

International Journal of the Commons
Vol. 4, no 1 February 2010, pp. 481–506
Publisher: Igitur, Utrecht Publishing & Archiving Services for IASC
URL:<http://www.thecommonsjournal.org>
URN:NBN:NL:UI:10-1-100221
Copyright: content is licensed under a Creative Commons Attribution 3.0 License
ISSN: 1875-0281

Public or private economies of knowledge: the economics of diffusion and appropriation of bioinformatics tools

Mark Harvey

Centre for Research in Economic Sociology and Innovation, University of Essex, UK,
mharvey@essex.ac.uk

Andrew McMeekin

Manchester Business School, University of Manchester, UK,
mharvey@essex.ac.uk

Abstract: The past three decades have witnessed a period of great turbulence in the economies of biological knowledge, during which there has been great uncertainty as to how and where boundaries could be drawn between public or private knowledge especially with regard to the explosive growth in biological databases and their related bioinformatic tools. This paper will focus on some of the key software tools developed in relation to bio-databases. It will argue that bioinformatic tools are particularly economically unstable, and that there is a continuing tension and competition between their public and private modes of production, appropriation, distribution, and use. The paper adopts an ‘instituted economic process’ approach, and in this paper will elaborate on processes of making knowledge public in the creation of ‘public goods’. The question is one of continuously creating and sustaining new institutions of the commons. We believe this is critical to an understanding of the division and interdependency between public and private economies of knowledge.

Keywords: Commons, instituted economic process, markets for knowledge, open-source software, public domain, public goods

1. Capitalism’s great divide

From the early 18th century, the earlier differentiation between public and private knowledge began to change and deepen into what we can now recognise as a great divide at the centre of the historical growth of knowledge and industrial

capitalism (David 2008). New institutional forms of public knowledge were being developed (encyclopaedias, public experimental demonstrations, development of shared mathematical languages across Europe) (Schaffer 1982; Shapin and Schaffer 1985; Golinski 1989, 1992; Stewart 1992). It could be argued that these were only tentative beginnings of the creation of a new ‘commons’ (David 2001a). Seminal accounts that have highlighted the significance of scientific (and technological) knowledge for the major historical transformation of industrial capitalism, however, have seriously underplayed the emergence of the great divide and the new forms of both public and private economies of knowledge (Mokyr 2002; Landes 2003).¹

During the 19th, and even more the 20th centuries, state investment in the production and reproduction of knowledge, through public education and public institutions of research, created an ever-expanding public domain of knowledge. These formed diverse public infrastructures of knowledge in new varieties of industrial capitalisms, even if the co-evolution of scientific and technological knowledge is recognised to be complex, involving feedback loops in both directions (Rosenberg 1992a, b; 1994). Public institutions of knowledge were at least as critical to the development of commercial, market capitalist activities as the infrastructures of health, law, or communications. It is difficult to conceive of continued industrial transformation without the growth of public economies of scientific and technological knowledge (Allen 2009). In retrospect, the new institutional forms of the public knowledge commons can be seen to initiate political economies that were multi-modal at their core, combining dynamic growth in many economic modes, not only the market. The characterisation of industrial capitalism as the historical beginnings of the commodification of all human activities, in which markets figure as *the* central institution (Nee and Swedberg 2005), is, in this perspective, one-sided. The dynamic growth of public domains, national and international, and especially of a ‘commons’ of scientific knowledge, has to be seen as a second, and far from subsidiary, motor of capitalist economic growth.

Yet, the emergence of new divisions between public and private domains of knowledge was, and continues to be, unstable and contested – evidence of a ‘fault line’ that runs through the history of industrial capitalisms. The boundaries are both shifting and “fuzzy”. At the extremes, the contrast between a public commons of knowledge (e.g. in a scientific journal) and private intellectual property rights (e.g. in a patent) can be quite sharp. But, as we shall show, there are many intermediate forms of ‘property rights’ over knowledge that blur any sharp lines of division. It could be argued that the legal establishment of private intellectual

¹ Mokyr, for example, projected a view of early science, where handfuls of individuals, the scientific notables of the day responsible for creating a society’s ‘epistemic base’, onto the subsequent history of the industrial revolution. The growth and expansion of the public epistemic base, eventually universal education, and major secondary and higher educational and research institutions, hardly figure in this account.

property was fraught precisely because such law also, by default, instituted what was *not* privately appropriable (Machlup and Penrose 1950; Kahn and Sokoloff 2001). The developments of public and private domains of knowledge have been closely intertwined. One of the key moments of institutional change in US patent law in 1836 involved the establishment of a panel of experts as the first publicly recognised ‘jury’ that judged whether or not a claim to private property rights was worthy of a patent (Sokoloff 1988): a co-institution of public experts and new forms of private *and* public property rights.² The emergent Patent Offices in many industrialising economies, and their subsequent developments from the mid-19th century, were janus-faced, even if the public face was on the shadow-side for the commercially-oriented gaze. Yet, it is ironic that somehow public institutions acquired much less public recognition, either in law or in economic theory.

2. The great divide and contemporary biological knowledge

This paper focuses on a new historical moment of instability at the fault-line of the great divide. In the late 20th century, biological knowledge underwent a revolutionary change in its epistemic practices and disciplinary characteristics (Cook-Deegan 1994; Zweiger 2000; Moody 2004; Harvey and McMeekin 2007). Although much of the emphasis concerning this revolution has rightly been placed on the genomics, post-genomics and microbiology revolution, new technologies of experimentation, digitisation of images, and computer-based *in silico* experimentation have extended this transformation across many biological domains. We argue that turbulence in knowledge is often accompanied by turbulence in the economic organisation of knowledge. The emergence of new knowledge forms, outside traditional and routinised practices of biological science and technology, presented challenges to existing economic institutions. Compared with the epistemic practices of the traditional ‘wet’ laboratory and the publication of a paper in the classic range of biological science journals, the landscape of biology and biotechnology has been transformed. In many domains, biology is now ‘big science’ and ‘big technology’, involving a step-change in funding, and new scales of international collaborative activity. A raft of new journals has emerged, redrawing boundaries and constructing new alliances with other disciplines: computing, engineering, physics, mathematics, to name but the most obvious.

The focus of this paper will be new forms of biological knowledge, and by this we mean especially the huge variety of bio-data deposited in bio-databases, and new forms of ‘soft’ scientific instrumentation, notably bioinformatic computing tools. By their nature, these new forms of knowledge and instrumentation broke with the established formats of hypothesis-testing experimentation. As we discuss in greater depth below, the separation of data production and deposition

² Patent legislation itself may contain various requirements of disclosure, and in so doing differentiates between dissemination and appropriation of knowledge, a point which we develop later.

from journal publication constituted a change in how biological knowledge was organised. This organisational transformation of biological knowledge, we argue, was dynamically related to changes in its economic organisation, and unsettled the boundaries between public and private economies of knowledge.

Assumptions that an easy distinction between 'basic' science and use-oriented, applied science, can be mapped unproblematically onto an equally easy distinction between public and private knowledge, has long been challenged (Nelson 1959, 1989, 2004, 2006; Stokes 1997). There is no clear or compelling economic or epistemic logic dividing basic science as public from technological knowledge as private and commercial. However, the turbulence in economies of knowledge that accompanied the revolution in biological knowledge was acute to the point where there was radical uncertainty as to whether vast swathes of biological knowledge would become privately appropriated for commercial use, or be established in the public domain. There were fears expressed that our biological heritage was at risk of being balkanised by competing commercial interests (Sulston and Ferry 2003) and counter-arguments that the advancement of biological knowledge could only be secured through market incentives and commercial investments (Venter 1998; Venter and Adams 1998).

