount of strategic dis-information by firms and organisations opposing the free and open source software movement).

In collaborations that really on a big number of collaborators from all over the world, the licensing approach adopted by the Globus Project is unlikely to be feasible.

Similar Approaches

One important lesson illustrated by the Globus Project case is that there is no necessity to stick with the IPR regime that is codified in law. It is perfectly possible to amend copyright and patent right with contracts, when some effort can be invested in the process and when project co-ordinators are endowed with some power to set the terms.

The development of novel licensing forms to meet particular circumstances is not inconsistent with the procedural recommendations developed in this report. On the other hand, the multiplication of many variant forms to meet special circumstances about which other potential users may have little or no experience-based information is a serious drawback of a laissez-faire approach in this regard.

Multiple Jurisdictions

A final complication worthy of notice arises when collaborations span multiple jurisdictions. In case of the Globus Project, copyright and patent right is defined in the U.S. at the federal level. Still, the fact that collaborators are working in different states may cause a problem when the validity of the Grant of License in Globus Toolkit Contributions is disputed because contract law is decided upon on the state level.

Despite devolution, the United Kingdom is still so centralised that similar problems within pure United Kingdom e-Science projects are unlikely to occur. Nevertheless, some problems are almost certain to arise when the United Kingdom e-Science project involves a partner from the United States, especially when databases are involved. Within the European Union, traditions on copyrights and patents diverge and directives are not implemented uniformly. In addition, there are big differences among member-states in the extent to which contracts are allowed to override the statutory limitations on copyright [Guibault, 2002]

Conclusion

The way the Globus Project manages IPR, its BSD-like license and demand for explicit contributor consent, is interesting and justifiable in the circumstances in which this novel license was designed. It is evidently the (intermediate) outcome of a process of give and take between the project's partners.
Therefore, it is important to note, first, that there is rather more symmetry between the resources of the Globus Project at Argonne and its principal industrial partners (IBM and Oracle) than is likely to obtain in the case of a typical university-based e-Science project.
Secondly, the Globus Project is essentially a monolithic organisation that has funded the development of the software, and its partners have entered to handle the distribution of the code. Whether or not e-Science projects will be able to emulate this approach depends on their characteristics.
A similar approach may work for some of the more highly unified among the larger e-Science pilot projects, but it is likely to provide more difficult in more ambitious and complex multi-institution collaborations.

References for Appendix 4

Allan, R. J., Boyd, D. R. S., Folks, T., Greenough, C., Hanlon, D., Middleton, R. P., Sansum, R. A., Mourao, Elson, Baxter, Rob, Henty, David, Brooke, Parsons, Mark John, Hewitt, W. T., Daw, Mike, MacLaren, Jon, Gibson, Jon, Riley, Graham & Pickles, Stephen, 'Globus and Associated Grid Middleware – Consolidated Evaluation Report from UKHEC Sites', Tech rep., CLRC e-Science Centre, http://esc.dl.ac.uk/TechReports/consolidate/consolidate/node25.html (2001).

Bollinger, Terry, 'Use of Free and Open Source Software (FOSS) in the U.S. Department of Defense', Presentation for "IT Applications "Community of Practice" at USAID – Third Meeting" (2002).

Bollinger, Terry, 'Use of Free and Open-Source Software (FOSS) in the U.S. Department of Defense', MITRE Report MP 02 W0000101, The MITRE Corporation, http://www.disa.mil/pao/dodfoss.html; http://www.egovos.org/pdf/dodfoss.pdf (2003).

Guibault, Lucie M. C. R., *Copyright limitations and contracts : an analysis of the contractual overridability of limitations on copyright* (The Hague: Kluwer Law, 2002).

University of Chicago and The University of Southern California, 'Globus Toolkit Public License', Version 2 (2003).

University of Chicago and The University of Southern California, 'Grant of License in Globus Toolkit Contributions', Prototype (2003).

Appendix 5
The Effect of Uncertainty as to Ownership: A Cautionary Tale
The Attwood Experience of Database Ownership

The difficulties of not agreeing *ex ante* where the fruits of particular research will lie are legion and can impact on the ability of a scientist, not only to commercialise her work, but to use it for her own subsequent scientific research. The following cautionary tale is extracted from a paper by Professor Teresa K. Attwood of the School of Biological Sciences and the Department of Computer Sciences at the University of Manchester:

"**The tale of a roaming PRINTS**

"PRINTS is a database of protein 'fingerprints' that can be used to diagnose family relationships in newly-determined sequences (*e.g.*, from genome projects). Fingerprints exploit groups of conserved motifs within sequence alignments to build characteristic family signatures; an uncharacterised sequence that matches all motifs can thus be readily assigned to a particular family. The diagnostic power of fingerprints, and the extent of documentation manually attached to each database entry, has lent PRINTS a significant role in protein sequence analysis and, ultimately, genome annotation. The fingerprint technique grew from my postdoctoral studies at Leeds University in 1989. With the help of colleagues between 1991 and 1993, I compiled 100 fingerprints into an un-named prototype resource. This work was not supported by a designated research grant; it was essentially a 'proof of concept' of the fingerprint method. In October 1993, I moved to UCL, having been awarded an independent Royal Society University Research Fellowship. This signified the first direct funding to support the resource, and I marked the event by naming the database PRINTS, formally identifying it with UCL. The first release contained 150 entries and, for a time, my colleagues in Leeds continued to give support. In August 1994, a Web site was established at UCL to serve PRINTS and related databases and tools. During this time, PRINTS grew slowly but steadily, with quarterly releases each adding 50 new entries. In 1998, in an effort to streamline its management, we migrated PRINTS into a relational database management system, and called the new streamlined resource PRINTS-S.

