Prometheus, Vol. 18, No. 4, 2000

Emerging Bioinformatic Networks: Contesting the Public Meaning of Private and the Private Meaning of Public

NIK BROWN & BRIAN RAPPERT

ABSTRACT *This paper explores the complexity of public/private identities in the emerging global economies of gene sequence mapping and analysis. In so doing we seek to offer a less over-determined acccount of what it means to describe institutional actors as either 'public' or 'private'. Instead, these 'codes' can be seen to offer actors a means of mutual positioning that, more usually conceals broader interdependencies within the world's bioinformatics networks.*

Keywords: bioinformatics, public and private, access, gene sequencing, Human Genome Project.

Introduction

Bioinformatics, the combination of genomics and advanced computing, is increasingly central to research in the 'new genetics'. The sequencing, storage and retrieval of genetic data, together with new silicone-based genechips, have all generated new possibilities for understanding the medical significance of genes and proteins. As part of a wider 'revolution' in genetics, innovators in bioinformatics promise an expanding range of drug candidates and reductions in adverse response rates. Recently, France,[1] Japan,[2] and Germany,[3] have announced major initiatives in bioinformatics. In the short term, the promise of this technology depends on completing the sequence map of the human genome through large public and commercial scale ventures, including the Human Genome Project (HGP) and more complex gene functionality and protein structure research.

That said, bioinformatics creates novel and pressing demands which represent radical departures from established ways of conducting research, communicating findings and producing therapies. This involves managing knowledge within highly accelerated computational systems requiring actors to pursue interdependent strategies across diverse global settings, including public sector research connected to the HGP as well as private sector activity in the pharmaceutical industry.[4]

Bioinformatics follows in the wake of many controversies surrounding the ownership and control of genetic information; for example, the patenting of DNA sequences, or the commodification of the cell-lines of indigenous peoples in the Human Genome Diversity Project.[5] There are continuing, and ever more acute, tensions relating to access, expertise, funding and intellectual property. Given the dual tendencies of greater

ISSN 0810-9028 print/ISSN 1470-1030 online/00/040437-16 © 2000 Taylor & Francis Ltd
DOI: 10.1080/0810902002000853 2

interdependence and tensions between public and private forms of research activity, appropriate public policy formation has become increasingly more difficult to establish.

The fear that bioinformatics and related areas of genetics research are becoming 'privatised' and thus less responsive to the 'public good' has rarely been far from the surface of commentaries about these fields. Whether or not this is the case, it is apparent that the boundaries of 'public' and 'private' science in bioinformatics are less than clear. Although concerns over public–private relations are not new, these have been the subject of more debate in recent years because of commercialisation pressures, changing rationales for the public funding of research, and the impact these dynamics will have on the circulation of knowledge. For example, a blurring of boundaries through commercial strategies in public research can disorientate existing policy by confusing the rationale for a public research capacity. Open science, often equated with public sector research, is supposed to produce public goods that are 'freely available'. However, pressures for economic exploitation of research experienced by academics and others give reason to question such a premise. This also has implications for the structures of ownership and reward.

The basis of the relationship between public and private sector science is continually negotiated between a wide range of social groups, raising questions of what distinctions exist (regarding roles, functions, and operations) and what relevance such distinctions have in the present political and economic environment. The philosophy of free access, for instance, has tended to direct much of the thinking behind policy making in genomics and the HGP more generally.[6] Operationalising this philosophy in practice, however, has been far from straightforward. To a large extent, this difficulty derives from the policies of many countries to capitalise commercially on public sector research. The anxiety felt over the need to exploit scientific research has given a significant momentum to restructuring and commercialisation policies. Publicly funded research is increasingly seen as an investment, expected to yield an economic benefit. As we will see in the case of bioinformatics, this has not always led to the establishment of coherent policies. Therefore, researchers in the field can face contradictory pressures in deciding the research and commercial paths they pursue.

Bioinformatics is but one instance of much wider tensions in the public and private status of knowledge and has attracted considerable interest from academics about how to conceptualise such formations.[7] This paper seeks to extend this literature by examining disputes about the distinction between 'public' and 'private' in bioinformatics in the light of constructivist arguments about the status of knowledge. Our purpose is to identify the major sources of tension and what these suggest for the purpose, meaning, and legitimacy of public investment in bioinformatic research. A number of key questions will be addressed.

- To what extent are bioinformatic data being made openly available and what are the criteria for judging open access? What constitutes research as a 'public science'?
- What are the current and likely future points of friction in the relationship between 'public' and 'private' sector databases?
- How do various actors create and exploit divisions between public and private identities?
- What function does publicly funded research play in fostering innovation by promoting the circulation of knowledge?

These questions are approached by covering a number of substantive areas, including the use and accessibility of databases and the ownership and appropriation of research in genetics.

Much of the argument presented here is derived from a 1999 report by the authors and others to the Science and Technological Options Assessment (STOA) Unit of the European Parliament.[8] For this report, a number of site visits and interviews were conducted in early 1999 with key members of the public and commercial European bioinformatics community, as was a survey of 40 industrial R&D directors, clinical geneticists, and database managers. The study generated a rich picture of the relationships between public and private bioinformatic actors which both extends and critiques existing accounts in sociological literature. The area of bioinformatics is developing at a rapid pace: relations between different research institutions are constantly open for negotiation, and technological advances render pressing issues faced one day irrelevant the next. This article represents only a snapshot of the developments in bioinformatics. As the authors complete revisions on this article in mid-2000, a good deal of posturing has gone on around the first 'working draft' of the human genome. Whatever the particular alignment between actors and the latest public spin to the HGP, though, the underlying dynamics related to the 'public' and 'private' control of knowledge will remain core to debates about bioinformatics.

Contested Terms—The Private Public and the Public Private

Several recent academic commentaries on the status of knowledge claims demonstrate that the availability of scientific knowledge for 'public good' can, at best, not be presumed. On the contrary, the public status of knowledge has always to be worked for, built and rebuilt since the tendency is seemingly towards increasing commercial pressures on the private protection of knowledge. Neither is there anything inevitable about the classification 'public' or 'private' being attached to pre-given kinds of actors, whether they be sponsored by the state or the market. An incorporated gene sequencing company, for instance, is not exclusively private any more so than a public health genetic diagnostic laboratory is exclusively public.