Behaviours by major organisations, quite unpredictable from traditional assumptions, further de-stabilised existing economic organisation. The National Institute of Health (NIH), one of the largest funders of public science in the US, led the way in patenting genomic material in the early 1990s. A major pharmaceutical company, Merck, countered the NIH by rapidly depositing vast quantities of similar data into the public domain so undermining potential patenting, at considerable expense to its own balance sheet (Marshall 1999; Eisenberg 2000; Eisenberg and Nelson 2002). Many academics were fearful that, following the 1985 Bayh-Dole Act in the US encouraging patenting by universities, the public domain would be irreversibly eroded (Mowery and Sampat 2001; Mowery et al. 2001; Coriat and Orsi 2002; Mowery and Zeidonis 2002; Rai and Eisenberg 2003).

We focused our research on and around the fault-line, in order to better understand the dynamics of division and interdependence between public and private domains of knowledge. Many areas of knowledge, where historically the division lay at some distance in the past, remained undisturbed. We are not arguing that all biological knowledge was equally troubled, either epistemically or economically. At the fault-line, *ex ante* it was certainly unpredicted and unpredictable by any of the major players, where the divide would fall. Even now, from our analysis, it is clear that the boundary is contingent, and could well have been drawn otherwise. In the case of bioinformatic tools, the main focus of this paper, there is continuing uncertainty and instability, both for public and market forms. But what we want to stress above all is that we are examining ongoing processes of differentiation and interdependence. The question is not one of which or how many elements of new knowledge fit into old institutional divisions, private or public. The process of differentiation and interdependence creates new divides, and new institutions of both public and private knowledge. The novel

institutionalisation of public and private knowledge is at least as important as the other major finding of our research, namely that public domains of biological knowledge were resilient, innovative and expanding. For, far from retreating under the onslaught of commercialisation and the rise of new biotechnologies, the remarkable feature of recent history has been the emergence of new institutions, the public databases and open-source software. But in order to understand why this has been so, we were challenged to re-examine some of the very foundations of economic organisation, the nature of property in capitalism, indeed in the case of the public domain, when and whether it is appropriate to speak of public goods as public property.

3. Rethinking the commons

Many of the debates around the contested nature of the new biological knowledge have used the ancient term of the ‘commons’. Allusion has already been made to the way that the public domain, or the public good characteristics, of knowledge – or indeed other areas of the ‘public’ economies of capitalism – have been under-characterized in social and economic theory, and underdeveloped in law. There has been a burgeoning sub-discipline and publications on intellectual property rights (IPR), almost exclusively directed at, and indeed often only indicating, private property rights. In this context, the public is often portrayed as the negative mirror image of the classic definitions of private property (Demsetz 1967): the public is the absence of exclusivity, not divisible, or non-rivalrousness in use, and not under anyone’s control. Once produced, and available for copying, this is sometimes almost presented as an intrinsic property of knowledge, because of the alleged relatively low costs of copying, and lack of means of preventing copying once deposited or distributed in the public domain (Nelson 1959; Arrow 1962; Dasgupta and David 1994).³

³ There are several assumptions raised by this presentation of public knowledge as inherently difficult to privatise, following the famous metaphor of Jefferson’s flame (see Boyle 2008), which is replicated ‘costlessly’. Firstly, for the flame (knowledge) to be transferred from one taper (a knowledge bearer) to another, a shared (epistemic) infrastructure with common properties that enable transmission is assumed, supporting transmissibility. This is far from guaranteed, and is certainly not cost free. Secondly, the metaphor deals only with distribution, not with appropriation – being publicly distributed is not at all the same thing as being publicly appropriated as we shall demonstrate more fully. Public distribution is a necessary but not sufficient condition for public appropriation. Thirdly, the continued burning of both flames (‘public’ sharing and maintaining knowledge *over time*) assumes an adequate common oxygen resource in order for copying not to be rivalrous (expenditure on education is not an infinite resource, and there are competing claims). Without unlimited resources of knowledge-sustaining ‘oxygen’, copying can deplete the original. Fourthly, an assumption is made that all copies maintain their similitude and identity in the process of copying without diminishing the worth of the original. However, the integration of a copy into the transferee’s knowledge frames may typically lead to differentiations and novelties competitive with the original, so potentially reducing the original’s value. Such differentiation also opens the door to possible private appropriation. In short, these discussions seriously under-characterise the conditions that characterise and underpin the ‘public’ character of a public good.

In a similar vein, the concept of the ‘commons’ with its connotations of unrestricted access to uncontrolled use by any and everyone is in effect only the negative image of private property, as defined by exclusivity and rights of control over use. Indeed, it is strange that a feudal institution has figured so large in the notion of contemporary concepts of public goods:⁴ in its origins, the commons were in fact owned by feudal lords, who devolved specific and limited rights of use to those circumscribed populations of the manor subordinated to them (Humphries 1990). The story of the ‘tragedy of the commons’ is then told as a moral tale linked to assumed universality of selfish individual interest and the *absence* of private control: the pursuit of individual interest without collective regulation inevitably leads to overuse and exhaustion of the common resource. The story has then been further buttressed by the ‘tragedy of the anti-commons’, another morality tale which has been deployed to convey two alternative (not necessarily incompatible) morals: both multiple partial claims to use and unclear rights of exclusion lead to underuse, resulting in the worst of all worlds (Hardin 1968, 1998; Heller 1998; Heller and Eisenberg 1998). The moral choice is either to restore the commons,⁵ or institute clear and unequivocal private property rights. David (2001b; also Lessig 2004) has effectively demonstrated the inadequacy of the concept of a commons when applied to knowledge, because knowledge is not a finite resource exhaustible by overuse. Knowledge grows and develops with use, especially when combined with collective examination and testing.⁶ Yet we are still left with a concept of commons defined by unrestricted access and non-exclusive use which is at bottom a negative rather than a positive account of the public character of scientific knowledge.

One of the aims of this paper and our book (Harvey and McMeekin 2007), therefore, is to develop a more adequate and multi-dimensional understanding of the public character of scientific and technical knowledge. The focus of the preceding discussion has intimated the need to distinguish between distribution and appropriation, and the necessity to include both in any adequate characterisation. Even within a narrow perspective of ‘property’ and ownership, we believe that a positive account of public appropriation is required to differentiate between different kinds of public good (Kaul and Mendoza 2003). Consequently, we need to elaborate a concept adequate to the evolving and developing nature of public scientific and technical knowledge. Essentially, we shall argue that public appropriation means public control. New forms of public control regulate the

⁴ The classic trope of the commons is that it is an open, public and collective resource, threatened by the menace of private ownership and ‘enclosure’, whether in relation to land in historical times or knowledge and ‘the commons of the mind’ in contemporary societies (Boyce 2003, 2008).

⁵ It should be remembered Heller’s first development of the anti-commons concept was in the context of the transformation of socialist into market economies, with the thrust of the argument clearly in favour of clear individual private property rights.

⁶ Interestingly, David here picks up on a notion of public knowledge in terms of collective production, as against unrestricted access and unlimited use. This is a point our framework develops further.

qualities, temporalities, representations and ontologies of knowledge, through a continuous and ongoing process of development of standards, testing, and enforcement procedures. In this paper, we aim to show that the institution of biodatabases, and specifically, the uses of bioinformatic tools, are at the centre of an evolving process of public control within key areas of biological science.

But we shall argue that there is much more to making knowledge public than public appropriation. The next section of the paper will summarise our broad analytical framework, necessary to encompass the multi-dimensionality of the public–private divide, and its historical and evolutionary character. We want to argue an apparently simple idea. A sharp division of labour between producers and users of knowledge opens up the possibility (no more) for stable patterns of distribution and exchange between them, and hence markets for knowledge *and/or* public domains of knowledge. Within capitalist political economies, markets for knowledge may indeed be established. But the growth of knowledge within collective communities of experts rests on an *absence* of divisions between producers and users of that knowledge for the production of further knowledge. So, there is an ongoing tension between establishing sharp divisions between producers and users, and ensuring the growth of knowledge through the productive and unrestricted flow of knowledge between producers and users within specific communities of experts. Hence, where markets for knowledge emerge, they remain dependent upon the growth of knowledge in the public domain. The instability of both private and public economies of knowledge for bioinformatic tools critically supports this argument by examining competing alternative experimentations in economic organisation, centrally focusing on conflicts around the emergent divisions between producers and users of bioinformatic tools.