For a while, the databases had a stable home but, in 1999, PRINTS and PRINTS-S followed me to a new position at the University of Manchester. Today, with 1550 entries, PRINTS31.0 is 50% larger than PROSITE, the 'gold-standard' protein family database. Last year, PRINTS, PROSITE, Pfam and ProDom were integrated into a unified family resource termed InterPro, (maintained at the EBI) which, in spite of its youth, has already had a key role in the annotation of the fly and human genomes.

Who owns PRINTS?

Like databases, academic environments are also evolving. Today, universities tend to protect their IP aggressively, whether or not they understand exactly what it is they're trying to protect. Accordingly, when I moved to UCL, Leeds adopted the position that, as my work had originated there, they owned it. In response to the European Directive on the legal protection of databases (EU Directive 96/9/EC), they placed licence and copyright notices on their Web site (see http://bmbsgi11.leeds.ac.uk/bmb5dp/homeprints.html). UCL immediately requested their removal, initially to no avail. Consequently, there followed a three-and-a-half-year battle between the organisations, at the end of which Leeds removed the licence restrictions, but left the copyright statements in place (http://bmbsgi11.leeds.ac.uk/ bmb5dp/prints.html). Ironically, these pages, and the underlying database, haven't been updated since the impasse in April 1997, and hence today provide an interface to a resource that's more than four years out of date.

Weathered by this experience, I feared similar battles on moving to Manchester. Fortunately, however, UCL has not defended its interest in either PRINTS or PRINTS-S, and Manchester has shown only a passing curiosity – they have understood its history and thought better of getting involved. In fact, they probably realised that claiming ownership would not be straightforward and any attempt to exploit the IP in PRINTS would be fraught with problems, especially if all 3 universities were to enter the ring. Ultimately, commercially, the databases are a minefield because who owns them is not clear – or, at least, has not been agreed."

Source: Teresa K. Attwood "Mobile, Metamoprhosing Academic Databases – Capturing IP on the move" in *Workshop Report on Managing IPR in knowledge-based economy – Bioinformatics and the influence of public policy* European Commission DG Research - Fifth Framework Pogramme (2001)

Appendix 6
The Effect of Inappropriate Licensing: A Cautionary Tale
The SWISSPROT Experience of Database Licensing

Background: Biologist A. Bairoch founded the SWISS-PROT protein database with support from the Swiss government in 1986. But the project proved to be expensive to maintain. In 1996, the Swiss government stopped funding the project. When the European Community declined to fill this gap, Bairoch and his colleagues set up a new company called Geneva Bioinformatics ("GeneBio"). In September, 1998, GeneBio began to charge non-academic users an annual subscription fee. Academic users continue to use SWISS-PROT without charge.

Terms for Commercial Users: *Commercial* subscriptions range from $2,500 to $90,000, depending on company size and number of users. Commercial database providers can redistribute the database provided that they (a) pay a fee, (b) provide a list of their customers, (c) update the list each month, (d) promise not to modify the database, and (e) require each customer to sign a separate license agreement. SWISS-PROT also requires anyone who copies its data to include a notice requiring for-profit users to pay a subscription.

Restrictions on Academic User: SWISS-PROT's disclaimer prevents non-profit users from modifying the database for any purpose. SWISS-PROT has said, however, that it will waive this provision if there is "a valid scientific or technical reason," depending on "how the information will be presented and distributed."

Unanticipated Reactions: The U.S. National Center for Biotechnology Information ("NCBI") is a United States government agency that develops databases for academic and commercial users. Prior to 1996, NCBI (a) distributed copies of SWISS-PROT to anyone who requested it, and (b) incorporated SWISS-PROT data into its own "Reference Sequence" and "Predicted Genes" databases. When SWISS-PROT began selling data, NCBI obtained permission to post the database on its web site in a non-downloadable format. Subsequently, however, NCBI stopped incorporating SWISS-PROT annotations in its own databases because it wanted those resources to remain unencumbered. NCBI eventually replaced SWISS-PROT with data from other sources.

Subsequent Policy: Despite its decision to become self-supporting, SWISS-PROT opposes similar arrangements for gene sequence data: "We consider that the funding model that has to be adopted to secure the viability of SWISS-PROT is not applicable to the international nucleotide sequence databases . . . even though these are also curated. Nucleotide sequences, from which SWISS-PROT entries are derived, must remain in the public domain in recognition of the fact that they are the primary data, and have been submitted to public-domain collections by individual scientists. This same consideration holds for primary databases of macromolecular structures (such as PDB)."