In this vein of thinking, Callon has argued that scientific knowledge is not a (quasi-) public good in the sense of being inappropriable and nontrial.[9] He advocates replacing the traditional distinction of public and private goods with a consideration of whether scientific activity takes place in local, tightly-bounded networks, or ones that are extended and have relatively free-flowing knowledge dynamics. The formation of extended networks is a necessary condition for public appropriation. Science taking place in local networks, irrespective of whether this involves researchers in the 'public' sector, is private science.

Following this framework, Cambrosio and Keating question the public nature of science and examine the processes by which public science is constructed, maintained, and stabilised by considering the complex infrastructure that enabled, in their case, the circulation of monoclonal antibodies.[10] Likewise, Hilgartner draws on constructivist theory to depict genomic research as a heterogeneous process where questions of whom, when, and under what terms data access takes place are crucial and often contentious.[11] Gieryn notes how it is that private territories of closed research activities and data are structured through the very physical architecture of the buildings in which 'publicly funded' science takes place.[12]

The general thrust of this type of analysis is that the 'public' status of knowledge can never be taken for granted; rather it must be constructed continuously. The discussion section of this paper asks how far this framework should be taken and whether the categories of public and private have a practical value for actors in bioinformatics on the one hand, and sociological accounts of developments in the field on the other. In order

to avoid confusion about the particular meaning of public and private, we draw upon three analytically separate types of public/private distinctions here: publicly and privately funded research, public and private sector research, and public and private science. The latter refers to the status of knowledge in terms of relative degrees of open availability, or what Callon refers to as the circulation of knowledge in local/closed and extended/open networks.

Following on from the themes mentioned above, we presume no simple relationship between different 'public' categories and the type of openness that might be said to follow. Publicly funded science may or may not be in the public domain and it may or may not be readily accessible. Also, it should be recognised that each distinction in itself is not easy to apply in practice. Some bioinformatics institutes located in universities receive funding from a wide range of sources, making their appeal to either public or private identity acutely problematic. Likewise, the growth in commercial activities (e.g. university spin-offs, licensing arrangements) in the 'public' sector frustrates hard and fast distinctions between the public and private sector characterisation. The recent announcement by the French Government to restrict the public release of information from its latest Human Genome Project initiative to ensure commercial benefits, for instance, illustrates the way in which public bodies facilitate a privatisation of research.[13] On the other hand, as we shall see, some new and largely commercial initiatives routinely lay claim to all the values traditionally associated with 'public science'.

That said, such categories continue to be used and have a practical value to actors seeking to create opportunities, negotiate patronage and identify themselves in relation to others. That is not to say that public and private hover outside the practices which produce them, waiting in the wings to be called upon when necessary to settle a dispute about ownership or access. Rather, our discussion will seek to illustrate how these 'codes' are produced in practical and rhetorical ordering practices that are integral to the dynamic innovation of new knowledge emerging in the tensions between actors who strategically trade upon differing conceptions of public and private.

The Contested Status of Gene Sequence Production

Bioinformatics is situated at the intersection of academic, government, and commercial interests, combining multiple sources of funding from research councils, charities, venture banking and private finance.[14] The mixture of different types of organisations in relations that combine competitive and cooperative incentives creates a blurring of roles regarding the structures of ownership, funding, reward, and accountability. Key bioinformatic actors may be characterised according to whether they are research institutions and alliances generating and providing data, whether they are involved in the provision, organisation and distribution of funding arrangements, or whether, like the pharmaceutical industry, they are purchasers of bioinformatic data services. In what follows, taking research formations whose funding is largely from public and charitable sources first, we introduce several key research actors and discuss features which challenge their claim to public or private identities.

In the first place, there are a number of major public research institutes producing genetic data which combine public and private sources of funding. In the context of European public research, the European Molecular Biology Laboratory (EMBL) in Heidelberg is of utmost importance to Europe's standing in genomics and bioinformatics. The EMBL operates three major outstations including the European Bioinformatics Institute (EBI), Europe's largest repository of non-proprietorial gene sequence data. The EBI database is updated every 24 hours with new sequence data from the Sanger Centre

(see below), GenBank in the US and the DNA Database of Japan. It also operates an Industry Support Programme involving more than 20 of the world's largest pharmaceutical firms and is significant in the dissemination of interoperability standards.

The Sanger Centre, having sequenced as much as a third of the total known human genome, is the largest contributor to the HGP. Supported principally by the UK's Wellcome Trust, it exercises no proprietary rights to its data and seeks to prevent other actors from doing so by releasing as much human sequence data as possible. As Fortun rightly notes, the emphasis on speed and urgency, originally to prevent the delay of therapeutic progress, was central to persuading state and charitable sponsors of the legitimacy of significant public investment.[15] More recently, however, the arrival of commercial sequencing competitors provides much of the momentum for increased acceleration and higher investment in the HGP.

Still, publicly or charity-funded research should not be seen purely as a public good. Academics in competitive, high risk, and high reward areas of research do not always share information freely. As universities and research centres assume the entrepreneurial role prescribed to them in much of European and national innovation policy, separating academic from commercial activities will be more and more difficult. In bioinformatics, where the distinction between academic and commercially relevant research is blurred, ample opportunity exists for commercial exploitation. Recently, spin-off companies at University College London and the European Bioinformatics Institute are indicative of the commercial activities in ostensibly 'public' bioinformatics.[16] The formation and operation of such spin-offs involves a delicate process of negotiating the boundaries of 'public' and 'private'. Who controls and who benefits from the intellectual property created in publicly funded settings are key issues.

