

Online Research @ Cardiff

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: <https://orca.cardiff.ac.uk/id/eprint/56597/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Lewis, Jamie Thornton ORCID: <https://orcid.org/0000-0002-1065-6017> and Bartlett, Andrew 2013. Inscribing a discipline: tensions in the field of bioinformatics. *New Genetics and Society* 32 (3) , pp. 243-263. 10.1080/14636778.2013.773172 file

Publishers page: <http://dx.doi.org/10.1080/14636778.2013.773172>
<<http://dx.doi.org/10.1080/14636778.2013.773172>>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies.

See

<http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



Inscribing a Discipline: Tensions in the field of Bioinformatics

Jamie Lewis and Andrew Bartlett

Published in *New Genetics and Society*

Abstract

Bioinformatics, the application of computer science to biological problems, is a central feature of post-genomic science which grew rapidly during the 1990s and 2000s. Post-genomic science is often high-throughput, involving the mass production of inscriptions (Latour and Woolgar, 1986). In order to render these *mass inscriptions* comprehensible, bioinformatic techniques are employed, with bioinformaticians producing what we call ‘*secondary inscriptions*’. However, despite bioinformaticians being highly skilled and credentialed scientists, the field struggles to develop disciplinary coherence. This paper describes two tensions mitigating against disciplinary coherence. The first arises from the fact that bioinformaticians as producers of *secondary inscriptions* are often institutionally dependent, subordinate even, to biologists. With bioinformatics positioned as service, it cannot determine its own boundaries but has them imposed from the outside. The second tension is a result of the inter-disciplinary origin of bioinformatics – computer science and biology are disciplines with very different cultures, values and products. The paper uses interview data from two different UK projects to describe and examine these tensions by commenting on Calvert’s notion of individual and collaborative interdisciplinarity and McNally’s (2008) distinction between ‘black box optimists’ and ‘black box pessimists’.

Key Words

Bioinformatics, Secondary Inscriptions, Technical Work, Genomics, Disciplines

Introduction

‘You then reached a point where you suddenly had something awfully like the *Amalgamated Union of Bioinformaticists* starting to say ‘no we do [bioinformatics], you don’t’ (Dr Harrison).

Science, argued Latour and Woolgar (1986), is a process of inscription; laboratory apparatus are ‘inscription devices’ that “transform a material substance into a figure or a diagram” (p51). Since the beginning of the Human Genome Project (HGP), biology has increasingly become high-throughput, a process of *mass inscription*, dependent on computers and mathematics to manage and interpret an unprecedented amount of data (Bayat, 2002). The Information Age has come to biology (Hood, 1992; Castells, 2000; Kay, 2000; Zweiger, 2001; Moody, 2004; Diamond and Woodgate 2005; Garcia-Sancho, 2012). The earlier successes of molecular biology, however, used apparatus such as microscopes, gels, mass spectrometers, and bioassays to achieve direct access to the components of life. This led some biologists to feel that an understanding of statistics and computing was irrelevant (Shrager, 2010). Our interviews with biologists in the UK confirm that many feel that their training as biologists has left them without the skills and knowledge required by this new era.

Since the mid-1990s, we have seen the rapid growth of bioinformatics as a central part of the post-HGP life science landscape (Boguski and McIntosh; Moody 2004; Diamond and Woodgate 2005). Bioinformatics – in its broadest sense, the application of computer science to biological problems – has been called on to clean, store, manage, and analyse the vast amounts of data produced by the high-throughput inscription devices of –omic, particularly genomic, science. This is especially true in the genetic dissection of complex disease by methods such as Genome Wide Association Studies (GWAS), which involves the genotyping of thousands of cases and controls at hundreds of thousands of genetic markers (Arribas-Ayllon, Bartlett and Featherstone (2010).

It is difficult – if at all possible – to identify the beginning of a historical process. It is equally difficult to refrain from presenting these processes as linear, purposeful developments punctuated by moments of scientific accomplishment and acclaim. Nevertheless, despite its significant growth since the 1990s, bioinformatics has a history that extends further into the past, and incorporates more than simply the analysis of genomic-era high-throughput data. Historians point towards independent developments in computing and DNA sequencing in the 1970s as precursors to the emergence of today’s genome-orientated bioinformatics (Suarez-Diaz, 2010: Garcia-

Sancho, 2012). Others identify pioneering work in protein research in the 1960s as a starting point (Smith, 1990). In particular, the protein sequencing work of Margaret Dayhoff is celebrated as the first attempt to use computers to solve biological problems (Dayhoff, 1969; Persidis, 1999; Strasser 2010). The late 1960s and 1970s also saw significant developments in the ARPAnet project – the precursor to the Internet. In 1973, Cerf and Kahn began linking the ARPAnet to other networks to create an ‘inter-network’ (Abbate, 2001). While early attempts to use the network to share biological databases were curtailed because ARPAnet was, primarily, a defence project, these developments in computer networking brought new ideas of interconnectivity and of data curation and storage; ideas which bioinformatics would put to use.