Having set out our analytical framework, the empirical core of the paper then employs it in order to analyse major historical bioinformatic cases. We first explore the process of differentiation and interdependence between new forms of private and public knowledge in the early history of bioinformatic tools and biodata. Then we present case studies of two of the most significant bioinformatic tools of the 1990s, exemplifying rapidly shifting, unsettled and blurred boundaries. We then conclude by concentrating our focus on the ‘instituting of the commons’, by arguing that public economies of knowledge entail the dynamic relations between collective production, forms of distribution, modes of control, and varieties of use. Appropriation – or property rights – can only be fully understood within this broader analytical framework of economies of knowledge.

4. The ‘instituted economic process’ (IEP) approach

One of the key features of the revolution in biological science has undoubtedly been the emergence of markets for knowledge. It has long been recognised that a society’s stock of knowledge, its ‘epistemic base’ (Mokyr 2002), is necessary for the production of all kinds of goods, many traded in markets. Markets for

knowledge – as distinct from labour markets for knowledge-bearers – entail the creation of knowledge entities that are themselves tradable. For biology, this was most evident as soon as data contained in biodatabases became a separate and novel form of biological knowledge, partly as a consequence of new technologies of data production, such as high-throughput sequencing.⁷ The emergence of new and distinct ‘economies of knowledge’, both public and private, present challenges to economic analysis, stretching and expanding our analytical frameworks. Knowledge is notoriously elusive as an object of socio-economic enquiry, and is ill-served when corralled into frameworks designed for understanding the more conventional ‘economies of products and services’.

The ‘instituted economic process’ (IEP) approach was developed out of an anthropological perspective on economies (Polanyi 1957a, b),⁸ that brought with it the advantage of neutrality with respect to market or non-market, primitive, historical or contemporary, economies (Harvey et al. 2003; Harvey 2007; Harvey and McMeekin 2007). Presented here schematically, the approach posits economies as the combination of four transformational processes: transformations of qualitative characteristics, transformations of spatial location, transformations of functionality, and transformations of control. Taken separately, none of these four transformational processes are in themselves economic. Only when instituted in combination with each other do they become constitutive of economies, recognisable under the guise of respectively, production, distribution, use and appropriation. So when applied to economies of knowledge, the IEP approach involves the institution of processes of the qualitative transformation of knowledge (knowledge growth and accumulation); the distribution or dissemination of knowledge; the changes in use of knowledge; and the control of knowledge, formal and informal, legal and technical, by individuals, legal entities, collective communities, or wider societies. Economies are instituted insofar as they successfully reproduce themselves through historical combinations of these four processes. Again, in terms of knowledge, this involves the idea that social knowledge is produced, distributed, used and controlled within a society over a period of time. Historically, we know that there are no guarantees that an existing level of knowledge is maintained. Knowledge is also unequally distributed, used and controlled both within and between societies.

⁷ Prior to this separation, data had more typically been presented along with experimental results from hypothesis testing, or as evidence supporting taxonomic categorisations, for example.

⁸ Following *The Great Transformation* (Polanyi 1944), Polanyi developed a novel framework for understanding the variety of historical economies, especially with his concept of economies as instituted process. For him, there were only two processes necessary to specify the formation of an economy, displacements in space and transfers of control. His failure to consider transformations of quality (“production”) or use (intermediate or end consumption), led him to give undue centrality to markets in contemporary economies, as the sites of both distribution and changes of ownership. Innovation in production and changes in consumption thus largely fall outside his analysis (Harvey and Metcalfe 2004).

But over the past three centuries, we have witnessed the growth of the modern sciences, absorbing vastly expanding societal resources, and an ever-changing division and hierarchical differentiation of scientific activities: new disciplines, new organisations of research production and teaching reproduction, new modes of funding, and so on. The creation of specialist public institutions, alongside the formation of private, knowledge-intensive corporations, has been ever more evident in the historical development of capitalism. Some of these institutions, notably those of the sciences, have been internationalised or globalised. The databanks that now hold huge quantities of genomic data in the US, Europe and Japanese are contemporary examples of new global institutions of science, with global interchange and harmonisation of data.

When analysing the emergence of new public institutions and markets in biological knowledge from an IEP perspective, two particular theoretical developments proved useful. The significance of the differentiation between production and use, and associated divisions of labour, has already been mentioned. For example, the specialised producers of bioinformatic tools, and in particular the creators of software packages, enable users to avail themselves of technologies without necessarily understanding the workings of the algorithms within a computer programme. The first development is to distinguish between two axes of the four economic processes, production and use, on the one hand, distribution and appropriation on the other (Figure 1, below).

The articulation of economies by two axes reinforces the earlier theoretical point that the four processes only become economic when in combination. When a relatively stable differentiation between production and use arises, this leads to a socio-economic division between groups of producers and groups of users. But for this to occur and be sustained over time, complementary processes of knowledge distribution between the groups emerge, and along with that, processes

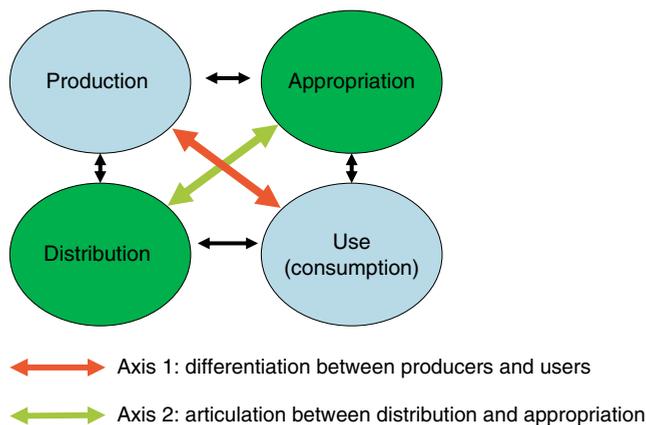


Figure 1: The two axes of instituted economic processes.

appropriation develop, by and between groups and individuals. In short there is an axis of differentiation and a second of integration,⁹ which, if leading to continuous and sustainable reproduction, constitute ‘an economy’.

The division between production and use led to the second development, particularly significant in the production of knowledge: knowledge as an output can either be used for the further production of further similar knowledge or for dissimilar knowledge production, or finally for non-knowledge purposes. Thus, a bioinformatics tool can either be the basis of further developments of similar bioinformatic tools (modified, for example, to be included in an interoperable suite of tools), or by protein structure specialists and deployed in producing analysis of protein structure, or by pharmaceutical company scientists identifying possible drug targets. Many of the productive outputs are therefore polyvalent in use, and given this multiplicity of alternative uses, competing economies of knowledge emerge, depending on how these different uses are combined, separated, or restricted.

Within the IEP framework for analysing economies of knowledge, therefore, it is important to stress that there are no prior assumptions about how and where new divisions may develop between knowledge producers and users, how and in what manner the resultant knowledge is distributed amongst potential users, or appropriated, collectively, corporately, or individually. Whether, and over what time spans, new economies of knowledge may be instituted is an empirical matter, and so it is to the history of the development of bioinformatic tools that we now turn.