More generally, Hilgartner has documented the tendency for publicly funded researchers in the Human Genome Project to close off their research from others.[17] Rather than engaging in the sharing of data for 'academic' and commercial reasons, publicly funded researchers failed to make information available. In very few settings have those responsible for the management of exchange policies been able to construct relations between data and material producers and users that facilitate both openness and yet 'ownership'. One positive example of this is the European Commission's Yeast Sequencing Programme, which prescribed clear rules about the production, ownership and exchange of data within and between various data streams.[18] These rules combined incentives regarding ownership rights over certain sequence territories on the map and penalties regarding failure to disclose data where ownership rights would be cancelled. That such policies had to be initiated at all indicates something of the contested public status of publicly funded research. The current debates within the US National Institutes of Health (NIH) over the sharing of research tools further illustrates the difficulties of assuming that academic research is shared widely.[19]

While distinctions between public and private sector actors can be useful for understanding some of the differences and dynamics in terms of the accessibility of information, access is arguably a somewhat crude basis for making distinctions between the two. Following on the points raised above, both public and private actors play a role in making available bioinformatics data as an accessible good, though with differing reasons and on the basis of different terms and time frames. Since its inception, Human Genome Sciences Inc. (HGS), for instance, has been at the centre of disputes about the public status of genes and bioinformatic knowledge. The company was formed in the early 1990s on the basis of research by Craig Venter (formerly of the US National Institutes of Health) on expressed sequence tags (ESTs). After failing to be granted patents on its ESTs, HGS made its data available to academic researchers on

condition they agreed that the company would be granted first refusal rights on any derived products. The subsequent academic backlash at these conditions and the threat of the pharmaceutical company Merck to make a database of ESTs freely available was enough for HGS's parent company to terminate the contract with its subsidiary.

More recently, new controversies are developing that involve many of the same actors and questions, but on a potentially much larger and competitively intense scale. In May 1998, an alliance between Craig Venter and the biological devices manufacturer Perkin-Elmer resulted in the formation of an enterprise to compete with the HGP in creating a sequence map of the human genome by 2001. Venter had applied 'whole genome shotgun sequencing' to the human genome, a strategy which he was prevented from pursuing whilst at the NIH. The strategy involves repeatedly fracturing the DNA of an entire organism into small pieces and matching the bases at the end of each fragment with one another. On the other hand, the more deliberate and arguably slower 'clone by clone' method favoured by contributors to the HGP involves copying short strings of DNA repeatedly within a replicating organism until the sequences within the string have been mapped. Celera, Venter's new company, promises to supersede the efforts of the HGP by as much as 2 years, generating severe criticism of the overall efficiency of the methods upon which the HGP has proceeded to date.[20]

Several significant areas of concern have been raised in relation to Celera, the most serious of which is its capacity to exercise exclusive proprietary control of large portions of human sequence data in copyright and patent law. In response to these perceived risks, the publicly funded parties to the HGP, particularly the Sanger Centre, announced major increases in sequencing capacity and the strategic use of shotgun sequencing (as opposed to whole shotgun sequencing) where appropriate. This includes major resource investments by other actors in the same Elmer Perkin sequencing machines provided by the device's manufacturer to Celera.

On the other hand, Venter sought to circumvent criticisms which might otherwise alienate the research communities of the HGP threatening his and Celera's credibility and scientific prestige. The approach taken is to create an inherently ambiguous public–private status for Celera. In most respects, Venter endeavours to appeal to the codes or criteria which will allow the company to alternate strategically between claims to public/academic and commercial identities. In his words:

> The scientific community thinks this is just a business project, and the business community thinks it's just a science project. The reality is it's both. This is a private company paying to sequence the human genome and give it to the public… There was never any disagreement that if we were going to sequence the human genome that it would be morally wrong to hold that data hostage and keep it secret.[21]

This leaves the question of how it is that the company will be able to create a return for its financiers on their $200m investment while satisfying all the free and open access codes of an academic research identity.

To be sure, the revenue generating methods suggested by Venter have troubled Celera's claims to public status. First, many of the world's major pharmaceutical firms will be paying annual fees of $US5 million each to get early privileged access to the sequence data before the data are released more widely. However, that the data will eventually be made 'publicly' available—under what terms is not clear—means that Venter is still able to appeal to widely held criteria of what counts as public scientific knowledge and reap the 'scientific' credit which derives from that. What is known about the subtleties of these terms is that they are likely to be modelled on that of an information agency selling access to, rather than exclusive ownership of, genetic data. In

late 1999, Celera took the initiative to apply to the US Patent Office for patents on 6500 sequence sites, an application totally disproportionate to the number of gene patents so far granted in the US. Indeed, whether as a direct consequence of this or not, the US Patent Office has now issued revised guidelines refining definitions of utility in genetic-related applications. Nevertheless, both of the above strategies have been taken as signs that Celera has increasingly had to curb both the rhetoric and the practicalities of its early 'academic' aspirations.

However, public and private continue to be deployed and appropriated strategically by organisations, defying easy categorization by other actors in bioinformatics. Indeed, until recently Venter was still in discussions with the NIH over the possibility of depositing sequence data in GenBank, but under certain undisclosed conditions. The NIH has so far declined.[22] More recently still, Celera announced that, upon completion, the sequence map will be made available under a 'non-redistributional arrangement' similar to those covering software licensing agreements. This effectively allows Celera to discriminate between its different users, thus creating its own map of the public–private interface and using that to set access prices accordingly.[23]

The case also illustrates the broader networks across which the categories of public and private are implicated. This can be best seen in the relationship between Celera and its parent company, Perkin–Elmer. On the one hand, the subsidiary has been at pains to signify co-operation with the aims of the HGP. On the other hand, Celera creates incentive pressures for HGP partners to accelerate their sequencing and thus the purchasing of the same state-of-the-art machines Perkin–Elmer has effectively 'given away' to Celera. The commercial dynamics in which Celera is embedded are not exclusively related to simple questions of whether its sequence data will cost consumers or not. Since roughly 98% of the world's sequencing activity is done using Perkin–Elmer machines, the parent company can only benefit from the added revenues created by the actual or perceived threat that Celera represents to HGP collaborators. Indeed, the Sanger Centre and NIH laboratories have both stepped up their procurement of Perkin–Elmer machines, demonstrating the more general point that public and private dynamics become ever more complex as one extends further away from the laboratory and into the wider innovation networks beyond.