By the early-1980s, several groups were working on molecular biology databases. In April 1982, the publicly funded European Molecular Biology Laboratory (EMBL) was founded in Heidelberg, Germany, to co-ordinate molecular biology research in Europe. During the same period, the National Institute of Health (NIH) funded a US DNA sequence database, which became the Genetic Sequence Data Bank (GenBank) (Moody 2004). Towards the end of the 1980s, the institutional growth of biological information databases was solidified. In 1988, the European Molecular Biology Network (EMBnet) was established (EMBnet, 2006) producing a network linking local nodes - European laboratories using bio-computing, bio-statistics and bio-informatics - to a centralised international facility (Attwood and Parry-Smith, 1999). In the US, the National Center for Biotechnology Information (NCBI) was created at the National Institute of Health (NIH). Today, the NCBI produces cutting-edge research in computational biology, while developing and promoting standardised computer software tools for genomic and post-genomic data analysis (NCBI, 2005).

This short version of a long history shows bioinformatics and big biology as being epistemologically and institutionally tied together. In 1994, the European Bioinformatics Institute (EBI) was opened at the site of the Wellcome Trust Sanger Centre in Cambridgeshire, establishing an infrastructural resource for bioinformatics services and research in Europe at the home of ‘big biology’. High-throughput molecular biology has transformed biological objects into a mass of *primary*

inscriptions. It is the task of bioinformatics to make sense of this data by producing comprehensible *secondary inscriptions*.

Social scientists have shown an interest in the episteme that emerges when computer science plays an increasing role in the life sciences (see Hine, 2006; Leonelli, 2009), in technology as a boundary object or meeting zone (Cook-Deegan, 1994; Leonelli, 2010; Heeney and Smart, 2012), and in institutional and infrastructural aspects of bioinformatics and data curation (Wouters and Beaulieu, 2006; Baker and Millerand, 2010). We, however, are interested in the social tensions apparent as the field of bioinformatics stabilises. This paper describes the ways in which people working in bioinformatics understand their disciplinary identities; what it is to *do* bioinformatics, and to *be* a bioinformatician. We find that for some bioinformaticians, the fact that their work is dependent on the data produced by biologists, who have an established, powerful, disciplinary identity, places them in a subordinate position. This produces a situation which recalls Keefe and Potosky's (1997) description of technicians, for whom "professional career expectations formed in educational programs have rarely conformed to their job experiences" (p55). And, as with technicians (Shapin, 1989; Iliffe, 2008), the work of bioinformaticians, while essential, often takes place in the background. We document the variegated ways in which bioinformatics is understood by those working in the field, and those outside, and the effects that these judgements have on those who practise bioinformatics. There are tensions within bioinformatics; there are different ideas about the present and future of bioinformatics, and about the role of bioinformaticians.

The paper concludes by discussing the implications that such tensions might have for those working in the field. We remark on the possible trajectories of field, and of those working in it; trajectories which are a product of contrasting aspirations, epistemologies, and practices.

Defining (and Disciplining) Bioinformatics

Bioinformatics, as an inter-discipline, combines the skills and techniques of biology, computer science and statistics. For definitions of this kind, 'bioinformatics is broad, covering anything from epidemiology, the modelling of cell dynamics, to its now

more common focus, the analysis of sequence data of various kinds (genomic, transcriptomic, proteomic, metabolomic)' (Harvey and McMeekin, 2002).

These descriptions leave bioinformatics as a broad field of practices. Turner argues that one of the most important processes in disciplinary formation is the acceptance of a definition – the establishment of edges, of an inside and an outside – by the community in question (Turner, 2000). Scientists working in bioinformatics, however, present themselves in many guises; not only as bioinformaticians, but as biologists, computer scientists, statisticians, etc. The establishment of the National Institute of Health Bioinformatics Definition Committee (NIH BDC) in 2000 was an attempt to bureaucratically close the question 'what is bioinformatics'? Opting for an inclusive definition, they did not produce the boundaries that might produce disciplinary coherence. Bioinformatics was categorized as any...

...research development or application of computational tools and approaches for expanding the use of biological, medical, behavioural or health data including those to acquire, store, organize, archive, analyze or visualize data. (Huerta *et al.* 2000).

The broad nature of definitions such as this creates a space in which there are a variety of understandings and ambitions for bioinformatics, inevitably accompanied by tensions. The question of what bioinformatics encompasses, or perhaps more importantly, what it excludes, has simply not been answered. Indeed, Harvey and McMeekin argue that we 'should expect to find all kinds of different things going on under the rubric of bioinformatics, a fair amount of lack of integration between them, and no clear boundaries' (Harvey and McMeekin, 2002, p. 11). Fenstermacher (2004) has noted a similar proliferation of sub-divisions within bioinformatics¹.