5. Early process of differentiation and integration within bioinformatics

In *Public or Private Economies of Knowledge* (Harvey and McMeekin 2007), we analysed the development of biodatabases, bioinformatic tools, and a microbial genome race (*Agrobacterium tumefaciens*). By restricting the focus of this paper to bioinformatic tools, we relegate to ‘pre-history’ a number of important developments, summarised here. Up to the 1990s, it would have been difficult to treat bio-databases separately from bioinformatic tools, and within the latter, the differentiation between algorithms and their operationalisation in computer programmes was quite underdeveloped. Many of the early bioinformatic software tools were distributed along with the databases, on tapes. Once internet access was developed for the main databases, the tools could be distributed separately from biodata, and in most cases they were, for both genomic and protein databases (Bairoch 2000). Secondly, in the early days of high-throughput sequencing, the manufacturers of the hardware, notably for the ABI Perkin-Elmer first generation

⁹ For Polanyi (1957a), how economic processes of distribution and appropriation of resources are organised within a society is critical to their integration. He distinguishes between three modes of integration, according to whether they are based on reciprocal obligation, hierarchical redistribution, or market exchange.

sequence assemblers, attempted to embed the software with the hardware. The public science Human Genome Project resisted this ‘privatisation’ of what was considered to be critical scientific aspect of the nature and quality of data produced by the software. Hence, Staden and Sulston reverse engineered the software, creating their own sequence assembly bioinformatic tools, independent of commercial hardware, and supported within the public domain. This independence of bioinformatic tools from hard technology, therefore, was secured as a necessary guarantee for the scientific robustness and quality-testing of the data (Sulston and Ferry 2003, 93–4). However, if only to emphasise the contingency of securing a public economy of knowledge along with the independence of tools from machinery, an almost identical controversy has arisen with the second generation high-throughput genome sequencers (the 454, for example), and once more independent public software has been developed and for essential the same scientific rationale (O’Rourke 2006).

Thirdly, before the widespread use of computer programming and desktop computers, the algorithms that were to be the key workhorses in the new technologies of similarity searching were published in scientific journals, along completely accepted and normal routines and practices of the day: there were no thoughts or prospects of commercialisation. The path-breaking algorithms of Needleman and Wunsch (1970) and Smith and Waterman (1981) are cases in point. As these were developed to enhance and accelerate these techniques, the algorithms were incorporated into software programmes, notably FASTP (Lipman and Pearson 1985) and later BLAST (Altschul et al. 1990), publicly available on sites hosted by the main bio-databases. The biological significance of similarity searching for the understanding of biological function across species was recognised to be one of the major scientific gains of the period (Doolittle et al. 1983; Waterfield et al. 1983; Hodgman 2000). For our argument, perhaps the most significant implication of this development was that any restriction of access to biological data was a restriction on the scientific development of understanding biological function through similarity searching. Open access to, and integration across, biological databases carried a scientific imperative underpinning the need for biodatabases to be public domain, and global in scale. At this early stage, therefore, the public character of both major databases and bioinformatic tools was taken for granted: it was ‘carry on as usual’.

Ironically, however, as soon as the process of differentiation in epistemic practices became established, and ‘bioinformatics’ and bioinformatic tool production established themselves as distinct activities, the usual no longer appeared usual. Questions were raised as to whether these activities were truly ‘scientific research’ worthy of public funding. Many of the early forms of databases and tools were instituted within quite novel ‘economies of knowledge’, maintaining themselves through licenses and charging both academics and commercial enterprises, if at different rates. There was open access, including to the source-codes of the software. But, because of the unwillingness of public funding bodies to support these activities, access was at a price. An exchange

was instituted in return for a right of access, while use was unrestricted for either public or private users. This destabilisation of the normalities of public economies of knowledge sets the scene for the two main empirical cases analysed below.

The resistance to publicly funded bioinformatic tools (and databases) is related to a fundamental question about the nature of these tools, as novel epistemic entities. In what ways are bioinformatic tools searching databases similar to or different from microscopes or telescopes searching the universe or a various microcosms? This relates to the question of whether there are distinct producers and users (the tool-makers, and the tool-users), and the related question of whether scientific tools are commercially produced (as have all PCR or high-throughput sequencers, for example) or publicly provided. As bioinformatic tools developed, and were operationalised in computer programmes, there were various groups of users increasingly no longer needing to understand how the algorithms worked, any more than a biologist using a microscope needs to understand advanced optics. Up to a certain point, therefore, bioinformatic tools seem similar to microscopes in the way the epistemic functions of tool-maker and tool-user are differentiated. However, as we have already seen with the Staden-Sanger sequence assembly software, there is also an important difference. Bioinformatic tools are deeply implicated in the construction of bio-data, the standardisation, quality control and rigour of the data, as well as underpinning the new ontologies of biological understanding. The development of the tools is continuous and interdependent with the development of new analytical and empirical scientific knowledge. So, while they are useful tools for some users, they are much more than that, and serve different epistemic functions at the core of a scientific research development for others. For the former, there is no need to understand how the tools work in order to use them; for the latter, tool development is central to the research activity, requiring access to, and understanding of, the inner workings of the tool. This polyvalence of use – maybe only a transitory phase but one that continues to the present – is, we believe, at the heart of the instability that persists in the economies of knowledge for bioinformatic tools.

Before leaving the discussion of the ‘pre-history’ of bioinformatic tools, however, this must be placed in the wider context. Although differentiated as epistemic entities, many bioinformatic tools are primarily dedicated to the analysis of bio-databases. We cannot present the evidence or analysis here, but these biodatabases emerged as dynamic, growing, and novel institutions overwhelmingly in the public domain. A whole new range of databases at the National Centre for Biotechnology Information (NCBI), European Bioinformatics Institute and DNA Databank of Japan (DDBJ) constitute a dominant, hegemonic presence, whose existence now appears securely established. As a consequence, whatever the economic path of development of bioinformatic tools, they achieve their value, and prove their use, only in strict interdependence with the ongoing growth and expansion of public domain data. The big picture for many commercial developments of the biological sciences in pharmaceutical or agricultural corporations, is thus one of dependency

on a vibrant and expanding public domain. There could be no clearer evidence of multi-modal capitalist growth.

6. Contrasts and conflicts: two trajectories of bioinformatic tool development

Two of the major bioinformatic software tools of recent decades, the Sequence Retrieval System (SRS) incorporating GeneQuiz and the Wisconsin package or GCG, underwent quite dramatic, yet contrasting, historical developments, epistemically *and* economically. Both had similar origins. They emerged as spin-outs from public science institutions, the European Molecular Biology Laboratory (EMBL) and the University of Wisconsin, respectively. But, situated on different continents and institutional settings, including financing arrangements, their subsequent pathways could scarcely be more contrasting. As we write, in one case, the software continues residually in the public domain after a long experimentation with hybrid private and public forms of the software. The story is a rollercoaster of commercial success and failure. In the other, privatisation and commercialisation provoked the emergence of a public alternative, so that we now witness almost equivalent private and public economies of knowledge, in potential or actual competition with each other. The contrast itself illustrates the possibility of different outcomes: there is nothing inherent or pre-established either about epistemic or economic organisation that led to these outcomes. But, it also supports our analysis that similar dynamic processes of differentiation and interdependence between producers and different users underpin both trajectories, and the consequent possibility of either market-commercial or public institutional economies of knowledge. This is the key purpose of our analysis of these two cases, namely, to offer an explanation of the contrasting trajectories.

The two contrasted trajectories are summarised – very schematically – below, with a particular emphasis on the articulation between the two IEP axes, production and use, distribution and appropriation.

6.1. GeneQuiz, BioScout and SRS¹⁰

The linked historical trajectories of a sequence analysis bioinformatic tool for automated annotation and the Sequence Retrieval System (SRS), both key technologies, are remarkable for their emergence out of a public economy of knowledge, a relatively brief but spectacular experimentation with commercial economies of knowledge, culminating in a return into the public domain, revived with a new and dynamic future. Both tools originated in EMBL, and, following their initial development in the mid-1990s, were considered to be technical instruments vitally useful for research, but no longer to be funded as

¹⁰ We are grateful to Reinhard Schneider, Georg Casari and Thure Etzold of Lion Bioscience, Peter Rice of Lion Bioscience and EBI, Iain Mattaj at EMBL and Janet Thornton, Director of the EBI for the interviews that informed this account.

objects of scientific research. European Commission research policy, influenced by developments in the US, was also directed to stimulating commercialisation of academic achievements where appropriate. GeneQuiz was one of the new generation of tools designed to meet the challenge of the explosion of sequence data, by providing high quality, high speed, annotation (biological interpretation especially of functionality). Consequently, although GeneQuiz remained in use in the public domain, a spin-out firm was created, Lion Bioscience, which had a license to develop and commercialise the tool, particularly for the expanding commercial market of pharma- and agri-biotechnology (McMeekin et al. 2004). The commercial version, named BioScout, was developed within Lion under license to EMBL. Then in 2002 its source-code was deemed significantly different from GeneQuiz, as developed within the EBI. BioScout then became an independent software package, with protected source code, and an established market. In terms of our IEP analysis, one tool had divided into two: GeneQuiz remained a tool within academia where a community of experts had access to the source code, and developed it in conjunction with a community of users. The tool was an integral part of knowledge production, within a shared community. It was freely distributed to, and was under the control of, that community. BioScout, by contrast, established a clear division between the tool developers (internal and restricted to the company), and users, the market clients. Both market clients and the academic community were cut off from the source code, and Lion distributed the tool through market exchanges, and had full control of their intellectual property. The functionalities of the two versions had become differentiated, and, up to a point, enabled the institution of two distinct economies of knowledge around the tool, one private the other public.

