

Organisational and technological challenges of large-scale multi-disciplinary scientific research

Marina Jirotk₂, Elaine Welsh₁, David Gavaghan₂ and Sharon Lloyd₂.

₁Department of Sociology, University of Warwick, Coventry, UK. CV4 7AL.

₂Oxford University Computing Laboratory, Wolfson Building, Parks Rd, Oxford, UK.
OX1 3QD.

Abstract

In recent years, a new approach to scientific endeavour has emerged resulting from the breakdown of the traditional barriers between academic disciplines and the application of technologies across these disciplines. Parallel and interleaved developments in science and technology have provided the potential to transform the way science is done. This new approach has profound implications for the entire research process and presents challenges both in technological design, and in our understanding of this new form of science, for the provision of infrastructure and training, in the organisation of research groups, and in providing suitable research funding mechanisms and reward systems. These challenges need to be addressed if the promise of this new approach is to be fully realised. In this paper, we will draw on preliminary observations of the Integrative Biology project to discuss organisational and technological challenges of large-scale multi-disciplinary scientific research.

1. Introduction

Over the last few decades, innovations in technology have enabled great advances to take place in the scientific research process. New techniques have been developed for probing and analysing matter and organisms, from the sub-atomic to the astronomical scale, generating data and information on an unprecedented scale. These technological developments in information technology and computational capability have in part been shaped by the needs of scientific research. A new approach to scientific endeavour is emerging as a result of the breakdown of traditional barriers between academic disciplines and the application of advanced technologies across those disciplines. This approach characterised as e-Science or e-Research can be seen operating through the emergence of distributed computer networks and computing-intensive forms of 'Big Science' research (Price, 1963; Galison and Hevly, 1992; and Button and Sharrock 1998). Though the emergence of this new endeavour is clearer in some scientific disciplines than in others, the first fruits of this new approach are beginning to be seen in, for example, genomics, (bio-) nanotechnology, and systems biology.

In this paper, we draw upon a discipline where a most remarkable transformation has recently taken place – biology. Since the late 19th century when traditional disciplinary boundaries

were more formally institutionalised (Kohler, 1982), biology has primarily been a discipline of description and classification, with the development of underpinning quantitative (mathematical) descriptions being limited by the sheer complexity of biological systems. This is now changing very rapidly, and with the completion of the sequencing of the human and other genomes over the last five years, the primary goal of post-genomic research in the life sciences has now shifted to the determination of biological function – how and why do the processes that together constitute a living organism arise from the constituent parts (fundamentally atoms and molecules) that make up that organism? These processes are sufficiently complex to require large, often international, and always multi- and inter-disciplinary teams if progress is to be made.

This paper reports on our experience of one such large-scale interdisciplinary project in this field – the Integrative Biology (IB) Project. The initial experiences in setting up the IB Project and the lessons learned from determining the issues to be explored have suggested that this new approach has profound implications for the entire research process. These issues, which raise challenges in the provision of infrastructure and training, in the organisation of research groups, and in the provision of suitable research funding mechanisms and reward systems, are related to a series of

concerns in the social sciences, most particularly the relationship between technological development, organisational change and social context. The issues raised from the IB project, and the e-Science Programme more generally, resonate with previous attempts to transform organisational work and practices through technological innovation, most notably those that have both enhanced and attempted to encourage new forms of collaboration and communication. We shall draw upon this literature to discuss the problems that were encountered when the new technology was deployed in organisational contexts and how the utopian visions for the technology often ignored the detailed understandings of organisations and social practices that are fundamental to the widespread acceptance of a new technology. Thus we intend to discuss the challenges that must be overcome to enable the e-Science programme to create new forms of scientific practice. In the following we will draw upon this research to examine the implications of this new approach for the ways in which we currently undertake research. We shall examine what is required of an IT infrastructure to support international virtual organisations collaborating across organisational and national boundaries and shall offer initial suggestions of how to train future generations of scientists to work in this new research structure.

2. The Integrative Biology Project

The Integrative Biology (IB), funded by the Engineering and Physical Sciences Research Council is a very large-scale international project with researchers drawn from a wide range of disciplines (including computer science, mathematics, medical and software engineering, biophysics, biochemistry, physiology, genetics, molecular biology, and several areas of clinical medicine). The primary aim of the IB project is the development of the IT or Grid infrastructure to support the entire research process of integrative systems biology - from experimentally derived hypotheses, through the model-building process and HPC-enabled simulation, to experimental and simulation data capture, storage and analysis, and on to model validation and the subsequent design of new wet lab and in-silico experiments. To determine the requirements for this infrastructure, the IB project has chosen to focus its initial efforts on the needs of two clinical areas, cardiovascular disease and cancer, which together account for over 60% of all UK deaths.