What is bioinformatics, then? Or, more importantly, what will it be? Will it simply be a description of a heterogeneous collection of practices involved in the application of computing to biological research? Or will bioinformatics become something with a

¹ This is not the same as the general fissiparous tendency noted by scholars of higher education such as Delamont, Atkinson, and Parry (2000), in which an *established* discipline develops sub-fields, the boundaries of which harden over time.

coherent set of practices, and perspectives? If the latter, from where will bioinformatics draw its practice and perspective? Is it an application of computer science, or is it a biological discipline? What will its relationship with biology be? Is it to adopt the peculiar reward system and ethos of academia, or is it to be a specialist, technical service to biology?

The answers to these questions are inseparable from the social process of disciplining bioinformatics. Social scientists have long been interested in disciplinary formation and demarcation (cf. Merton, 1970; Lemaine *et al.*, 1976; Gieryn, 1983, 1999; Calhoun, 1992; Weingart and Stehr, 2000; Delamont, *et al.*, 2000). Their work highlights the way in which scientific boundaries are the product of social processes, rather than an organisational mirror to divisions in nature. Disciplines are not defined by their ideas, their intellectual coherence, or even their practices, but by institutional power. They are social and intellectual arrangements that shape our scientific and educational practices as well as our understandings of the world (Weingart and Stehr, 2000). Disciplinary identities are therefore understood as resemblances that are made object by funding agencies and assessment exercises that allocate resources on the basis of disciplinary membership (Turner, 2000). A bioinformatics that is not able to define and discipline itself will struggle to make the case for the allocation of research council funds, the appointment of bioinformaticians to senior positions, and the establishment of higher education programmes geared to producing academics. It will be defined from the outside, in a way akin to the manner in which the nature of technical work is determined not by its practitioners, but by the professional field from which technical work is ‘hived off’ (Keefe and Potosky, 1997).

Within the accounts that we collected, we found a kind of ‘boundary-work’ (Gieryn 1983) in operation, as scientists mark the intellectual activities of science as different from the technical and mechanical activities. As a fledgling inter-discipline, bioinformatics contains internal tensions as practitioners arrive at the field with different epistemes, interests, and expectations. As a producer of *secondary inscriptions*, bioinformatics is sometimes positioned subordinate to biology, seen as a service rather than a research discipline.

Methods

The paper draws on empirical evidence from two research projects conducted between 2004 and 2008, providing a snap-shot of post-HGP bioinformatics. The first project (2004) was a small-scale qualitative examination (10 interviews) of scientists working in the field of bioinformatics at one UK institution. These included interviews with respondents who identified themselves as biologists, as computer scientists, and as bioinformaticians. The majority worked on genomics and related research, but one interviewee was a biologist with a broader interest in the application of computing to biological problems. Most respondents were mid-career researchers, though one PhD student in bioinformatics was interviewed. All respondents were working in bioinformatics at a single research intensive university; the research focused on the disciplining of bioinformatics within that institution.

The second project (2004-2008), was a qualitative study of genomic scientists working at a number of UK universities. 31 interviews were supplemented with fieldnotes taken during conferences, workshops, bioinformatics courses, and informal meetings. Interviews were conducted with researchers from five different research-based UK universities working in the area of bioinformatics and proteomics. Respondents were a mixture of mid and later career researchers, with the exception of two PhD students. A senior manager involved in distributing bioinformatics funds was also interviewed. The perspective of bioinformatics presented in this paper is therefore from academic bioinformaticians in research-intensive universities working on the prominent branch of bioinformatics that deals with data from genomic and genomic-related data. The institutional background of the interviewee and the year in which the interview was conducted are provided after each extract to provide the reader with additional context.

Interviews and fieldnotes were analysed by content for emergent themes (Weber, 1990). Extracts used in this paper have been anonymised.

The Amalgamated Union of Bioinformaticists

The ‘boundary-work’ (Gieryn, 1982) in these accounts is practical as well as rhetorical. In order to become a discipline, boundaries must be established. From an

open, heterogeneous beginning, the boundaries of the field have hardened, according to Dr Harrison, a biologist and senior manager at a UK funding council:

When you have something very new that comes along, there is a sort of process of growth which is quite interesting [...] When bioinformatics first started, people started saying this is all very interdisciplinary [and] we have got to have no boundaries, we are drawing on skills from all sorts of people, we are terribly eclectic and it is open to all-comers and it is a very new field. It then starts to attract funding and then starts to develop a professional infrastructure of its own where people go to conferences, and they meet one another, and they start forming ideas of whether the people they are meeting are the same as them or not.