However, there was a continuing instability at the fault line between public and private domains. GeneQuiz and BioScout had both been developed within the SRS technology platform, a tool that provided an homogenous interface to 80 biological databanks at the time (Etzold et al. 1996). SRS was the dominant, indeed standard, bioinformatic technology for integrating a vast new and disparate data landscape. In a sense, SRS underpinned the integration of multiple databases into a single public space. For Lion Bioscience, this meant that their core business was constrained by its dependency on the SRS environment. At the time, moreover, there were considerable conflicts over funding SRS in the European Commission. So, as with GeneQuiz, a license was given to Lion BioScience to develop SRS and commercialise it. This time, however, no doubt because of its significance for integrating public domain data, the license set conditions that precluded separate development. Moreover, new software tools and applications, whether developed in the public or commercial sector, also required access to the SRS source code if they were to operate within the SRS data environment. There was to be a single SRS, but with polyvalent uses, serving both public and commercial users. This hybrid economy of knowledge was reflected in the strange arrangement that Lion's CEO, Thure Etzold, retained a position at the EBI, straddling private and public organisations. SRS

was provided to commercial clients by Lion, while at the same time being freely available within the academic community.

For a while, this hybrid economy of knowledge proved viable, and for Lion, it was seen to provide considerable commercial leverage, because SRS secured a market position by being the gold standard for this tool functionality, acquiring an almost monopoly position. By the same token, BioScout also traded on this advantage. However, this hybrid economy also proved unstable, and eventually unsustainable. A commercial logic was pushing for SRS to become integrated into a commercial package of software, a 'one-stop-shop' for major clients. As a consequence, through a succession of acquisitions, Lion attempted to construct a bioinformatic platform including chemo- and medico-informatics, as well as integrative middleware, providing access to relational databases. This process of integration – never technically achieved – entailed major licensing agreements with Bayer and Nestlé, and for a time Lion appeared destined to become 'king of the (market) jungle', attracting investments of more than \$100 million. As the market grew, and major clients were viewing the Lion platform as possibly central to their whole knowledge management, so too did the viability of a significant component of the platform remaining open-source and in the public domain appear less commercially interesting. At the same time, the academic community were becoming increasingly concerned that the increasingly market, drug development, adaptations of SRS would eventually lead to a fate similar to that of BioScout, or even that SRS would be entirely privatised. The integrated technology platform, incorporating SRS, was therefore becoming unsustainable from all sides. Within a short space of time, the SRS market imploded, and in 2006, Lion Bioscience's bioinformatic business was sold for a mere \$5 million.

But SRS survived and thrived, as did GeneQuiz. Brought back fully into the public domain within the EB1, it now has 31 registered public science institution servers, integrating public domain data across 1104 libraries. It has been progressively developed, undergoing many new and quite substantial revisions. In terms of our IEP analysis, the terms set for the commercialisation of SRS, in particular the retention of open access to its source codes, both prevented clear functional differentiation *and separation* of uses, and, by the same token, compromised its tradability in knowledge markets. Selling a knowledge management platform to corporate clients that had a key component remaining open-source, meant that neither the package nor the company were open to full private ownership or control. Straddling the fault-line between public and private proved unsustainable. Equally, however, it should be emphasised that market creation was undermined by the growth and dynamism of the public domain, and the advantages of a technology that integrated data across the world in a single epistemic space. SRS provides a public asset, an expanded commons, open to use by both scientific and commercial communities. But, although this might now seem the obvious outcome of the trajectory, it cannot be overemphasised enough that many of those involved at the time considered this to be the least likely end-game. Moreover, our second example shows that there are indeed alternatives.

6.2. Dividing and competing: GCG and EMBOSS¹¹

The bioinformatic tool known as GCG from the Genetics Computer Group, or the Wisconsin package, from its host organisation the University of Wisconsin, was the pioneer suite of programmes for analysing nucleic acid sequence data. It was put in the public domain at the outset in one of the new journals in the field, *Nucleic Acids Research* (Devereux et al. 1984). It rapidly became a dominant bioinformatics tool, cited over 6000 times. Incorporating the Needleman and Wunsch, and Smith and Waterman algorithms, it created a unified computational environment, and was open source, allowing users to develop and customise it to their own purposes. In IEP terms, it was initially a producer-user/user-producer knowledge output. In the pre-internet period, as already mentioned, it was typical in that it was distributed on tapes along with data from Genbank. As one of the first examples of open-source software, however, it was also typical of the time in charging for distribution, ostensibly to cover costs, and at differential rates for commercial and academic users. Although copyrighted, if the user-producer modified it significantly (by 25%), the copyright lapsed, so providing an incentive to the scientific community to engage in its development.¹²

Quite early on, this public development or production process, through the open-source facility, was exemplified by the emergence in Europe of EGCG,¹³ within EMBnet linked to EMBL, a European adaptation of the package. Importantly, EGCG was made available without charge on the EMBL network file server and ftp site. The main difference between the US and European versions was that it was distributed without technical support, and the GCG group distributed this version to all its subscribers. From an IEP standpoint, the interesting aspect of this complicated form of exchange is that it entailed a cost of distribution with no rights of appropriation – clearly demonstrating the importance of distinguishing between these two processes of making knowledge (or any other entity) public or private.

Partly because GCG was at least self-financing, and partly, no doubt because of the very different culture of the US public science, there were no serious internal pressures to commercialise the package, in the full market sense embracing both distribution and appropriation. But this economy of knowledge was de-stabilised from outside by a company trading in a competitor software package that had spun out of Stanford, namely, Intelligenetics (IG). GCG were considered to be engaging in unfair, state-subsidised, competition, distributing their product at a lower price

¹¹ This account draws on interviews with John Devereux, ex-GCG, Scott Khan (Accelrys), Peter Rice and Alan Bleasby. We are grateful to them, and especially to John Devereux for providing us with correspondence and documentation related to the spinning out of GCG from the University of Wisconsin.

¹² It should be noted that these forms of property rights were quite ‘experimental’, untested by law, and preceded any formal conception, let alone legal status, of what later came to be known as “open source” software.

¹³ The E of EGCG at first signified European, later Extended.

than possible for a fully independent company. Law suits were threatened, and IG's parent company, Amoco, a major funder and sponsor of the University of Wisconsin, likewise put pressure for IG to be given rights to distribute GCG at full market price, on pain of withdrawing all financial support to the University.

The GCG response, ultimately, was to spin-out themselves, and create an independent company. Initially, no change was made in the mode of distribution or control. During the early 1990s, however, rivalry with EGCG, now distributed widely without charge to non-GCG license holders (even though unusable without a GCG license), increased to a point where a decisive break was triggered. GCG Inc black-boxed the source code, the key form of appropriation for many software markets. As a consequence, the producer-user/user-producer feedback loop so critical for the early development of the tool was also broken decisively: from now on, any development of GCG would be behind the company firewall and the equivalent version of EGCG was abandoned.