The flexibility of public and private identities in bioinformatics data can also be aptly illustrated by the decision of the Icelandic Government to use a private company as a broker in the sale of the Icelandic population's genetic register. The company concerned is the Delaware firm, DeCode Genetics, which has now entered a fee-based contract with Hoffman-La-Roche, giving the company access rights to the database. The small size and internal homogeneity of the Icelandic population, together with the existence of an exhaustive public health record extending back over many years, makes the register a valuable asset in tracing rare and common genetic disorders.

If there were any doubt over the unprecedented reconfiguration of public and private by the Icelandic case, its import is not lost on Iceland's Surgeon General: 'I don't think this country can just sit here and say, "No, sorry, we are going to stand on rules that existed in a different era for a different wor1d" '[24] The 'difference' to which this statement alludes largely rests on translations of the relative virtues of public and private. For example, the privatisation of genetic knowledge will be justified on the basis of the aggregate 'public good' likely to benefit the people of Iceland in terms of immediate health care revenues and longer term access to cheaper and more varied pharmaceuticals. The legal justification of commercial sale also follows the logic that the register is not owned by individual patients or institutions since, over many years, its production has been paid for by public funds.[25] Illustrating the plasticity of public/private codes, the

same line of reasoning has more usually been reversed in arguments that publicly funded data should not be property. In fact, this latter term of reference is the key focus of debate surrounding the Icelandic legislation.[26]

Whilst there can be no mistake that there are specific aspects to the Icelandic case, developments elsewhere have seriously called into question its assumed distinctiveness. In December 1999, the UK Court of Appeal overturned an earlier High Court ruling that the secondary use of anonymised patient data for commercial development had breached confidentiality.[27] The challenge, against the Department of Health, was initiated by a suitably hybrid public–private consortium including Source Informatics, representative organisations of the pharmaceutical industry, the General Medical Council and the Medical Research Council. The judgement applies to two databases, the UK Research Database managed by the Medicines Controls Agency, and the UK Primary Care Database, ownership of which has passed from Source Informatics to IMS Health. The last of these supplies data on over two million patients to pharmaceutical firms.[28] Controversially, the new ruling asserts that since the data is non-patient-identifiable, consent for secondary use is not required. This is but a small illustration of much more general lobbying to persuade government and health providers that patient records represent a significant, though as yet insufficiently utilised, commercial research resource.[29] The terms under which this information should be available will require negotiation, and actions will increasingly depend upon reformulating the terms under which records are controlled, including a redefinition of ownership, informed consent and privacy. Respondents to our study from the pharmaceutical sector were unanimous in pointing to the Icelandic case as a model of how the Department of Health could proceed in making population genetic data more readily available.

Again, the combinations of actors involved in these kinds of initiatives are highly telling of the plasticity of public/private identities, even for institutions that are otherwise seen as vociferous defenders of the public status of genetic information. For instance, the UK Medical Research Council and the Wellcome Trust have recently announced that work is to begin on two genetic surveys, the scale of which, involving more than 500,000 patients, is unprecedented in the UK. This, together with the realisation of the UK's Electronic Patient Record, are seen as necessary precursors to a commercially run database that will service both 'public health and industrial research'.[30]

When considering the status of global gene sequence producers, 'public' and 'private' serve as idealised codes to which various actors, whether they are universities or commercially funded research initiatives, can appeal (though with varying degrees of success). Indeed, it is their very instability that enables actors to appeal simultaneously to one or other of the terms. As a consequence, policymaking is beset with the problem of determining what role 'public' research and funding in particular should play given the increasing fluidity of these identities. We will return to this problem in the discussion after exploring in more detail the uses made of bioinformatic data by industrial actors. In particular, we are interested in how use reflects back upon the public/private character of genetic data provision.

The Contested Status of Bioinformatic Consumers

As the Icelandic case illustrates, commercial actors play a major role in bioinformatics, whether that be in basic sequencing or in enabling pharmaceutical firms to integrate genomic research within their portfolio. The development of genetic diagnostics requires tapping into large repositories of genetic information that need to be customised to differing information sourcing requirements. Most major pharmaceutical firms have

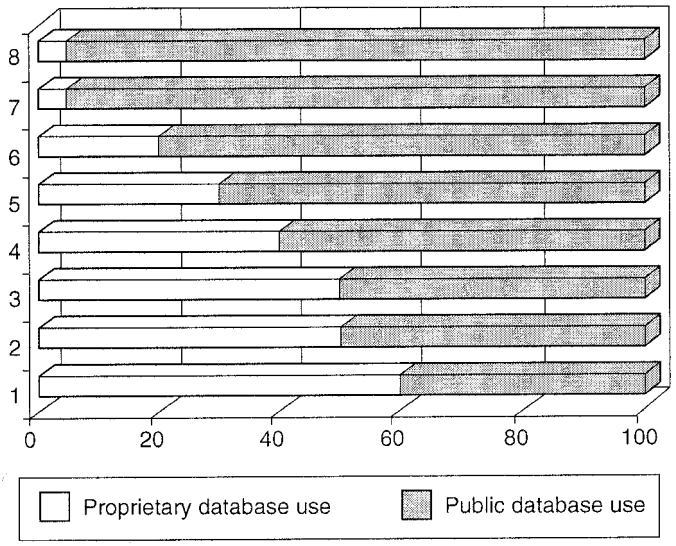

Figure 1. Percentage use of proprietary to non-proprietary databases by eight pharmaceutical firms.

formal arrangements for accessing and sourcing repositories from various sources, sometimes for free, sometimes for a fee. This includes in-house copies of databases (mirror sites) permitting speedy and confidential access to commercially valuable sequence information. Incyte Pharmaceuticals Inc. is one of the largest and most heavily subscribed proprietary databases. The Incyte databases had 17 subscribers in 1997, each paying as much as $US15–25 million per annum.[31] The sums of money spent in this area should not be seen as an indication that the information serves well defined functions. A good deal of uncertainty still exists in determining how bioinformatics can be put to use. SmithKline Beecham, for instance, withdrew its contract with the Human Genome Science (HGS) database (worth $US140 million) because the company needed more time to assess the significance of the data that it had 'mined'.