You then reached a point where you suddenly had something awfully like the '*Amalgamated Union of Bioinformaticists*' starting to say 'no we do it, you don't'. They put up little barriers and try to make sure they are fighting for their own corner, their own money and their own professional identity [...] The problem with bioinformatics is that it is not essentially biology so these were people in a field where they weren't doing the things that biologists did, they weren't doing experiments, not in the sense that biologists see experiments. So they were fighting the fact that they were up on the peer review panels so we would look at what they were doing and say this isn't biology, why are we paying for it? And other bizarre things like that (2005).

Dr Harrison describes the transition of bioinformatics from a field in which early inter-disciplinary idealism ruled, to one in which the need for coherency and strong self-identity has risen to the fore. The development of a disciplinary identity and a disciplinary infrastructure go hand-in-hand. A 'tribe' forms, which competes with others for the resources, material and otherwise, of academic society (Cook-Deegan 1994). This is a normal feature of the disciplining of research fields; staking out territory and building defences in order to keep others out (Becher, 1990). This disciplinary distinction is important as it feeds into university systems of reward and recognition (Bourdieu, 1988).

Dr Harrison, a biologist, describes bioinformatics as ‘not essentially biology’. His logic is based on the practices involved; bioinformaticians do not perform experiments, at least not in the way that biologists do. Their practice involves the manipulation of the *primary inscriptions* produced by biologists, rather than the transformation of the natural world into inscriptions. According to Dr Harrison, peer review panels conclude that many bioinformatics proposals should not be a priority for resources allocated to biological research, as they are not biology. A bioinformatician might read this as an argument for the development of an even stronger disciplinary identity – an *Amalgamated Union of Bioinformaticians* perhaps. For a biologist, the development of hard boundaries and a strong self-identity is unwelcome; rather than an autonomous discipline with an independent set of interests and perspectives, bioinformatics should remain in a symbiotic, subsidiary relationship to biology.

The suggestion that bioinformatics is not biology rests, in part, on the distinction between the ‘wet lab’ and the ‘dry lab’ (Penders, Horstman and Vos, 2008). As the labels suggest, the difference is not only one of location, but also one of practice. Both deal with biological problems, but they have different methods and approaches to biological objects. Bioinformatics work is often conducted away from traditional laboratories; at computers requiring no physical co-location with the biological objects under study. This entails a cultural, epistemic, and practical shift in the way of doing biology, from the manipulation of the material to produce inscriptions, to the manipulation of the symbolic; transforming existing *primary inscriptions* (Latour and Woolgar, 1986) into *secondary inscriptions* - a shift that mirrors wider changes in work in the Information Age (Castells, 2000).

Bioinformatics can be seen as a field of practice dependent on ‘wet lab’ biological sciences. Many participants questioned the idea that bioinformatics was, or could be, an autonomous discipline. Dr Woodbridge, a bioinformatician based in a biology department, thought of bioinformatics as a field of practice within biology:

I’ve got a slight problem with thinking of bioinformatics as this sort of separate, discrete subject in itself anyway. I think it’s an aspect of biology.
(2004)

Kirk, a research fellow in bioinformatics, also argues that bioinformatics is not an autonomous discipline. For Kirk, as with Dr Woodbridge, bioinformatics is a subordinate discipline, dependent on the existence of laboratory-based science to provide it with the material for its own explorations – data; the inscriptions of the wet lab.

I do not see bioinformatics to be a separate discipline [...] [It] has been hooking its nose into every discipline [...] I mean, it cannot be a separate discipline; it is a subset of every discipline. (2004)

The wet lab and the dry lab are very different places, in which very different forms of work are practised. The dependent, subsidiary, but essential nature of bioinformatics places it both at the epistemic centre of contemporary life sciences, and at the institutional edge.

While interested outside observers, such as Dr Harrison, have seen the boundaries of bioinformatics harden, this does not translate into a definition that every bioinformatician can share. Dr Hopper, a senior lecturer in bioinformatics based in a school of medicine, provided a typical response:

[I]t's a difficult question. I'm not really sure what the answer is [...] it's developing into a discipline, but it's not one specific thing, it's very wide-ranging (2006).

Importantly, when Kirk answered a similar question, he transformed it into a question bounded by practice.

[T]hen the second question comes, if you're doing bioinformatics, what are you doing? (2004)

The development of shared definitions and understandings helps new disciplines to form. This not only enables the sharing of concepts, data, and techniques, but also aligns ideas about the organization, trajectory, and self-identity of the emerging

discipline. If bioinformatics is to be a discipline, some degree of collective understanding of what the practice of the field entails is necessary. We found that the answer to the question, ‘what is bioinformatics?’ depends, to some degree, on the disciplinary hinterland of the bioinformatician. This background determines the value that they place on different kinds of data, knowledge, outputs, and practice, but also introduces tensions into the field.