From this decisive turning point, two major developments in economies of knowledge arose. GCG Inc, now with fully protected intellectual property rights over the bioinformatic software, themselves became targets for acquisition. The package was successively acquired first by Oxford Molecular, and then in 2001, by the newly formed Accelrys. One of the ironies of this process of business consolidation was that the firm that had triggered the spin out of GCG in the first place, Intelligenetics, had also been acquired by the same two companies. But the key development that occurred as a consequence of the black-boxing was a sharpening of the division between producers and users. Through the process of amalgamation, Accelrys was developing the one-stop-shop, comprehensive and integrated, bioinformatic platform, with increasing user-friendly interfaces, maximising the range of possible clients, and minimising their required bioinformatic expertise. A bioinformatic tool was developed where indeed the users of the scientific equipment required little or no understanding of its inner workings, any more than a user of Microsoft Word.

The second major development was the public response provoked by the black-boxing. Now cut-off from a major scientific resource, where tool development was integral to further advances in biological science, the original user-producers of EGCG created a public equivalent, on an open-source, free distribution basis, EMBOSS, launched in 2000 (Rice et al. 2000). Significantly, with the experience of recent history behind it, the new software suite of programmes was protected under the General Public License (GPL), developed and administered at first by the Free Software Foundation. Effectively, the GPL and even the subsequent Lesser GPL, secures public rights over the software, impeding or restricting any subsequent appropriation and privatisation by private companies or other individuals and bodies. Although initially struggling to maintain itself through public funding, EMBOSS is now supported by the UK's Biotechnology and Biological Science Research Council (BBSRC).

In this new, and reinforced public economy of knowledge, EMBOSS has now achieved full functional equivalence to the commercial GCG package, as evaluated

by the NCBI website. Moreover, as part of its expanded public funding, the software has developed two formats, both supported by a helpdesk, one for the general user, whether public or private, the other for user-producers, bioinformaticians involved in the expert public contributing to the further development of the tool, customised and adapted to scientific knowledge production.

At this point in time, therefore, we can see that a clear process of differentiation of functionality has occurred. There are now two versions of the original software package, one private, embedded in a broad technology life-science technology platform, the other public, available to general users but critically also for user-producers. Whether there is sufficient and stable differentiation of functionality remains an open question, and on that question rests the future of either version.¹⁴

To conclude the analysis of the GCG trajectory, we can return to our question raised as to whether a bioinformatic tool was similar or different from other types of scientific instrumentation, such as microscopes. The answer we can now give is that bioinformatic tools are relatively malleable: they can be crafted into tools like microscopes, where the user does not require knowledge of the workings of the tool; or they can be integral to further knowledge production processes in ways that not only require knowledge of the inner workings, but where such knowledge is necessary to analysis of the outputs of the instruments. This malleability makes it difficult to stabilise and secure functional differentiation into two different types of tool, and to eliminate their polyvalence of use. This epistemic characteristic is intimately bound up with the instabilities in economic organisation that we have analysed.

6.3. Comparing trajectories

To sum up this section, the two contrasting trajectories of SRS and GCG exemplify the analytical usefulness of the instituted economic process approach. The two axes of production and use, distribution and appropriation are articulated with each other in distinctive ways, with greater or lesser stability and durability. Some economies of knowledge became destabilised from within, almost as a consequence of their growth, as was the case with the early GCG package when resident within the University. Some were triggered by external pressures, notably the research funding climate in Europe and the initial resistance to understand and adapt to the novelty of digital, computer-based, biological science.

Both trajectories, each marked by transient economies of knowledge, reveal the dynamic connection between changing organisation of knowledge and

¹⁴ Interestingly, the unstable frontier between public and private forms of these software package has continued up to the moment of writing, with attempts by some GCG users to get the source code released into the public domain. Conversely, EMBOSS has defied the odds in successfully being developed by its supporters, with a new version just released. Ironically, it can be argued that EMBOSS presents commercial users with less of a risk to continued access and development than its private Accelrys rival. Peter Rice, personal communication, August 2009.

changing economic organisation. Changes in the producer-user relationships, the processes of differentiation within scientific activity, entail changes in processes of integration, through the distribution and appropriation of knowledge. Many of the transient forms can be seen as experimentation in economic organisation – such as the two-tier, non-commercial pricing systems for distribution, for example, or the new forms of open-source, license-protected, public knowledge.

Above all, the creation of a divide between public and private domains appears as a dynamic and ongoing process, where new forms of economic organisation on both sides of the divide have emerged. Nonetheless, new divisions did occur, and the least sustainable economies are those that experimented with straddling that divide. Moreover, the resilience and dynamic of public domain development, even in the face of competition with private commercial forms, is clear even with respect to bioinformatic tools, where usual preconceptions might lead one to anticipate otherwise. The story is one of private growth being interdependent with growth in the public domain, one of the multi-modality of capitalist political economies.

7. Instituting the commons

The view of capitalist economies as one-sidedly market-led economies, locating the source of dynamism and growth in commercial enterprise (both a Marxist and Schumpeterian view), has, we argue, resulted in a undervaluation and lack of analysis of public, non-market, sources of dynamism and growth, most notably with respect to knowledge. We noted that ‘the public good’ or ‘the commons’ are characterised frequently as a default case or negative mirror image of private property: that which is non-exclusive, non-rivalrous in use, or inherently resistant to private appropriation.

Instituting the commons, we have shown, is a much more complex and positive process. In concluding this paper, we wish to draw together some of the central features of the ‘instituted commons’, first by considering them in terms of instituted economies of knowledge, and then, more narrowly, to draw attention to distinctive modes of public appropriation, those that make some knowledge a form ‘public property’.

Our analysis of the formation of public economies of a few but central bioinformatic tools, notably BLAST, FASTP, SRS and EMBOSS, has demonstrated the importance of the articulation between four processes. Co-production by a community of experts was a major feature of all of these packages – and indeed of many others not discussed here. Open-source software led to a process of tool development, where new versions were generated, glitches resolved, and scope and scale of use of the tool expanded. Bioinformatic tools are not only developed through use, but also users become innovators, bringing about change in the tool construct. This then is one sense of public: production by a community of experts. If we were to broaden our focus, and explore the range of users and uses of bioinformatic tools, we could provide evidence of a notable expansion

in the scale of co-production of biological knowledge. To take but one example, a publication of a genome in a scientific paper entails a well-stocked toolbox of many bioinformatic software technologies, an extensive multi-disciplinary collaboration, and frequently, many different organisations, scattered across the world.¹⁵ The production and use of bioinformatic tools both by communities of user-producers, the bioinformatic experts, and by communities of users, within public biological sciences, has undoubtedly been expanding significantly over recent decades.

This expansion of both user-producers and users of tools for further knowledge production, however, is on its own an insufficient characterisation of an instituted commons. The possibility of creating an ongoing development requires also modes of distribution and appropriation, integrating the communities engaged in the process. Distribution by the World Wide Web, and the proliferation of networked PCs, has transformed access and availability of bioinformatic tools and data, along the general ITC revolution. However, we have seen that many early public software packages charged for distribution, both to academic and commercial users, as part of cost recovery. As with scientific journals, costed distribution, although unquestionably restricting access, does not in itself disqualify the distributed knowledge from the public domain. Only when distribution is linked to, made conditional upon, an exchange of restricted rights over use, or to the institution of public rights over use, does distribution become an issue of private or public.

For, open, unrestricted, possibly free distribution, does not of itself make what is distributed a public good (Benkler 2006). This is not only the case for newspapers distributed 'free' to customers (financed, of course, by advertising paid for indirectly by those customers). But the more critical contrast is with the personal blog site, where the content of the site is entirely under the control of the individual, but the distribution is completely free, and access is unrestricted. Or, yet more acutely, material is both illicitly and freely distributed on the web, it may both undermine private property and undermine public control, so becoming against the public good through exploitation of vulnerable people. So the final feature of the public economy of knowledge also involves modes of public appropriation. For the public commons of knowledge to be sustainable as an economy of knowledge, therefore, particular historical modes of instituting public production, use, distribution and appropriation emerge dynamically, as we have analysed in the foregoing section. The commons is an everchanging combination of these four processes.