Internationally leading experimental groups in each of these areas are members of the IB consortium. These two application areas are complementary in terms both of modelling - each involves multi-scale modelling of a complex biological system - and in terms of the required Grid infrastructure. The modelling of the human heart is the ideal test-bed for building such a system since it is in this area of physiology that the integrative approach is most mature, with the seminal paper, in the area of cellular modelling grounded in detailed experimental work, dating back to the early 1960's with the pioneering work of Denis Noble (Noble, 1962). Over the intervening decades, an international heart modelling community has built upon these foundations so that it is now possible to simulate both normal and abnormal physiology integrating effects from the molecular to the whole-organ level (Kohl et al, 2000, Noble 2002).

Cancer modelling has been chosen as the second application area, since, although there is a large cancer modelling community in the UK, there has as yet been no concerted attempt to take an integrative approach. The aim of the IB project is therefore to support this community in its initial attempts at building a comprehensive model of cancer development across multiple spatial and temporal scales.

The primary goal of the IB project, then, is to develop a virtual research environment, based on state-of-the-art Grid technologies. This environment will be used to support very complex and large scale research activities undertaken by international virtual organisations spread across three continents and drawn from multiple scientific disciplines. In this, it is typical of the "Big Science" research paradigm.

3. Engaging a user community in IB

A core aspect of the IB project is the development of suitable individual and collaborative tools to support the scientific process and working practices. In developing such tools and a technical framework to support their work, there is a need for the technologists to determine requirements for collaboration as part of a suitable development lifecycle. Where systems are developed based on assumptions, the process of acceptance and deployment is considerably extended and development undergoes extensive reiteration. Furthermore, even in the absence of specific problems with the delivered solution, user acceptance may be compromised if they have not had sufficient participation in the initial design, or at least to

an initial prototype. Bearing this in mind, IB decided at the outset that engaging with potential users early on in the project was critical to the success of technology delivery. Development was iterative in nature starting with objectives, vision and clearly defined roles and responsibilities for the project teams through the definition of a generally agreed project structure. Successful communication amongst the various project participants was considered essential to the success of the project.

As a second-generation project expected to utilise results from previous projects, an integration project and a production system development project, Integrative Biology raises a number of interesting project process and requirements gathering issues. The following sections describe the evolving process IB has followed and how user engagement and continual assessment of the scientific progression are important parts of this process.

3.1 Preliminary observations on collaboration in IB

IB benefits from a large consortium of scientific investigators from the UK, US, Auckland and Europe. These investigators have recruited both students and research assistants onto the project to support the scientific development and to strengthen the user team. These researchers typically work on their individual scientific agendas, or within a local research team but have extended their work to include new collaborations with global partners. The UK and Auckland heart modelling teams have collaborated over several decades to develop detailed, computationally intensive models that will eventually describe how biological systems function at all levels from the molecular to complete organs. These collaborations have been formed through the sharing of results and ideas and in some instances, an element of the scientific agenda for each of the participants. Researchers frequently work with the collaborating parties through funded extended visits and occasionally researchers are recruited between groups. Since the project started in February 2004, the project has attracted extensive interest from other heart modelling groups in both Europe and the USA. These groups have shown an interest in both benefiting from the tools and services expected to be made available through the development of the technology, and also from building joint scientific programmes with other heart modelling groups. The cancer modellers have

yet to develop such collaborative working experiences.

On the surface, the project appears to have a strong user community able to drive the requirements for building such a technology solution to enable them to advance their science. However, closer examination suggests that an element of prioritising plays a major part in the commitment to the development of the technology. Resonating, in part, with Grudin's findings (Grudin 1988), in the IB project, where researchers are either fully funded or expected to immediately benefit from engagement with the technologists, the commitment is clearly greater. An additional concept identified is that of career levels and the stronger willingness to collaborate once researchers have an extensive publication record and thus, a niche research area. An additional point to make is that the co-investigators and their researchers are the developers of these models, not clinical users of these models. Ultimately Integrative Biology may also need to consider the practices of a further group of users. These potential new users may exist within pharmaceutical companies or hospitals and may have very different needs. Our existing users range from those who develop models utilising tools like Matlab, to those who have developed their own modelling suites in Fortran or Delphi. The technical ability ranges widely from those comfortable with developing complex parallel codes to others who are keen to utilise tools and services developed by the Integrative Biology technologists.