Different camps coming from different directions

“[B]ioinformatics [is...] the progeny of the shotgun marriage between molecular biology and computer science” (Marijuán, 2002, p111). It is an ‘inter-discipline’. Bioinformatics, therefore, involves scientists from quite different backgrounds, who carry with them different interests, ambitions and perspectives. Or, as Heaton, then a PhD student in bioinformatics, put it, there are: “different camps coming from different directions”. Dr Griffiths, a reader in bioinformatics, was asked to define bioinformatics. His reply demonstrates how disciplinary background – in this case biology – can shape the view of the ‘inter-discipline’:

I always struggle with this one. For me it is a biologically driven problem. Essentially, I am a biologist that happens to use computers. I am more of an applied rather than a theoretical bioinformatician [...] we develop our own software and things but I am not a computer scientist, and I would never consider myself as one (2006).

Dr Griffiths’ interest in biological problems aligns his brand of bioinformatics with the perspectives of biology rather than those of computer science. He distances himself from computer science and, though he identifies himself as a bioinformatician, from *theoretical* bioinformatics too. Similar perspectives were offered by other bioinformaticians with backgrounds in biology. Jenkins, a PhD student in bioinformatics and molecular biology, describes how his disciplinary background shapes the way he works:

[O]bviously, coming from a biological background [...] I can look at it more sensibly by not having to ask trivial questions constantly to people in the group because I have the biological awareness (2006).

To Jenkins, the biological knowledge he has gained by virtue of his background is a prerequisite of the job. This perspective holds that, while bioinformatics draws on computer science, biology must remain the dominant partner in this ‘shotgun marriage’. With biology to the fore, bioinformatics will remain focused on questions of biological importance, avoiding ploughing the obscure computational furrows described by Dr Harrison.

By contrast, Dr Michaels has a background in mathematics and computer science. This brings a different perspective; that bioinformatics should remain an open ‘inter-discipline’, in which a variety of disciplinary perspectives are welcome:

Then I asked Dr Michaels for her definition of bioinformatics. 20 years ago, she said, it was a set of mathematical and statistical approaches to DNA analysis. Now, since the HGP, this has extended to also include data-base management and design. I asked her about the idea of a ‘bioinformatician’ identity developing, perhaps more homogenized than the varied group today, i.e. an identity where bioinformatician came first, rather than computer scientist or biologist. She remarked that this development might lead to a narrowing of outlook and that the beauty of bioinformatics is its differing perspectives of which linguistic mathematics is one. (Field Notes², 2004)

For Dr Michaels, biological data presents interesting computational and mathematical problems, aside from their biological significance. Given that a homogeneous bioinformatics could only be a bioinformatics closely tied to biological problems, an adjunct to biology, a heterogeneous future is one in which computer scientists can find academic fulfilment working in a broadly defined field.

Going further, Dr Campbell, a reader in bioinformatics and proteomics who has a background in computer science, argues that only a bioinformatics that is closely tied to computer science can claim disciplinary autonomy.

² Dr Michaels refused permission for the interview to be recorded, instead offering to speak slowly to allow notes to be taken.

I think that you can only define [bioinformatics] as a discipline, if you are really doing some cutting-edge research and you are using entirely new statistical computational or mathematical approaches in the area that hasn't been used before [...] I think then it becomes a discipline. Otherwise, I think it is a service or a facilitator for knowledge (2006).

In other words, the potential for bioinformatics to stand outside the control of biology lies not in its usefulness to, but in its difference from biology. The question of bioinformatics being positioned as a service to biology, and the tension between 'service' and 'discipline' conceptions of bioinformatics, is developed in more depth in the next section.

These extracts are not intended to persuade the reader that the 'camps' in which biologists, bioinformaticists and computer scientists pitch their tents are distinct intellectual objects with hard boundaries. However, divisions within a field of work have material implications. For Dr Campbell, the heterogeneity within bioinformatics prevents any one scientist becoming a 'complete' bioinformatician.

I don't think I have ever met people who have been equally interested in both [biology and computer science] and equally skilled in both. Almost by definition, people come from one discipline or the other. It is quite rare and unheard of, certainly in [the] top end of researchers, that they have been equally trained in both areas (2006).

For Dr Kennedy, a bioinformatician, the lack of genuine multi-skilled individuals is a problem for the specialism.

There are very few people who have got the ability to think biologically and also to think in terms of the computing needs. This is a problem (2006).

The existence of two camps demonstrates not just the epistemic problems of interdisciplinarity, but also the difference in interests. Dr Fairbrother, a lecturer in bioinformatics, argues that, as we found with Dr Michaels, many computer scientists

see biological problems as a proving ground for the theories and techniques of computer science, rather than interesting problems in their own right.