A few remarks relating to public 'ownership' of knowledge are in order. A positive conception of control, one that stresses the development of positive gains from control rather than a risk-avoidance conception, underlies our notion of

¹⁵ In Harvey and McMeekin (2007), we stress that this kind of public co-production can also involve combinations between public science institutions and commercial companies, though often with the tensions characteristic of proximity of the divide between public and private knowledge economies.

property, whether private or public. In considering public appropriation, the many modalities of control and their evolution are worth further exploration. Property, including most certainly public property, is a moving target. Here, three forms of public control, state-bureaucratic, formal legal, and expert community normative control, can be distinguished. Although frequently mentioned in the course of our analysis, the allocation of public resources to support co-production, even free distribution, are absolutely critical to the growth of the commons. Many of the funders of public research attach conditions to the allocation of resources to research projects for publication, use and dissemination. Broadly, this can be seen as the exertion of public control over public knowledge, underpinning co-production, distribution and use. In many countries, there has been experimentation with new forms of conditionality: how and in what ways knowledge is disseminated, how and in what ways knowledge is made usable and by which communities of users. The example of EMBOSS was given, where a help-desk and a generalist interface was stipulated in the research funding. Indeed, Stokes has argued for the need for a continuing renewal of the social contract around the financing of public research (Stokes 1995).

Alongside state bureaucratic modes of appropriation, there have been tentative and relatively underdeveloped legal instruments for instituting public appropriation. We have noted the use of General Public Licenses to underpin and consolidate open-source software for the public domain, by restricting opportunities for subsequent commercialisation. Another example, supported by the National Institute of Health in the US, is that of the Genetic Association Information Network (GAIN) certificates, which formally prevent the patenting of published genomic data. However, it is notable just how underdeveloped, in legal terms, these forms of public appropriation are compared with patent and copyright law.

However, what is perhaps most striking about the recent period is the emergence of new forms of public control from within the scientific communities of experts, especially the control over quality, standards and norms. These have extended well beyond the traditional and entrenched norms of peer review as a means of controlling public content and distribution of knowledge. Moreover, bioinformatic tools have been at the heart of this developing and innovative institutionalisation of public control. One of the most startling, indeed controversial, examples of such community-based public control was the establishment of the Bermuda Rules for genomic data deposition (Bentley 1996).¹⁶ This required any but the smallest strings of sequence data to be deposited on public databanks by all research communities, within 24 h. This not only ensured the continuous expansion of public databases and the scope of similarity searching by bioinformatic tools, but restricted opportunities for private appropriation and patenting. It was seen as a necessary underpinning to ensure international collaboration between laboratories, a common standard of practice in making data public. This 'institution' of the

¹⁶ Cf. http://www.ornl.gov/sci/techresources/Human_Genome/research/bermuda.shtml for the Bermuda rules.

commons, however, underwent significant changes over the next decade, with raw data being distinguished from finished or annotated data. Critically, Phil Green's quality score measures (phred and phrap) became the bioinformatic tools that ensure standards of quality of data worldwide. For data to be public domain data, on a global scale, new norms and conceptions of quality were institutionalised.

We have already remarked on the significance of SRS as a tool for integrating otherwise discrete bio-databases into a single public data resource. There is a proliferating range of bioinformatic tools developing standards for harmonising the 'ontologies' of data in diverse databases. The Microarray Gene Expression Data Society and the Macromolecular Structure Database as part of the worldwide Protein DataBank (wwPDB) both exemplify this process. Communities of experts develop common standards by which all abide, as a supportive infrastructure for knowledge production. The Gene Ontology Consortium project is aimed at constructing shared metalanguages enabling integration of genomic knowledge, enrolling a wide range of genomics laboratories across the world (<http://www.geneontology.org/>). Similarly, the Genomics Standards Consortium created a community of associates in 2005, laying down the minimum information standards for any deposition process incorporating data into public domain databases (<http://gensc.org/gsc/>). In this respect, a positive conception of control entails construction of shared constructs within which various communities agree to operate, as producers and users of biological knowledge. These examples demonstrate how the construction of norms for controlling the commons is an integral part of the scientific knowledge production, use and distribution process. As examples of 'self-regulation' within the public domain, they have shown continuing adaptation and evolution as the science itself develops.

In conclusion, instituting the commons is a complex, ongoing process, involving expanding societal resources supporting production, use, distribution and appropriation in ever novel ways. The commons is not a finite or fixed resource, to be used or exhausted, or a blank and given institutional space to be populated by public goods the attributes of which are universal or ahistorical. Alongside, and in dynamic tension with, the growth of private market and corporate organisations, the public commons has proved to be an evolving, innovative, institutional source of variation and experimentation. As such, it has been a primary motor of capitalist economic development.

Literature cited

- Allen, R. C. 2009. *The British Industrial Revolution in Global Perspective*. Cambridge: Cambridge University Press.
- Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215:403–410.
- Arrow, K. J. 1962. Economic welfare and the allocation of resources for invention. In *The Rule and Direction of Inventive Activity*, ed. R. R. Nelson, Princeton: Princeton University Press.

- Bairoch, A. 2000. Serendipity in bioinformatics, the tribulations of a Swiss bioinformatician through exciting times! *Bioinformatics* 16(1):48–64.
- Benkler, Y. 2006. *The Wealth of Networks. How Social Production Transforms Markets and Freedom*. New Haven: Yale University Press.
- Bentley, D. 1996. Genomic sequence should be released immediately and freely in the public domain. *Science* 274(5287):533–534.
- Boyle, J. 2003. The second enclosure movement and the construction of the public domain. *Law and Contemporary Problems* 66:33–74.
- Boyle, J. 2008. *The Public Domain. Enclosing the Commons of the Mind*. New Haven: Yale University Press.
- Callon, M. 1994. Is science a public good? *Science, Technology and Human Values* 19(4):395–424.
- Cook-Deegan, R. 1994. *The Gene Wars. Science, Politics and the Human Genome*. New York: Norton.
- Coriat, B., and F. Orsi. 2002. Establishing a new intellectual property rights regime in the United States. Origins, content and problems. *Research Policy* 31: 1491–1507.
- Dasgupta, P., and P. A. David. 1994. Toward a new economics of science. *Research Policy* 23:487–521.
- David, P. A. 1993. Intellectual property institutions and the panda's thumb: patents, copyrights, and trade secrets in economic theory and history. In *Global Dimensions of Intellectual Property Rights in Science and Technology*, eds. M. Wallerstein, M. Mogege, and R. Schoen, Washington, DC: National Academy Press.
- David, P. A. 2001a. From keeping “nature's secrets” to the institutionalisation of “open science”. Discussion Papers in Economic and Social History, 1–23, University of Oxford.
- David, P. A. 2001b. Tragedy of the public knowledge “Commons”? Global science, intellectual property and the digital technology boomerang. MERIT, Maastricht.
- David, P. A. 2004. Can “open science” be protected from the evolving regime of IPR protections? *Journal of Institutional and Theoretical Economics*, 160(1):9–34.
- David, P. A. 2008. The Historical Origins of “Open Science”: An essay on patronage, reputation and common agency contracting in the scientific revolution. *Capitalism and Society* 3(2), Article 5.
- Demsetz, H. 1967. Toward a theory of property rights. *American Economic Review* 57(2):347–359.
- Devereux, J., P. Haerberli, and O. Smithies. 1984. A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Research* 12:387–395.
- Doolittle, R. F., M. W. Hunkapiller, L. E. Hood, S. G. Devare, K. C. Robbins, S. A. Aaronson, and H. N. Antoniades. 1983. Simian sarcoma virus *onc* gene, *v-sis*, is derived from the gene (or genes) encoding a platelet derived growth factor. *Science* 221:275–277.

