Institutional Infrastructures for Global Research Networks in the Public Sector

Michael Spence
St Catherine's College, Oxford

Establishing effective scientific collaborations is not just about getting the *computing* science right, it is also about getting the *social science* right. If global research networks are to deliver the fruit that they promise, they need effective institutional infrastructures.

Take an example from a GRID based collaborative project. The UK eDiaMoND project was a pilot, intended to build a national database of mammographic images. The project was managed by the University of Oxford Computing Laboratory and involved the University of Oxford Engineering Department, two commercial partners (IBM and Mirada Solutions), and four public hospitals and the universities to which they were attached (three in England and one in Scotland). It was funded by the UK Engineering and Physical Sciences Research Council, the UK Department of Trade and Industry and an IBM Sur Grant. The project was based upon a head contract between the University of Oxford, IBM and Mirada solutions and a series of letters of agreement with partner universities. The legal agreements did not directly extend to the hospitals. Their involvement was covered by agreements with research and development departments within the relevant hospital trusts.

This was a relatively straightforward project in three ways. First, many of the parties and regulatory bodies were willing to forego dispute over potentially contentious issues because the project was a two year trial. Second, the project was national, rather than international. Third, as a project to construct a database and the platform to support it, it was one of the more institutionally simple of the categories of e-science project that I shall come to outline.

Nevertheless, as this project was put together it was not always clear that the institutional obstacles it faced would be overcome, and many crucial issues remain unresolved in a way that would be unsatisfactory at the next stage of its development. Thus the project had to deal with differences in the policy of government health, and information technology services, in England and Scotland. Permission to use mammographic images had to be obtained from the ethics committees of each of the relevant hospitals and from the UK medical research ethics committee. Questions about the extent of patient consent for the use of images would need to be re-examined if more family information were to be linked to images in the future. Attitudes to the ownership of images included in the project varied between project partners. The question of who owns the database rights in the database of the images as a whole is as yet unresolved. Further issues concerning the impact of competition law norms on the development of the library, and liability for any aspect of the database that might give rise to misdiagnosis were it to be clinically used, are as yet beyond the radar screen of those responsible for the pilot project. It is a testimony to their commitment to this project that the pilot was realised, but it is clear that inappropriate institutional arrangements could have been, and perhaps could still be, as fatal to the success of the project as the technical challenges that it faced and overcame. The design of institutional infrastructures is as important as the design of technical infrastructures for the next generation of collaborative science.

Because of this importance, I would argue that projects involving significant public sector funding should set the standard for those infrastructures: public funding bodies must ensure, not only that good science is done, but that the institutional context in which it is done maximises the return on their investment in the scientific commons. If the funding bodies have a role in the promotion of public science, then they must also have a role in the promotion of institutional best practice.

However, while this all sounds like common sense, it is not at all easy to set a standard for effective institutional infrastructures in e-science. The task facing the public funding bodies is a daunting one, that they might be tempted to leave alone, muttering (wrongly) about technical determinism and how the right institutional arrangements will follow the technical possibilities. The problem is three-fold.

First, there is considerable diversity in the types of e-science project. A taxonomy that Paul and I have offered in the circulated paper suggests four different types of project. Thus: (i) some are "community-centric", bringing together researchers for synchronous or asynchronous information exchanges, (ii) some are "data-centric", providing accessible stores of data drawn from different sources, editing and annotating them, (iii) some are "computation-centric", providing high-performance computing capabilities, either by means of servers accessing super-computers and parallel computing clusters, or making possible for collaborators to organise peer-to-peer sharing of computation capacity, and finally (iv) some are "interaction-centric", enabling applications that involve real-time interactions among two or more participants for decision-making, visualisation or continuous control of instruments. Of these varieties of project, the third and fourth obviously require far more complex institutional arrangements than the first two.

Second, there is considerable variety in the social contexts in which e-science projects are undertaken. This again makes the design of general models for the institutional infrastructure of such projects very difficult. Potential areas of dispute between the parties to an e-science project, or between those parties and outsiders, involve: (i) the materials and personnel that each party brings to a collaboration, (ii) the allocation of the resources, if any, to which the collaborative project will give rise, and (iii) the apportionment of any liability that the collaborative project might incur. Which of these potential disputes is more likely to hinder the establishment, or effectiveness, of any particular project depends upon a variety of factors that vary with the social context in which the project is undertaken. Thus, it depends upon the *scientific context* of the project. Ethical issues that central to the feasibility of an e-health project, will be less important to a project in astrophysics. It also depends upon the *normative context* in which the project is undertaken. A project may be effected by norms of formal law, of what is sometimes known as 'soft-law' (such as institutional policies), and of conventions that operate within particular stakeholder communities (such as conventions regarding the attribution of credit for a particular type of research project). Norms may vary between the countries in which a global research network operates and between the various disciplinary communities involved in a multi-disciplinary research project. Finally, the range of potential disputes to which a given project is likely to give rise depends upon the *institutional context* in which it is undertaken and, in particular, the range of actors involved.