Let's say for instance we are supposedly solving these biological problems and all these sorts of things. But the computer scientists come in and it is more of the case of here is a problem that I can apply my pet technique that I have been working on in the past ten years. I can get some money to work on it, but it doesn't actually have to produce anything useful [...] It is quite difficult because I don't think the biological community has the expertise to drive things forward where it wants to be going. And if you can't drive yourself and you are being pulled by somebody who is not quite doing it for the right reasons you end up with this problem with what we are having. I don't think there is an obvious solution (2006).

For Dr Fairbrother, the future of bioinformatics will be determined by the interests of those who have the expertise, and power, to drive (or pull) bioinformatics forward. The dearth of computing and statistical expertise among 'wet lab' biologists results in bioinformatics being led by those with an education outside biology. Some biologists believe that these bioinformaticians have only a tangential interest in the underlying biological problems. Dr Fairbrother suggests that a lack of expertise in the biological camp and a lack of understanding or interest in biological problems in the computer science camp has led to squandered opportunities; 'you end up with this problem we are having'.

These extracts illustrate some of the difficulties of forming an 'inter-discipline'. Concentrating mostly on what Calvert (2010) terms 'individual interdisciplinarity', these participants argue that very few, if any, bioinformaticians have mastery of both biology and computer science. Two distinct camps of bioinformaticians result and members of these camps have differing epistememes, ambitions and interests. The idea of being trained equally in biology and computer science, of achieving individual inter-disciplinarity, was seen by many as not feasible. We argue that this is particularly the case during the early unfoldings of the field, when there was no formal bioinformatics education infrastructure. Bioinformatics has depended on individuals coming together from different disciplines – collaborative

interdisciplinarity (Calvert, 2010) – with all the problems of cooperation that are inherent in collaboration³.

Service versus Discipline

There is a widely accepted hierarchy of esteem in scientific disciplines; physics at the top, followed by chemistry and then biology to make up the ‘hard sciences’, which are collectively higher in this hierarchy than the ‘soft’ or ‘human’ sciences, which themselves have a hierarchy of hard to soft (Storer, 1967; Cole, 1983). However, the disciplines lower in this hierarchy are not subordinate to the ‘harder’ disciplines. Bioinformatics faces the question of whether it is to be an autonomous, academic, research discipline, an independent sub-discipline, or a service subordinate to biology. In attempting to establish bioinformatics as an autonomous field of research, bioinformaticians are dependent on the weight and value (both esteem and material) given to bioinformatics by biologists. People trained in the ‘wet lab’ are the strategic research planners, make up peer review panels, and dominate department hierarchies, which leads to an understanding of bioinformatics taking hold that is in the interests of those who do not work directly in bioinformatics. Perhaps this is why Dr Harrison finds something like the *Amalgamated Union of Bioinformaticians* has (informally) formed. We found that the ways in which biologists understand and position bioinformatics was a particular concern of bioinformaticians. From their view, biologists regard bioinformatics as a highly-skilled technical service. Dr Kennedy was asked whether he thought bioinformatics was a service or a discipline.

That is an excellent question. It is both. If you asked the bioinformaticists at my level, more often than not they will turn around and tell you it is a discipline. If you asked a biologist, they should say it is a service. They probably won’t, but they should, because 90% of biologists I deal with or have dealt with in the past view, whether they know it or not, bioinformatics as a service (2006).

³ It should be noted that collaborative interdisciplinarity is not always voluntaristic. It is not always the result of the collaboration of (more or less) autonomous scientists, but often a relationship between employer and employee, another mirror of the scientist-technician relationship. We prefer to downplay the equality implied by idealisation of ‘collaboration’, and instead suggest ‘collective interdisciplinarity’.

It is of note that Dr Kennedy, a bioinformatician, responded by describing the way in which biologists see, or more importantly *use*, bioinformatics. While a new discipline might work to shape itself – as in Dr Harrison’s description of hardening boundaries – it is often being positioned by those with existing power. Dr Kennedy suggests that even if biologists do describe bioinformatics as a discipline, the way in which they interact with bioinformaticians is to treat it as a service.

Dr Kennedy was asked to further describe how he, as a bioinformatician, perceives bioinformatics.

I view it as a discipline and it is a discipline because I suppose bioinformaticists at this level have their own research and are an interface between biology and computer science. So whether they [are] biologists or computer scientists, which they tend to be, it is still at that interface. In that respect, it is distinct as a research area [...] I mean research needs papers and the papers that come out are bioinformatics in nature, and they can be applied. They can be computer science in nature or they can be involved in novel biological-type data. So, it is cross-disciplinary in that respect (2006).