One important consideration is the relative use of proprietary and non-proprietary genetic databases by different actors. Each type of database is valued for different reasons and is used strategically according to the R&D interests of bioinformatics consumers. For example, the fact that many public databases are updated every day and are easily accessible is generally highly valued by the research community as a whole. Commercial constituencies, including proprietary database providers themselves, have depended heavily upon much of the 'public' sequencing initiatives, which, ironically, have made possible the 'privatisation' of genetic knowledge. Our study of database use by pharmaceutical firms showed a tendency for actors to draw heavily on free access databases, combined with the targeted use of commercial sequencing databases where appropriate. Of the major pharmaceutical companies surveyed as part of the STOA report, only one firm showed a higher rate of use for proprietary over non-proprietary sources. A not insignificant finding is that, without exception, all of the respondent companies have found it necessary to subscribe to private databases (see Figure 1).

Among pharmaceutical respondents there was very strong agreement that the use of public databases is set to increase. While publicly funded databases make a fundamental contribution to therapeutically relevant sequences, subscriptions to their databases will

continue to be important. As one respondent said about the future activities of commercial interests in the field:

> Our company data is currently proprietary, but some estimate that most of it will be in the public domain inside two years [because of the activities of other researchers]. Innovation of new data products that add further value—e.g. genetic variants, genotype/phenotype correlates—will become the data products of tomorrow and are likely to be only achievable with private investment bolstered by industry subscription in the hunger for new data to build competitive advantage.

Although commentaries about bioinformatics generally recognise a strong need for both publicly and privately funded databases, each serves different functions and operates under unequal terms. The commercial companies we approached for the STOA report were more optimistic about the future of this co-operation than public sector constituencies. There are several reasons for this. The first relates to the unequal share of investments put into staff development. When public sector actors move into companies, whether through the creation of firms or through the movement of personnel, they take with them both expertise and knowledge which needs to be replaced. While in policy terms this might be seen as part of the rationale for public research in its contribution to the knowledge base, actors in universities are more likely to interpret this in terms of having continually to rebuild expertise and of losing needed revenue. Illustrating this, the US NIH are considering a multi-billion dollar agency for funding biocomputing, in part designed to help stem the flow of expertise into industry.[32]

Secondly, respondents from the public sector indicated that whole genome shotgun sequencing, as done by privately owned commercial firms such as the Venter initiative, is a major threat to public databases and to the quality and accessibility of data. The technical rationale is considered to be less accurate than more deliberate clone-by-clone strategies, since it relies so heavily on repetitive sequencing. This is especially the case in large organism research, where a significant proportion of the genome has no known function, making the assembly of a genetic map particularly imprecise. In assessing the contribution of organisations to advancing bioinformatics, the quality of the data is just as important as the speed by which it is gathered.

Thirdly, while our survey of major public and private database providers and pharmaceutical firms in Europe found a reasonable consensus that the cost per sequence is likely to fall considerably over time, this drop was not expected in the short term. Given the costs associated with the bioinformatic research tools, molecular biology is widely held to have assumed the 'big science' scale more usually associated with nuclear physics and space research. In turn, this shifts the scale of the resources necessary for the creation of new research alliances and specialisation, and places demands on research institutions that try to maintain their status in molecular biological fields. The current high costs associated with attaining a competitive advantage through bioinformatics mean larger firms have greater capacities for utilising gene sequencing and access to genetic data for pharmacogenomics and combinatorial chemistry. While the price of producing sequence data is expected to fall, nearly all respondents agreed that the cost and complexity of analytical bioinformatics will continue to rise. Matching these higher R&D costs competitively certainly implies a necessary change in scale, such as that seen in recent merger negotiations between Glaxo-Wellcome and Smithkline-Beecham.

This raises the question of who will be able to tap into bioinformatics sources. High costs are concentrated particularly in: access to skilled (and expensive) bioinformaticians; institutional commitment to high cost investments in new gene sequence and array equipment; and new data platforms for storage and analysis. Most respondents con-

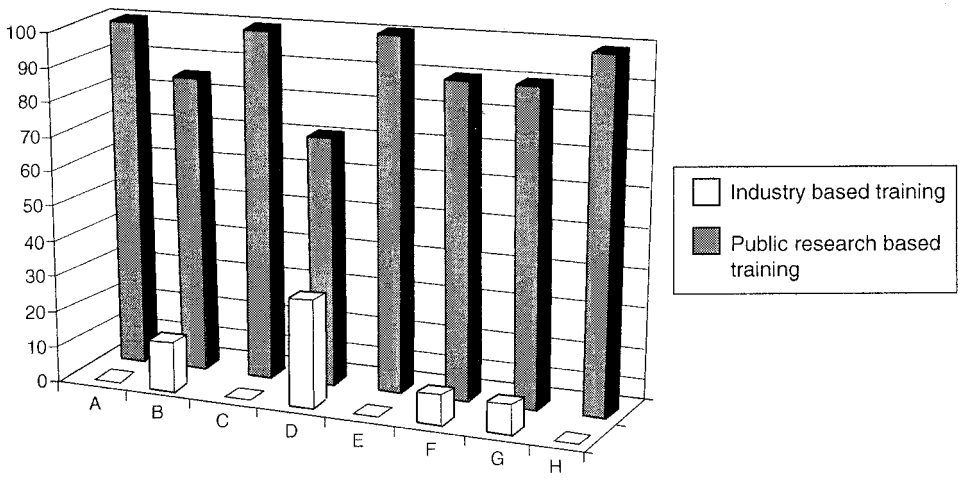

Figure 2. Percentae ratio of public to privately trained bioinformaticians in eight pharmaceutical companies.

sidered that, if left unattended, the relative position of public to commercial research would deteriorate. The description above gives a clear indication of the potential for bioinformatics to increase asymmetries in resources between commercial and public research institutions. This general situation raises important questions about the position of public sector databases.