Source: Stephen M. Maurer, "Promoting and Disseminating Knowledge: The Public/Private Interface," NRC Conference on *The Role of the Public Domain in Data and Information*, National Academy of Science, Washington, D.C.,September 5-6 2002. Policy statement quoted by Maurer (2002) from SWISS-PROT Webpage.

Appendix 7

OXFORD INTERNET INSTITUTE

Participants List for E-Science Workshop 02.05.03

Mr Iain Banfield	Department of Trade and Industry	Iain.Banfield@dti.gsi.gov.uk
Professor Michael Blakeney	Herchel Smith Professor of Intellectual Property Law, Queen Mary, University of London, Director of the Queen Mary Intellectual Property Research Institute	m.blakeney@gmul.ac.uk
Ms Leandra Briggs	Head of Research Contracts, Research Services, University of Oxford	lea.briggs@admin.ox.ac.uk
Professor Jean-Michel Dalle	Associate Professor, University Pierre-et-Marie-Curie, Researcher, IMRI-University of Paris Dauphine	Jean-Michel.Dalle@admp6.jussieu.fr
Professor Paul David	Professor of Economics, Senior Fellow, Institute for Economic Policy Research, Stanford University, Senior Fellow, Oxford Internet Institute, University of Oxford	paul.david@all-souls.oxford.ac.uk
Mr Peter Davies	Visiting Fellow, Oxford Internet Institute, University of Oxford Barrister	Peter.Davies@oii.ox.ac.uk
Mr Matthijs den Besten	St Antony's College, University of Oxford OII Research Office – e-Science Project	matthijs.denbesten@sant.ox.ac.uk
Mr Matthew Dovey	Technical Manager, Oxford e-Science Centre, University of Oxford	matthew.dovey@oucs.ox.ac.uk
Professor William Dutton	Director, Oxford Internet Institute, Professor of Internet Studies, University of Oxford	William.Dutton@oii.ox.ac.uk
Dr Thomas Finholt	School of Information, University of Michigan	finholt@umich.edu
Dr Michael Fraser	Head of the Humbul Humanities Hub Research Technologies Service, University of Oxford	mike.fraser@computing-services.oxford.ac.uk
Dr Jeremy Frey	Senior Lecturer in Chemistry, University of Southampton	j.g.frey@soton.ac.uk

Mr Rishab Aiyer Ghosh	MERIT, University of Maastricht Editor and Publisher of *Indian Techonomist* Co-Founder and Managing International Editor of *First Monday*	rishab@dxm.org
Professor Clive Gringas	E-Commerce Partner, Olswang	clive.gringras@olswang.com
Dr Yike Guo	Professor of Computing Science, Department of Computing, Imperial College, University of London	yg@doc.ic.ac.uk
Professor Tony Hey	Director, e-Science Core Programme, Engineering and Physical Sciences Research Council (EPSRC)	Tony.Hey@epsrc.ac.uk
Professor Paul Jeffreys	Director, Oxford e-Science Centre, Oxford University Computing Services	paul.jeffreys@oucs.ox.ac.uk
Dr Marina Jirotka	Associated Director, Oxford e-Science Centre, Oxford University Computing Laboratory	Marina.Jirotka@comlab.ox.ac.uk
Dr Anselm Kamperman Sanders	Associate Professor of Intellectual Property and Trade Law, Faculty of Law, University of Maastricht	A.KampermanSanders@PR.unimaas.nl
Dr Bryan Lawrence	Council for the Central Laboratory of the Research Councils	b.n.lawrence@rl.ac.uk
Ms Catherine Quinn	Director of Research Services, University of Oxford	catherine.quinn@admin.ox.ac.uk
Professor Ian Lloyd	Professor of Information Technology Law, University of Strathclyde	i.j.lloyd@strath.ac.uk
Dr Andrew Martin	Lecturer in Software Engineering, Oxford University Computing Laboratory Deputy Director of the Centre of Excellence in Software Engineering	Andrew.Martin@comlab.ox.ac.uk
Mr Mustafizur Rahman	ICT/Begbroke Directorate, Oxford University	mustafizur.rahman@begbroke.ox.ac.uk
Mr Sebastian Rahtz	Information Manager, Oxford University Computing Services	sebastian.rahtz@oucs.ox.ac.uk
Dr Michael Spence	Faculty Associate, Oxford Internet Institute, Chair of the Law Board, University of Oxford, Consultant, Olswang; Faculty of Law, Boston University	michael.spence@law.oxford.ac.uk

Professor W. Edward Steinmueller	Professor of Information and Communication Technology Policy, SPRU, University of Sussex	w.e.steinmueller@sussex.ac.uk
Dr Arthur Thomas	Managing Director, Proteus Associates	ajt@proteus-associates.com
Professor David Vaver	Reuters Professor of Intellectual Property and Information Technology Law, University of Oxford	david.vaver@law.ox.ac.uk
Dr Daniel Zizzo	Lecturer in Economics, Christ Church College, University of Oxford	daniel.zizzo@economics.ox.ac.uk