As a second round e-Science project, Integrative Biology proposed utilising the results of previous projects. This proposition assumed that the outputs from these projects partly met the requirements of a potential user base thus defining a potential starting point for prototypes. Whilst this approach promised to enable the project to release early prototypes to users for evaluation and feedback, the overlap between these projects is in some cases extensive. Coupled with a variation in the choice of standards and technologies adopted across these contributing projects, the challenges for developing a prototype, which meets the users' requirements are extensive.

A key part of the IB project is having a user community with key 'brokers' who are bridging the gap between the pure science and the pure technology development, where these users attend technical design and review meetings and present the use of the developed tools at technology conferences. Continuing to foster

these relationships will be crucial going forward to ensure wider buy-in from new users.

4. Organisational Challenges

Our initial experience of attempting to facilitate collaboration within the IB project has met with mixed success. In part this may be accounted for by considering the ways in which organisational structures in universities, laboratories and research groups are currently produced and how they fit with e-Research. At the heart of a broadly defined academic culture is individual merit. A career hierarchy may often foster competition amongst individuals and between Universities, in which individuals often develop a particular area of expertise within the boundaries of a specific discipline, and rewards are given to individuals rather than groups of people. Clearly, there are opportunities to work with, and communicate research results to, others in a research community. Fundamentally, though, much contemporary research is organised by, and rewards given to, individuals rather than groups of people working together towards a shared goal. This approach to academic life continues to be dominant despite recent developments in inter-disciplinary science and the parallel commercialisation of the results of much of this scientific endeavour (Owen-Smith & Powell 2001). Our preliminary analysis suggests that the 'Big Science' approach to multi-disciplinary research, of necessity, compromises this model of academic work, and instead requires individuals to think about scientific problems across disciplinary boundaries, whilst working with colleagues – on an equal, rather than hierarchical basis – to solve these problems. This can be illustrated by examples where scientists can see a benefit in meshing their research activity with that of colleagues to advance the science further, to the advantage of both parties, or where the technologists aim to utilise the results or outputs from existing projects as part of their solution ensuring that they do not have to repeat existing work. By doing so, the aim in both cases is to leverage existing results to move forward at a faster pace and by doing so, each collaborating party benefits from this engagement. For example, in the TBioSim project, researchers from the Universities of Bristol, Southampton, Oxford and UCL have individual expertise in biological simulation ranging from the quantum scale, through molecular dynamics, to the meso-scale. By bringing together their expertise, aided by infrastructure developments in Oxford and

Manchester, they are aiming to gain a detailed understanding of the interactions that take place between drug molecules and membrane proteins.

The move towards a systems, rather than a reductive, approach to biological research has created a demand for large research teams, where the expertise is distributed in time and space around the world. Producing and creating new science within such large research teams (such as IB) relies upon issues of interdisciplinarity, cooperation and trust. In IB, existing long term individual and group collaborations have fostered relationships of interpersonal trust. These relationships have taken years to develop. Nascent collaborations between less mature research groups need to first overcome perceived trust barriers before detailed working arrangements can be determined. Further research is needed to identify the ways in which interpersonal trust is produced and maintained in such collaborations in order to determine whether some of the essential features may be designed into the technology itself (Jirotko et al forthcoming; Hartswood et al 2005).

In the e-Research vision, researchers are expected to work together on projects that cut across national, institutional and cultural boundaries in unprecedented numbers. Though computational capability may provide technological support for researchers to collaborate to solve particular scientific problems, they face the challenges of working within international and interdisciplinary teams, whilst simultaneously working within the constraints of a particular local organisational context. Consider for example, the work of the cancer modellers on the IB project who are working towards a common objective for this exemplar on the project. Initial meetings with these modellers were focused on identifying any overlaps between the research of each contributing team and on determining unique elements for each to ensure competition between the groups was minimised and joint research was encouraged. This aspect was also visible in the development activity where existing projects were coming together to produce a combined solution, but in many cases, the solutions overlapped in terms of functionality, thus requiring difficult decisions in terms of whose software was actually used.

At the centre of the e-Research vision lies co-operation and collaboration; the traditional career structure which rewards individual merit may have to be re-thought to take account of the different kinds of achievements that active

members of interdisciplinary research groups accomplish. Individual merit may be replaced by group merit where researchers are required to think outside the traditional disciplinary boundaries and to make intellectual links between their own area of expertise and that of others.

The international nature of this approach to large-scale research means that it will be necessary to develop systems of reward that do not disadvantage those involved in interdisciplinary, international and inter-group research. We need to consider the implications of viewing the academic research community from a global perspective, and also reward it from such a perspective. Our funding opportunities, in the UK have begun to take this new approach on board and are actively encouraging inter-disciplinary research.