It is worth emphasising the complexity of this last factor. Most obviously, public research institutions and their private sector partners often have very different incentives and very different ways of working. Public sector research is characteristically different from its private sector counterpart. It can involve higher levels of task uncertainty, be conducted by parties who have higher levels of autonomy and be supported by public funds that have a two part structure, with the fact that researchers enjoy a basic salary not dependent upon research outcomes exacerbating the effect of their higher levels of autonomy in determining the goals and timetables of research. And most importantly, public sector scientific research is traditionally "open" in the sense that it facilitates further research by making its findings generally available as soon as possible. This culture of public sector scientific research is obviously under threat, but it may be regarded as appropriate if public sector institutions are the best place to conduct basic research. While that culture survives, it obviously renders the institutional context for developing collaboration agreements such as eDiaMoND, collaboration agreements that normally involve both public and private sector input, one that is particularly challenging.

Further, while the difficulties of public-private partnerships in research are well known, it is not often enough also recognised that different players within the same public sector institutions—scientific departments, legal departments, research services departments, technology transfer sections—are also characteristically operating with different incentives and under different constraints. There are institutional tensions, not only between public sector researchers and their private sector partners, but also between different actors within the public sector institution. So, for example, simply to take account of the interests of 'the university' in designing institutional infrastructures for global research networks is an insufficiently nuanced approach. And to rely upon the advisors of individual public sector institutions (their lawyers and research services offices) to establish the best conditions for collaborative public sector science is foolishness.

The variety of social context in which large scale collaborations take place, renders the establishment of general models for institutional infrastructures extremely challenging.

The third reason that setting a standard for institutional infrastructures in e-science is difficult, concerns the need to achieve model infrastructures that are both sufficiently certain and sufficiently flexible. Large scale collaborations need certainty. And lawyers and managers know that certainty requires careful planning. If all potential disputes are anticipated and principled resolutions of them determined in advance, then, to their way of thinking, institutional arrangements will be optimal. But scientists are likely to take a different approach. They know that anticipating every dispute that might occur is unlikely to foster the type of trust upon which effective collaboration depends. They also know that the world of scientific research is extremely fluid. As researchers move between institutions, as institutional research priorities change and as the nature of a given project develops, the partners to a collaboration are likely to vary. Institutional infrastructures need to be open to partners to a project joining and leaving it. Finally, scientists know that projects develop in ways not quite anticipated at the outset. Thus questions about the use to which material brought to the project will be put and the allocation of resources to which it gives rise are likely to be far more complex than can be adequately anticipated.

Setting standards for institutional infrastructures therefore means finding models that can take account of the variety of e-science projects and the social contexts in which they are undertaken and that can balance the certainty and flexibility required for a successful collaboration. This arguably renders two different approaches institutional infrastructures untenable.

First, solutions involving the imposition of codes should be resisted. Solutions involving codes might be legislative. They might also involve standard form contracts that are simply applied to a wide variety of projects. The beauty of codes is that they reduce transaction costs for individual participants. Scientists don't want to be bothered with institutional arrangements and an off-the-shelf solution seems very attractive. They also increase certainty by offering a standard set of institutional arrangements that can be learned by repeat players. But the danger with solutions involving codes is that they may be inappropriate to the particular context to which they are applied. Moreover, codes tend to have the ossifying effect that standards can have in technology. We are only just beginning to understand how these collaborations might play out, to introduce a set of standard form agreements in this context might hinder the development of appropriate institutional arrangements. Standard form agreements only tend to work well in contexts such as domestic property conveyancing or international shipping in which certainty is paramount and there are a large number of repeat transactions involving a known range of risks.

But second, bespoke arrangements for every new ecsience project ought also be avoided. One of the thing that those involved in the eDiaMoND project faced was the task of negotiating every aspect of their agreement. This was thought to be very burdensome, even for a pilot project. This burden was increased by differences in the attitude to negotiation of different players. Bespoke solutions also reduce certainty. One of the beauties of codes is that, not only their text, but also its interpretation, can be settled within a given community. So, for example, not

only are the texts of the standard form contracts used in much international shipping settled, but the way in which those texts are read by courts has come to be relatively predictable.