Fourthly, of major concern to both public and private constituencies is the question of whether public databases will be able to secure funding in the future, especially for maintenance and up-dating software programs. Funding issues are said to be 'biased' in favour of the private sector. This sector profits from the availability of publicly funded data which is often crucial to the development of its commercial products. In questionnaires and site visits, public research actors across Europe were consistently anxious about their capability to maintain a research agenda which was relevant, up to date and able to take advantage of innovative developments in bioinformatics. As with changes in scale evident in commercial pharmacology, public sector diagnostic actors foresaw necessary changes in resource sharing on a much larger, even interstate, basis as the only way to maintain research capacity and service provision. Maintaining the status of public sector institutions is particularly important given their educational role. As Figure 2 indicates, the percentage of bioinformaticians in major pharmaceutical companies who had received public sector training was typically in the region of 90–100%.

The Appropriation of Research?

As already mentioned, one of the main sites where debates about policy, public/private boundaries, and the capturing and capitalising on research have become apparent is in the area of intellectual property rights (IPR). In bioinformatics-related fields like genetic diagnostics, seeking patents requires finding a link between genetic characteristics and disease in a way that conforms to the criteria for patent approval (novelty, utility, and non-obviousness). The commercial pay-off of bioinformatics and the role of IPR in realising this has never been far from the global bioinformatic agenda.

While discussions about the openness of research typically revolve around discussions of intellectual property rights, these should be balanced against wider considerations. The

attractiveness of IPR lies, of course, in its providing inventors with exclusive rights. It would therefore seem reasonable to assume that the exchange of ideas and materials between companies is highly restricted. While no doubt this is the case in the latter stages of R&D, the pharmaceutical industry and related innovation networks are characterised by fairly extensive exchange systems.[33] At a general level, the sector defies an easy characterisation as open or closed. While intellectual property rights and proprietary claims to knowledge are key aspects of ensuring competitiveness, companies in this field are close knit, where the sharing of information is quite common. A substantial regulatory and commercial infrastructure is in place whereby companies acquire information about each other and possible avenues of research, for instance through clinical trials, news agencies, conferences, patent searches, mutual) academic contacts, and testing and trials. As the work of authors such as Hilgartner[34] and Faulkner and Senker[35] has illustrated, a delicate process of negotiation is taking place between what is public and private. Industry often acts as an important source of generic scientific and technological knowledge.[36] Secrecy is, of course, utilised among those in bioinformatics, but this should not be seen as the blocking the flow of information. This sector exemplifies the strategy noted by Hicks that private actors engage in a limited public dissemination of information in order to 'construct the distinction between public and private boundaries in such a way as to provide maximum advantag'.[37] Corporate publishing enables such organisations to participate in the barter exchange relationships of researchers, to establish their credibility, and to signify their competencies.

Much of the European policy discussion about IPR related to bioinformatics has commented on the perceived inadequacies of IPR regimes in Europe vis-á-vis the US. Given the points made above, however, it is perhaps not surprising that most of the individuals contacted as part of this study were not worried that Europe's relatively open access policy would detrimentally affect its ability both to exploit and regulate bioinformatics-based health technologies. This was either because European provisions were not judged to be inferior, or because formal rights were not seen as important. As one of the interviewees claimed, 'I cannot see a US company will inhibit EU research because they patented data that the EU has been giving away freely'. If actors reacted on the issue at all, they saw that a much more fruitful way forward would be to create opportunities for 'intellectual property' claims (taken in the broadest sense of the term) that did not change the open access policy. As one of our interviewees explained, providing free access to information could also be an advantage to researchers. In this particular example, researchers had made available a set of markers that became well used by forensic laboratories all over the world. The broad diffusion of these markers via the Internet provided the researchers with a huge research network and extensive contact with other researchers. This network, according to the researchers, was in itself a very valuable and important resource for further research. The bioinformatics research system is one that mixes symbolic credibility cycles and formal strategies. Commercially motivated companies must find a way of engaging in both to be successful.

The private sector is not homogenous, rather it is characterised by a diversity of appropriation strategies. Bioinformatics companies are not simply servicing larger firms with specialised expertise. Research establishments, such as those in bioinformatics, have shifted their resources from sequencing to developing compounds for specific genomic targets.[38] They have to think of their market location in terms of building up from data to producing diagnostic applications, or from identifying genes and relevant proteins to developing new treatments. To do so successfully requires finding ways of securing their position, in part by seeking formal intellectual protection. Some of those we interviewed

were attempting to seek IPR on the functions of genes, whereas others outside these research establishments felt that securing IPR would require associating a genetic marker with specific clinical outcome. Such clinical determinations are expensive and likely to be undertaken only by or in collaboration with major pharmaceutical firms.

Respondents from within the pharmaceutical industry were divided on whether there were any inhibiting effects of proprietorial claims to sequence data on drug R&D. However, this needs careful qualification not least because it is not necessarily in the interests of such actors to contribute to ongoing public and political reservations about the patenting of sequence data. This then must be balanced against the perspectives of other constituencies indicating perhaps more considerable implications of patenting on genetic innovation. For example, a recent study of genetic screening services in the US found that a quarter had been notified by the law firms of biotechnology companies that the pathologies for which they were screening were now under proprietary protection.[39] As many as 50% had withdrawn one or more services because of substantial licence fee costs associated with being legally obliged to use proprietary diagnostics. Myriad Genetics, for instance, is now compelling screening services to use exclusively its test for BRCA1&2, the gene etiologically implicated in some breast cancers.

Nevertheless, in the broader context of the genomic mapping, many pharmaceutical firms have found it to be in their interests to cooperate on a formal basis with the HGP in preference to seeing increasing pressures from proprietorial bioinformatic actors. An example of this is the consortium between the Wellcome Trust and the pharmaceutical firms to map roughly 15,000 single nucleotide polymorphisms, enabling better understanding of the genetic factors involved in adverse drug responses. The SNP consortium is one such example of where public–private alliances have opportunistically formed to protect what might otherwise seem to be competing interests. In the case of the pharmaceutical industry, this lies in the protection of a market share threatened by the sequencing industry, but combines with the Wellcome Trust's commitment to open and free knowledge transfer. Indeed, the fortunes of companies such as Genset in France have arguably suffered as a result.[40] This latest twist is indicative of the general conflicts over IPR in a closely integrated research system.