Dr Kennedy reiterated that bioinformatics is a discipline by virtue of having its own, autonomous research area. Bioinformaticians that are able to conduct research themselves, rather than offer skilled technical assistance to the research of biologists, are able to conduct work that is innovative and creative. Bioinformatics might be distinct, but, as Dr Kennedy describes, it is also cross-disciplinary, located at the boundary between computer science and biology. Just as bioinformaticians can be found in different ‘camps’, so the research papers produced by bioinformatics ‘can be computer science in nature or they can be involved in novel biological type data’.

A common feature of our interviews was that our participants did not feel that bioinformaticians are a homogenous group. Later in the interview, Dr Kennedy expands on his ‘at this level’ comment, constructing a boundary between research-driven and service-driven bioinformatics. By doing so, he distinguishes himself from others involved in bioinformatics-work who occupy a lower level in the hierarchy of academic esteem.

[A] bioinformatician tends to be on the service side as a graduate MSc student or maybe a Research Fellow who has gone to work as part of a team, but does not come up with their own research. Essentially, they are providing a service; a data analysis service.

[Whereas] a bioinformaticist is viewed as, and these definitions are all mine, a bioinformaticist tends to be someone who actually carries out the research. So there is definitely a distinction. There are bioinformaticians out there and far more bioinformaticians than bioinformaticists, if that is how they are termed. So yes that distinction does exist (2006).

That there are far fewer ‘bioinformaticists’ – those with a research identity – than there are service-orientated ‘bioinformaticians’ might be an ordinary feature of any field of human labour – that there are far more ‘doers’ than ‘thinkers’ – but some of our interviewees felt that this bias towards service provision in bioinformatics was an obstacle to the coherence of bioinformatics as a research discipline. For example, could it be that the positioning of bioinformatics as a service denies those working in the field the autonomy and opportunities to drive the field forward, to bring the biological and computing epistemes closer? The idiosyncratic terminology that Dr Kennedy uses has seemingly spread among the local community of bioinformaticians. Dr Fairbrother, a lecturer in bioinformatics, also uses ‘bioinformaticist’ and ‘bioinformatician’ as an ordering device to illuminate distinctions within bioinformatics.

I think the one that I am told I should use is bioinformaticist. The last definition I heard was that a bioinformatician is someone who uses bioinformatics tools, a bioinformaticist is somebody who develops them (2006).

Jenkins, a PhD student in bioinformatics and molecular biology, has also adopted these terms, as they have spread down the local hierarchy.

If you speak to my supervisor he would say there is a difference between a bioinformaticists and a bioinformatician...I think of myself as more of a developer than a service. So I am not necessarily the person that a biologist should come to... [even though] it might appear that I am. I actually have done that initially in my PhD where people come along and say I want to find a gene through an analysis of that. Will you find it please? But [this] is an area that I want to move away from and to move towards [...] developing applications rather than using them and giving people the results. I want them to do that rather than me (2006).

Jenkins wants to be viewed as a research bioinformatician, developing new applications and programs, rather than using existing ones to provide a technical service to biologists. However, he explains that he has been positioned, against his interests, perhaps as a product of the way bioinformatics is defined, as a service provider to biologists, being asked to conduct work over which he has no control.

The term collaboration implies a voluntary agreement between equals. Service bioinformaticians, however, are often not ‘collaborating’, but are contracted. The relationship is one of employer and employee, or, at least, service provider and client, not that of autonomous scientists collaborating in their mutual interests. Of course, few collaborations approach this ideal (Parker, Penders and Vermeulen, 2010), but in the case of the service bioinformaticians, the power and economics involved in their relationships with biologists is all too apparent. The nature of these relationships, in turn, has an impact on the rewards and recognition enjoyed by bioinformaticians.

The positioning of bioinformatics as a service recalls studies of professions and of technicians in the sciences. Discussing the professions, Abbott (1988) argues that each profession makes claims to all the knowledge and practice within its exclusive domain. He might point to a subordinate relationship between the offspring, bioinformatics, and the parent, biology (Abbott, 2001). That bioinformatics is dependent on the *primary inscriptions* produced by biologists – who belong to an existing, well-established discipline – means that their contribution is often set in the background. In the sciences, ‘routine’, but skilled and essential, procedures have long been delegated to ‘invisible’ technicians (from Shapin, 1989 through Keefe and

Potowsky, 1997, to Iliffe, 2008), and it is professional (or in this case, disciplinary) groupings that hold the social power to delegate low esteem tasks to other groups of workers.

When bioinformatics provides a service it allows the ‘black boxing’ of large parts of the process of making knowledge claims. Dr Nielson, a molecular biologist, describes his own relationship with bioinformatics and laments the difficulty of individual ‘interdisciplinarity’ (Calvert, 2010).

There are two questions there. Am I interested [in bioinformatics]? Yes. Do I have the time? No. That is the big problem, I am an amateur bioinformatics person... I do want to know how these things are doing. I think it is important from the point of view of interpreting papers so you can understand the differences, the interpretations of what are happening and what genes are being expressed and how they selected these things out. So I think it is very important that you understand at least the minimum part of bioinformatics for any research so you can understand that.