If public funding bodies are to promote best practice in the institutional infrastructure for e-science collaborations, something with neither the dangers of codes, nor of bespoke arrangements for each new project, needs to be found. In our circulated paper, Paul and I have argued that the best solution might entail the establishment of a new public actor to promote best practice in institutional arrangements for various types of e-science project. This body would function in effect as a "collaboration service" by addressing itself to three crucial questions. It would do so free of the pressures under which either public sector institutions or their commercial partners operate, and have as its goal the advancement of collaboration in a context in which science is kept as open as possible.

The first question to which such a body would address itself, is that of the legal shape that arrangements for different types of collaboration might take. Such discussion as there has been, has tended to assume that contractual arrangements among the parties to an e-science project are optimal. However, there will be many situations in which a project might best be incorporated as an independent legal entity – a "company." That doesn't imply a profit-making purpose, as the entity could also be a company limited by guarantee operating on a not-for-profit basis with individual scientists acting as guarantors. This may be desirable in circumstances in which issues of potential liability are important; or where there is a need to provide a stable mechanism for the arrival and departure of parties to the research, and for the development of its work. Such a company could also provide a repository for the assets contributed to, or those created by, the project. Alternatively, the collaboration could take legal form as a normal company with shareholdings for the partner institutions, and a mechanism for distributing its profits. In this case individual scientists could be involved as directors of the company giving them a clear hand in controlling its direction. One role for the proposed independent collaboration service, therefore, would be to explore contexts in which incorporation was desirable for a large scale project and to offer advice as to the appropriate form of its Memorandum and Articles of Association.

The second class of questions that our new public actor would address concern the substance of either contractual arrangements or memoranda and articles of association. The independent service could work to build consensus between major research institutions and funding bodies about the ways in which large scale collaborations should be built. It could also develop model contractual clauses or memoranda and articles of association for use in particular contexts. For example, it has been pointed out that although open source licensing of copyright software can be effective in encouraging the distribution of advances in the computing industry, analogous open source licensing of patents in fields such as biotechnology would be much more difficult to achieve. Advocacy for a culture of open science is a priority, but so too is an understanding of which potential solution of the many currently on offer is the one that will best facilitate collaboration, and the development of a body of experience based upon careful legal work to structure licensing in ways that provide and sustain effective access to research tools, materials and published results.

This organisational approach to consensus building, advice and model clause development has already begun in relation to dealing with the problems presented by intellectual property, and particularly copyright, in collaborative science. These are issues that will be addressed by Larry Lessig and John Wilbanks in later presentations to this meeting. Creative Commons, and its recently launched Science Commons project, are particularly exciting developments. However, intellectual property problems are, as we saw with the eDiaMoND project, only one of the very many different types of institutional issues facing the establishment of effective e-science collaborations.

The third area to which this independent public actor would need to attend concerns the resolution of questions about the interpretation and enforcement of arrangements regarding collaborations in ways that are sympathetic to, and make use of the informal codes and behavioural norms of, the particular scientific communities that are involved. e-Science, and especially transnational e-science, must find suitable *fora* for resolving disputes that take into account the often complicated and subtle features of the working cultures of productive collaborative research projects. This might be achieved in the way in which it is achieved in relation to contracts for international shipping—that is, by parties habitually choosing the law of particular jurisdictions as the law governing their relationships and the courts of particular jurisdictions as the fora in which disputes are to be heard. Alternatively, the independent collaboration service that we are proposing might develop specialised dispute resolution fora to deal with issues arising from e-science collaboration agreements. These *fora* would be well situated to acquire particular expertise, and accumulated a body of documented experience regarding the sources of seriously disruptive conflicts, and the means of resolving them a variety of institutional settings, digital information environments and research domains.

In all these activities, the new public actor would need to harness expertise in particular areas both of scientific activity and the social sciences. The role of lawyers committed, not to advancing the cause of those they represent, but the cause of open collaborative science would be particularly crucial. As Paul and I conclude in our circulated paper:

> "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 mobilize and maintain the necessary technical expertise: first, to select and standardize [particular patterns of institutional arrangement] ... and then to assess the performance of the various collaboration [arrangements] 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 realizing the ... promise of [global research networks in] ... e-Science."

Getting the social science of global research networks right is something that we cannot afford, from the outset, to fail to do.