However, some of our other mechanisms compromise this. The Research Assessment Exercise (RAE) assesses research output within 'units of assessment' that are broadly discipline, and thus department, based. The funding formula encourages departments to develop critical mass within these units of assessment rather than also encouraging it between them. Additionally, departments within a single unit of assessment compete with each other for funding; it is in the interests of departments to have as few other departments as possible being awarded the highest possible grade as the more units of assessment awarded a higher grade, the smaller share of the overall financial pot each department will be given (Bessant et al, 2003). Thus, prestige gained from the RAE rests on individual, discipline specific research that may cause difficulties for individuals working in this new research paradigm.

e-Research also presents certain educational challenges. With inter-disciplinary research and group research expanding, it may be that a different kind of PhD is needed to provide researchers educated in this new way of working. The current structure does not encourage a researcher trying to develop an innovative way of modelling the heart as a basis for their thesis, to work in conjunction with other researchers for fear of diluting the impact of their work. The preliminary observations of collaboration in IB suggest that where researchers do not see any immediate benefits to collaborating they continue to work primarily alone. Though small and large scale science are mutually complementary and will always remain so, we do need to consider how to organise our research communities to maximise the potential for cutting-edge research output in

both fundamental small-scale science, and in the new large-scale interdisciplinary science. The IB project is attempting to address this issue in collaboration with Oxford's EPSRC-funded Life Sciences Interface Doctoral Training Centre (see www.lsi.ox.ac.uk). 8 PhD students funded through IB have spent the first year of their PhD studies in an intensive training programme that has given them the opportunity to develop the necessary biological and practical (generic) research skills to be able to contribute fully to an interdisciplinary research project such as IB. In particular, it is becoming increasingly acknowledged that students need to be exposed to, and gain a knowledge of, different scientific disciplines. In IB, there is a need for our potential users of high performance computing services to be able to develop models that migrate to such platforms. With a background in mathematical modelling, should we expect these researchers to also be skilled in creating MPI code which runs on hundred of processors? Can the science progress without these skills? How can a researcher validate their models without an in-depth understanding of the experimental science from which these models have been developed? These issues are being addressed in a specially developed module at Oxford for the IB students entitled 'Software Engineering for Complex Systems Biology'. Over the course of their PhD programme students need to gain an awareness of, and respect for, different disciplinary cultures and to see the possibilities of drawing on different disciplines to further collaborative scientific research.

5. Studies of collaborative technologies and organisational change

Collaborative scientific research practices shape and in turn are shaped by new technologies. The success of e-Research relies fundamentally upon collaboration, and the design and deployment of technologies and infrastructures developed to facilitate collaboration across both local and global communities of scientists. Previous literature analysing the relationship between new technology and organisational change highlight the many claims made that new technologies will transform, quite radically, how organisations are structured and carry out their business (see for example Hiltz and Turoff 1978; Davidow and Malone 1992). When such predictions have been examined more closely, the anticipated changes have not been realised. Conclusions suggest that the relationship between technological development and

organisational change is not clear (Brown and Duguid 2000). Understanding organisational change requires detailed understanding of organisational cultures and practices regarding how individuals and organisations collaborate. Similarly, in fields such as Computer Supported Cooperative Work (CSCW), analyses of particular technologies in use (see for example, Olsen et al 1985, Heath et al 2000) have suggested a lack of sufficiently detailed accounts of how certain types of work activities to do with collaboration and communication are produced in organisations. The implications of findings from CSCW for the development of advanced technologies for large-scale collaborations have yet to be debated. e-Research systems attempt to support large-scale collaboration - and, these interactions are increasingly about supporting real-time distributed collaboration over the output of the scientific work – data, simulations, models, graphs, global networks of streaming data from environmental sensors to name but a few. We need to understand how to support such interactions over these artefacts and how to represent data collected from one set of scientific studies to be relevant to other scientific domains. Though doing ‘Big Science’ may yet transform existing practices in some scientific settings, we will need to understand the details of scientific practices and tacit knowledge in order to inform the design of the advanced technologies that will enable e-Research to flourish. Within the IB project, there are a wide range of scientists with a vast set of skills and preferences. Different scientists have for example, preferred visualisation packages that have been developed by the groups themselves and from which they are unwilling to depart at present. By working with the research groups, we have understood how tightly coupled their work is with these tools and the extent to which they can rely on them to be effective and efficient in the work they are undertaking. IB solutions therefore have to consider these constraints to ensure that the work of the scientists is not negatively impacted by the use of the IB technology. Our work on the development of a virtual research environment for these scientists will look closely at the day to day working practises, including the management of publications, the policies for data management and the way in which these scientists interact.