Making sense of the dynamics of appropriation between ostensibly public and commercial actors also forces us to reconsider assumptions about assumed inequities. On the one hand, knowledge dynamics in the context of sequence production reflect often expected patterns of dependence that are in many respects asymmetrical. Not surprisingly, since the HGP data are placed in the public domain daily, commercial data providers are able to take advantage without being obliged to reciprocate. Celera's endeavours to map and supersede the HGP have been made possible only by, deriving as much as 80% of their sequence map from the public domain.[41] Further, that Celera continues to wait upon HGP data releases softens claims that it can feasibly 'overtake' the latter. On the other hand, the inequities of this relationship need to be qualified not least because both seem to be generating data at roughly the same pace. In addition, redressing the often cited simple reading of disproportionate advantage, it can be argued that, in terms of the prosperity and unprecedented investment in public sequencing, the HGP has profited considerably from the incentive pressures created through commercial enterprise and the justifiable threat of proprietary appropriation.

In the future these negotiations about the appropriation of knowledge and the structure of the relations between and within public and private arenas are likely to flare up over access to health genetic registers. Probably one of the most striking areas of consensus we found between the constituencies interviewed, particularly in respect to the pharmaceutical industry, was the almost unanimous agreement that R&D will increas-

ingly depend on making arrangements to access public health genetic registers. The increasing importance of patient data being circulated through the global bioinformatic network is readily evident in both the UK and Icelandic cases. Such arrangements highlight the premium placed on new forms of partnership both within the clinical setting and between clinical and non-clinical constituencies as well as the terms (e.g. with regard to ownership and access) under which such collaborations should be undertaken.

Conclusion

Bioinformatics is situated at the intersection of information technology and life sciences, where academic, government, and commercial interests overlap and sometimes merge. Such networks are ever more likely to serve as the future foci for commercial competitiveness as proprietorial capital increasingly shifts towards 'biowealth' investments. These emerging tendencies require new understandings of the relationships between actors and a corresponding reappraisal of the role of public research and, indeed, of the meaning of 'public'.

The mixture of different types of organisations in a setting that combines competitive and co-operation incentives presents problems for the circulation of knowledge and the degree of openness by which bioinformatics might be characterised. This account has shown that the 'free availability' of genomic sequence data as a public good is far from inevitable. Indeed, the public status of knowledge and even its presumed merits have been tested in the context of increasing commercialisation.

A number of factors influence the degree of openness of research, including the source of funding and accompanying terms and conditions for funding, the type of research and its applicability, the extent of codification, publication practices, the character of exchanae relationships. IPR policies, and commercialisation arrangements. It should be clear that examining the source of funding or the site of research or the IPR practices are not in themselves enough to appreciate the accessibility of research. Research takes place in heterogeneous networks that in different ways structure 'openness'. Any discussions about the openness of bioinformatics databases should acknowledge the systemic relation between actors. Public bodies funding research can set rules of dissemination which, however supportive they are in terms of circulating research, actors in commercial organisations do not have to follow.

This paper has sought to illustrate some of the complexities that attend a mapping of innovation dynamics according to public/private typologies. It has been possible to see how, for the purposes of fostering the success of their networks, actors trade upon public/private codes that, upon closer scrutiny, are far less stable and far more contingent than colloquial use in modern accounts of biomedicine would suggest.

The account of bioinformatics presented here illustrates how organisations increasingly combine multiple forms of public/private activity, engaging in both gift and proprietary economies simultaneously. This is most evident in the way Celera's founder, Craig Venter, persistently requests that the wider world, and the public research community specifically, put aside conventional notions of a profiteering private sector at odds with the rationale of public research. Venter's collaborative rhetoric is one of a number of examples where it has been important for commercially funded organisations to present multiple public/private identities in the interests of appeasing potential sources of opposition (whether that be government proprietary legislation, funding policy or wider public support more generally). On the other hand, it has been just as valuable for worldwide contributors to the HGP to present themselves as unquestionably public while engaging in collaborative ventures with industry.

This over-determination of the public/private identity within organisations is just as misleading as it is when used to map the relationships between organisations illustrated by such terms as the public or private 'sector'. The analysis we have presented here has highlighted some of the mutual interdependencies between ostensibly opposed actors that point to the need for a broader framing of networks. Actors may present themselves (or are presented as) mutually antagonistic according to homogeneous accounts of public/private, but it is not the case that conflicts over knowledge are entirely non-reciprocal. The triangulation of the relationship between the HGP, Celera and Elmer Perkin is one such illustration of a broader network where the pace of change and the size of investment necessarily depended on the actions of others. Celera's progress towards its drafting of the human genome has depended heavily on existing HGP sequence data, the recruitment of once-public research personnel, and so on. In turn, public research investment in the HGP has swelled considerably under the threat of the commercial appropriation of the human genome presented by private sequencing activity.

How is public policy in the newly emerging biosciences to respond to changes in public/private composition and definition of activity? As the above discussion illustrates, the question is particularly thorny in the context of a policy regime increasingly confronted with the need to understand better what the practical value of public and private codes are to its own changing role and that of other knowledge producers. From the perspective of public policy and public investment in sequencing, it is still difficult to tell whether some years hence the pressures created by commercial sequencing companies may been seen to be instrumental in mounting an effective policy response and public investment in the HGP. Undoubtedly, the research economy of the HGP itself will be judged to have thrived as a result of the fears associated with the private appropriation of human sequence data. Whether this increased prosperity will be seen to have realised the goals of free access, fewer 'walled gardens' and steady state health costs is a different matter. Indeed, perhaps more likely and even paradoxically, public investment in the HGP may have hastened the proprietary commodification of genetic data and even unwittingly contributed to an escalating crisis in the state financing of healthcare.