- Eisenberg, R. S. 2000. The public domain in genomics, unpublished manuscript, www.law.nyu.edu/ili/conferences/freeinfo2000/abstracts/eisengberg.html.
- Eisenberg, R. S., and R. R. Nelson. 2002. Public vs. proprietary science: a fruitful tension? *Daedalus* 131(2):89–101.
- Etzold, T., A. Ulyanov, and P. Argos. 1996. SRS: Information Retrieval System for Molecular Biology Data Banks. *Methods in Enzymology* 266: 114–128.
- Golinski, J. 1989. A noble spectacle. Phosphorus and the public culture of science in the early Royal Society. *Isis* 80:11–39.
- Golinski, J. 1992. *Science as Public Culture: Chemistry and Enlightenment in Britain, 1760–1820*. Cambridge: Cambridge University Press.
- Hardin, G. 1968. The tragedy of the commons. *Science* 162(3859):1243–1248.
- Hardin, G. 1998. Extensions of the “Tragedy of the Commons”, *Science* 280(5364):682–683.
- Harvey, M., S. Quilley, and H. Beynon. 2002. *Exploring the Tomato: Transformations in nature, economy and society*. Edward Elgar: Cheltenham.
- Harvey, M., and J. S. Metcalfe. 2004. The ordering of change: Polanyi, Schumpeter and the nature of the market mechanism. *Journal des Economistes et des Etudes Humaines* 4:87–114.
- Harvey, M. 2007. Instituting economic processes in society. In *Karl Polanyi: New Perspectives on the Place of the Economy in Society*, eds. M. Harvey, R. Ramlogan, and S. Randles, 163–184. Manchester: Manchester University Press.
- Harvey, M., and A. McMeekin. 2007. *Public or Private Economies of Knowledge? Turbulence in the Biological Sciences*. Cheltenham: Edward Elgar.
- Heller, M. A. 1998. The Tragedy of the anticcommons: property in the transition from Marx to markets. *Harvard Law Review* 111(3):621–688.
- Heller, M. A., and R. S. Eisenberg. 1998. Can patents deter innovation? The Anticcommons in biomedical research. *Science* 280(5364):698–701.
- Hodgman, T. C. 2000. A historical perspective on gene/protein functional assignment. *Bioinformatics* 16:10–15.
- Humphries, J. 1990. Enclosures, common rights, and women: The proletarianization of families in the late eighteenth and early nineteenth centuries. *The Journal of Economic History* 50(1):17–42.
- Kahn, B. Z., and K. L. Sokoloff. 2001. History lessons: the early development of intellectual property institutions in the United States. *Journal of Economic Perspectives* 15(3):233–246.
- Kaul, I., and R. U. Mendoza. 2003. Advancing the concept of public goods. In *Providing Global Public Goods*, eds. I. Kaul, P. Conceicao, K. Le Goulven, and R. U. Mendoza. Oxford: Oxford University Press.
- Landes, D. S. 2003. *The Unbound Prometheus. Technological Change and Industrial Development in Europe from 1750 to the Present*. 2nd Edition. Cambridge: Cambridge University Press.
- Lessig, L. 2004. The comedy of the commons. IT Conversations, <http://itc.conversationsnetwork.org/shows/detail349.html>).

- Lipman, D. J., and W. R. Pearson. 1985. Rapid and sensitive protein similarity searches. *Science* 227:1435–1441.
- Machlup, F., and E. Penrose. 1950. The patent controversy in the nineteenth century. *Journal of Economic History* 10(1):1–29.
- Marshall, E. 1999. Drug firms to create public database of genetic mutations. *Science* 284:406.
- McMeekin, A., M. Harvey, and S. Gee. 2004. Emergent bioinformatics and newly distributed innovation processes. In *The Economic Dynamics of Modern Biotechnologies: European and Global Trends*, eds. J. Laage-Hellman, M. McKelvey, and A. Rickne, 235–261. Cheltenham: Edward Elgar.
- Mokyr, J. 2002. *The Gifts of Athena. Historical origins of the knowledge economy*. Princeton: Princeton University Press.
- Moody, G. 2004. *Digital Code of Life: How Bioinformatics is Revolutionizing Science, Medicine and Business*. New Jersey: Wiley.
- Mowery, D. C., and B. N. Sampat. 2001. University patents and patent policy debates in the USA, 1925–1980. *Industrial and Corporate Change* 10(3):781–814.
- Mowery, D. C., and A. A. Zeidonis. 2002. Academic patent quality and quantity before and after the Bayh-Dole act in the United States. *Research Policy* 31:399–418.
- Mowery, D. C., R. R. Nelson, B. N. Sampat, and A. A. Ziedonis. 2001. The growth of patenting and licensing by US universities: an assessment of the effects of the Bayh-Dole act of 1980. *Research Policy* 30(1):99–119.
- Nee, V., and R. Swedberg. 2005. *The Economic Sociology of Capitalism*. Princeton: Princeton University Press.
- Nelson, R. R. 1959. The simple economics of basic scientific research. *The Journal of Political Economy* 67(3):297–306.
- Nelson, R. R. 1989. What is private and what is public about technology? *Science, Technology and Human Values* 14(3):229–241.
- Nelson, R. R. 2004. The market economy and the scientific commons. *Research Policy* 33(3):455–471.
- Nelson, R. R. 2006. Reflections on “The simple economics of basic scientific research”: looking backward and looking forward. *Industrial and Corporate Change* 15(6):903–917.
- O’Rourke, K. 2006. A phred for every platform? Next generation sequencers assess quality of data. *GenomeWeb News*.
- Polanyi, K. 1944. *The Great Transformation. The Political and Economic Origins of our Time*. Boston: Beacon Press.
- Polanyi, K. 1957a. The economy as instituted process. In *Trade and Market in the Early Empires*, eds. K. Polanyi, C. M. Arensberg, and H. W. Pearson. New York: The Free Press.
- Polanyi, K. 1957b. Aristotle discovers the economy. In *Trade and Market in the Early Empires*, eds. K. Polanyi, C. M. Arensberg, and H. W. Pearson. New York: The Free Press.

- Rai, A. K., and R. S. Eisenberg. 2003. Bayh-Dole reform and the progress of biomedicine. *Law and Contemporary Problems* 66:288–314.
- Rice, P., I. Longden, and A. Bleasby. 2000. EMBOSS: The European molecular biology open software suite. *Trends in Genetics* 16(6):276–277.
- Rosenberg, N. 1992a. Scientific instrumentation and university research. *Research Policy* 21:381–390.
- Rosenberg, N. 1992b. Economic experiments. *Industrial and Corporate Change* 1:181–203.
- Rosenberg, N. 1994. *Exploring the Black Box. Technology, Economics and History*. Cambridge: Cambridge University Press.
- Schaffer, S. 1983. Natural philosophy and public spectacle in the eighteenth century. *History of Science* 21:1–43.
- Shapin, S., and S. Schaffer. 1985. *Leviathan and the Air-Pump. Hobbes, Boyle and the Experimental Life*. Princeton: Princeton University Press.
- Sokoloff, K. L. 1988. Inventive activity in early industrial America: evidence from patent records, 1790–1846. *Journal of Economic History* 38(4):813–850.
- Stewart, L. 1992. *The rise of public science: rhetoric, technology and natural philosophy in Newtonian Britain, 1660–1750*. Cambridge: Cambridge University Press.
- Stokes, D. 1997. *Pasteur's Quadrant: Basic Science and Technological Innovation*. Washington, DC: Brookings.
- Sulston, J., and G. Ferry. 2003. *The Common Thread. Science, Politics, Ethics and the Human Genome*. London: Corgi Books.
- Waterfield, M. D., G. T. Scrace, N. Whittle, P. Stroobant, A. Johnsson, A. Wasteson, B. Westermark, C. H. Heldin, J. S. Huang, and T. F. Deuel. 1983. Platelet-derived growth factor is structurally related to the putative transforming protein p28sis of simian sarcoma virus. *Nature* 304(5921):35–39.
- Venter, J. C. 1998. *Prepared statement to US House of Representatives Committee on Science*.
- Venter, J. C., and M. D. Adams. 1998. Shotgun sequencing of the human genome. *Science* 280(5369):1540–1544.
- Zweiger, G. 2000. *Transducing the Genome. Informaton, Anarchy, and the Revolution in the biomedical sciences*. NY: McGraw Hill.