6. Conclusion

At present there is little research evidence on which to draw firm conclusions about the working practices of “Big Science” research groups. We need to know what factors enhance this research approach and what may act as inhibitors to its function. From such studies we might determine what opportunities emerge from the formation of new scientific communities and the factors to put in place to promote them both organisationally and technologically. In response to these and other research issues that have emerged from the IB and related e-Science projects, the Oxford e-Social Science project (OeSS): Ethical, Legal and Institutional Dynamics of Grid Enabled e-Sciences (ESRC) has recently formed to address some of the areas outlined above. For example, drawing upon a combination of disciplines such as Law, Bio-ethics, Social Science and Computer Science, current policies and regulations as they apply to the possibilities enabled by e-Research will be identified, and possible suggestions for modifications or new policy introductions may be recommended to the relevant institutions. Findings from this project will also inform the e-Research community more generally.

Many challenges lie ahead. In the context of the issues raised in this paper we need to continue investigations into the best ways to train scientists in the new approach whilst also taking into account the skills of current scientific training, particularly where tacit knowledge and information is communicated. Most critically we must understand how the new forms of scientific collaborations will impact on the scientific record, and the ways in which new scientific practices will determine the acceptance or otherwise of scientific studies.

7. References

- Bessant, J; Birley, S; Cooper, C; Dawson, S; Gennard, J; Gardiner, M; Gray, A; Jones, P; Mayer, C; McGee, J; Pidd, M; Rowley, G; Saunders, J; Stark, A. (2003). The State of the Field in UK Management Research: Reflections of the Research Assessment Exercise (RAE) Panel. *British Journal of Management*, 14, 51-68.
- Button, G. and Sharrock, W. (1998). The Organizational Accountability of Technological Work. *Social Studies of Science*. 28 (1): 73-102.
- Clark, B.R. (2004). *Sustaining Change in Universities: Continuities in case studies and concepts*. Maidenhead: Open University Press.

- Davidow, William H., and Michael S. Malone. (1992). *The virtual corporation*. New York, NY: Edward Burlingame Books/Harper Business.
- Galison, P. and Hevly, B. (eds). (1992). *Big Science: The Growth of Large-Scale Research*. California: Stanford University Press.
- Grudin, J. (1988) Why CSCW applications fail: problems in the design and evaluation of organizational interfaces. In the Proceedings of the 1988 ACM conference on Computer-supported Cooperative Work. Portland, Oregon. USA.. Pages: 85 – 93.
- Hartwood,, M., Ho, K., Procter, R., Slack, R. and Voss, A. (2005) Etiquettes of data Sharing in Healthcare and Healthcare Research. In Proceedings of International Conference on e-Social Science. Manchester UK. June 2005.
- Heath, C. and Luff, P. (2000). *Technology in Action*. Cambridge: Cambridge University Press.
- Hiltz, S. R. and Turoff, M. (1993). (first published 1978) *The Network Nation: Human Communication via Computer*: second edition. Cambridge, MA: MIT Press.
- Jirotko, M., Procter, R., Hartwood, M., Slack, R., Simpson, A., Coopmans, C., Hinds, C. and Voss, A. (forthcoming). Collaboration and Trust in Healthcare Innovation: The eDiaMoND Case Study. In the International Journal of Computer Supported Cooperative Work. Springer
- Kohl P, Noble D, Winslow R & Hunter PJ. (2000). Computational modelling of biological systems: tools and visions. *Philosophical Transactions of the Royal Society A* 358: 579-610.
- Kohler, R.E. (1982). *From Medical Chemistry to Biochemistry*. Cambridge: Cambridge University Press.
- McLennan, G. (2003). Sociology's Complexity. *Sociology* 37 (3): 547-564.
- Noble D. (1962). A modification of the Hodgkin-Huxley equations applicable to Purkinje fibre action and pacemaker potentials. *Journal of Physiology* 160: 317-352.
- Noble D. (2002). The Rise of Computational Biology. *Nature Reviews Molecular Cell Biology* 3: 460-463.
- Olson, J., Olson, G., and Meader, D. (1995). What mix of video and audio is useful for small groups doing remote real-time design work? Proceedings of the SIGCHI conference on Human Factors in computing systems.
- Price, D.J. (1963). *Little Science, big science*. New York: Columbia University Press.
- Seely Brown, J. and Duguid, P. (2000). *The Social Life of Information*. Harvard Business School Press.