Notes and References

1. 'Genomics boom', *Science, 285*, 16 July 1999, p. 309.
2. 'Japan set on five year genome research plan', *Nature*, 400, 22 July 1999, p. 305.
3. 'Germany to boost bioinformatics', *Nature*, 400, 8 July 1999, p. 93.
4. S. Gardner, Z. Paolo and T. Flores, 'The evolution of bioinforatics', Synomics Ltd online papers (ww.synomics.com/about/ev_bio.htm); L. Hwa and T. Butt, 'Bioinformatics takes charge', *TIBTECH*, March 1998, pp. 104–7
5. P. R. Wheale, 'Human genome research and the Human Genome Diversity Project: some ethical issues', in P. R. Wheale, R. von Schomberg and P. Glasner (eds) *The Social Management of Genetic Engineering*, Ashgate, Aldershot, 1998; L. A. Whitt, 'Biocolonialism and the commodification of knowledge', *Science as Culture*, 7,1, 1998, pp. 33–68.
6. National Research Council, Mapping and Sequencing the Human Genome, National Academy Press, Washington, DC, 1988.
7. B. Rappert, *Privatising Research: Intellectual Property and Construction of the 'Public' and the 'Private'*, PhD Dissertation, Anglia University, 1999.
8. N. Brown, A. Webster, A. Nelis, B. Rappert and G. J. V. Ommen, *Bioinformatics: A Technological Assessment of Recent Developments in Bioinformatics and Related Areas of Research and Development including High-throughput Screening and Combinatorial Chemistry*, Report for the Science and Technological Options Assessment (STOA) Unit, European Parliament, May 1999.
9. M. Callon, 'Is science a public good?' *Science, Technology, and Human Values*, 19, 1994, pp. 395–424.

10. A. Cambrosio and P. Keating. 'Monoclonal antibodies: from local to extended networks', in A. Thackray (ed.), *Private Science*, Pennsylvania State University Press, University Park, PA, 1998.
11. S. Hilgartner, 'Data access policy in genome research' in A. Thackray (ed.), *op cit.*
12. T. F. Gieryn, 'Biotechnology's private parts (and some public ones)', in A. Thackray (ed.), *op cit.*
13. D. Butler, 'French plan to exploit genome sparks row', *Nature*, 402, 1999, p. 569.
14. For a more in-depth description of funding see the following DG XII related publications: M. Hallen (ed.), Human Genome Analys Supported under BIOMED 1, IOS Press, Amsterdam 1998; M. Hallen and A. Klepsch, *Biomedical and Health Research Series: Human Genome Analysis Programme* Vol. 8 IOS Press, Amsterdam, 1995; S. Suhai (ed.), *Theoretical and Computational Methods in Genome Research*, Plenum Press, London, 1997.
15. M. Fortun, 'The Human Genome Project and the acceleration of biotechnology', in A. Thackray (ed.), *op cit.*
16. 'UK academics, researchers, launch firm to meet bioinformatics outsourcing demands', *Bioinform News Service*, 2, 23, 1998; 'Bioinformatics experts' start-up will integrate research systems', *Bioinform News Service*, 2, 17, 1998.
17. S. Hilgartner, 'Data access policy in genome research' in A. Thackray (ed.), *op cit.*
18. *Ibid.*
19. M. Wadman, 'NIH strives to keep resource sharing', *Nature*, 27 May 1999, p. 291.
20. *Science News*, 153, 23 May 1999, p. 334; 'Genomics boom', *Science*, 285, 16 July 1999, p. 309.
21. *Science*, 18 June 1999, pp. 1906–9.
22. *Science*, 28 May 1999, p. 1445.
23. Nature, 403, 1999, p. 23l.
24. See Mannvernd (*Icelanders for Ethics in Science and Medicine*). www.mannvernd.is.
25. See Council of Europe Steering Committee on Bioethics, 'The Icelandic act on a health sector database and the Council of Europe conventions', Ministry of Health and Social Security, *Strasbourg*, 1999 (CDBI-CO-GT2); Ministry of Health, *Bill on a Health Sector Database*, Ministry of Health, Reykjavik, 1998.
26. R. Chadwick, 'The Icelandic database—do modern times need modern sagas?', *British Medical Journal*, *319*, 1999, pp. 441–4.
27. R. v Department of Health *ex parte* Source Informatics Ltd; *Times Law Reports*, 14 June 1999.
28. *British Medical Journal*, 320, 2000, p. 77.
29. R. Fears and G. Poste, 'Building population genetics resources using the UK NHS', *Science*, 284, 1999, pp. 267–8; 'Swimming or drowning' (Editorial), *New Scientist*, *5* December 1998, p. 3; *Times Higher Educational Supplement*, 11 February 2000.
30. See George Poste in submission to the House of Lords Select Committee on Science and Technology: Second Report, 199–2000.
31. *Chemical Engineering News*, 24 August 1998, p. 11.
32. D. Malakoff, 'NIH urged to fund centers to merge computing and biology', *Science*, 284, 11 June 1999.
33. A.H. van Reekum, *Intellectual Property and Pharmaceutical Innovation*, PhD Thesis, Labyrint, Gronongen 1998.
34. S. Hilgartner, 'Data access policy in genome research', in A. Thackray (ed.), *op cit.*
35. W. Faulkner and J. Senker, *Knowledge Frontiers*, Oxford University Press, Oxford, 1995.
36. R. Nelson, 'What is private and what is public about technology', *Science, Technology and Human Values*, 14, 1989, pp. 199–24l.
37. D. Hicks, 'Published papers, tacit competencies, and corporate management of public/private character of knowledge', *Industrial and Corporate Change*, 4, 2, 1995, pp. 401–24.
38. *Nature*, 391, 1998, p. 621; also, delegates at the annual meeting of the American Society for Human Genetics (27–31 October 1998) discussed the problem that newly generated DNA sequence data have outpaced bioinformatic capability to make use of the data. See *Bioinform News Service*, 1998.
39. *Guardian*, 15 December 1999, p. 1l.
40. D. Butler, 'French geneticists raise worries over the use of new genome funds', *Nature*, 20 May 1999, pp. l85–6.
41. *Nature*, 403, 2000, p. 119.