Do I have enough time and do I understand computing enough? Well, no to both of them. I understand a lot more but I don’t have the time to practise on them. We buy in a couple of ready-made programmes like Gene Spring and we have got Array Assist and... I get ones downloaded... which comes from TIGR, the Institute for Genome Research, which is very good, and it is free which is even better. It is a matter of playing around with them and making sure you know how they work, but I mean how they go through and do all these things no [I don’t understand]. The big one I don’t know and what I should really know is R or Bioconductor from the R programme, and I feel as though somebody else should do that for me. But we don’t really have that much bioinformatics support (2007).

Writing in 1990, Leroy Hood stated ‘...the HGP is training a new type of interdisciplinary biologist who understands technology as well as biology. I believe these multidisciplinary biologists will be among the scientific leaders of the 21st century’ (Hood 1990, p13). Nearly twenty years later, Bruun (2007) found that “most

bioscientists lack formal competence in bioinformatics, computer science, statistics and mathematics” (p190). Dr Neilson suggests that the answer to this is a closer relationship with bioinformaticians; a ‘collaborative interdisciplinarity’ rather than ‘individual interdisciplinarity’ (Calvert, 2010). However, Dr Cherry, another molecular biologist, exemplifies the way in which many bioscientists currently engage with ‘black boxed’ bioinformatics:

...I will use bioinformatics as a computer. I will use bioinformatics programmes, but I don’t know what the algorithms are and how they are constructed (2006).

Collaborative interdisciplinarity can involve the use of programmes or the expertise of people as black boxes, with buttons pressed and answers demanded, no knowledge of the workings required or expected. This is not an inherently bad thing; modern, technological society demands black boxing. However, the degree of black boxing that is desirable, in a science in which the goal is novelty and understanding, is open to debate, as is the feasibility of developing the individual interdisciplinarity envisaged by Hood.

While a service-orientated bioinformatics might frustrate the aspirations of some bioinformaticians, such a development would not necessarily be harmful to science. Service-orientated bioinformatics is, in effect, the black boxing of bioinformatics knowledge. McNally (2008) makes a distinction between ‘black box pessimists’ and ‘black box optimists’. An optimist argues for the benefits of black boxing knowledge within technologies and routinised processes. These technologies and processes free scientists from the need to be masters of that which can be mechanised or ‘hived off’. A black box need not be a techno-fix, but could just as easily be a rearrangement of the social. For a black box optimist, the role of bioinformaticians is to take concerns over data management, mining, and analysis away from biologists, either by offering their skills as a service or by creating routinised, user friendly applications, thus allowing biology to proceed apace in the genomic era. Collaborative or collective interdisciplinarity is acceptable. On the other hand, black box pessimists believe that there is a danger in closing off from scientists the assumptions and hypotheses that are contained in black boxes. Understanding the contents of the black box – in the case of

bioinformatics, the statistical theory and the specific algorithms that allow scientists to make knowledge claims – is necessary in order to do good science. Seen like this, individual interdisciplinarity is the ideal⁴.

Discussion

Canguilhem (1988), writing on the classification of scientific activities, argued that the history of science was the victim of ‘accepted’ classifications and that the real problem for historians and sociologists of science was to critically unravel why those classifications exist in the first place. What is accepted as bioinformatics practice, and what is not, determines its trajectory and the size, shape and distinctiveness of its contents.

The determination of bioinformaticians to avoid being categorised primarily as providers of a service recalls the findings of studies of technicians. Technicians are highly educated, but their knowledge is dependent on the discipline within which they work. Their knowledge is delegated knowledge. However, the education of technicians produces expectations that they ought to be part of a horizontally ordered work setting, such as that found in academic life, not that they are positioned in a vertically ordered work setting – providing a service (Nelsen, 1997). As Keefe and Potosky (1997) write, for technical workers, “professional career expectations formed in educational programs have rarely conformed to their job experiences” (p. 55). In the terms used by Flecker and Hofbauer (1998), the socialising process of an education in science “generates ‘superfluous’ subjectivity, that is, aspirations, desires for self-actualisation and so on, which exceed what is required” (p. 121). Bioinformaticians, with post-graduate degrees, socialised into the structures of academia, will find that their expectations will not match the employer-employee relationship of service provision.

Large-scale genomic science has created an increasingly acute division of labour (Arribas-Ayllon and Bartlett, *in press*) within the production of *primary inscriptions*. When we discuss the relationship between bioinformaticians (the producers of

⁴ Individual interdisciplinarity is a step on a path of infinite regress. Should scientists be able to understand the workings of all their own instruments? And the components for those instruments?

secondary inscriptions) and biologists (the producers of primary inscriptions), we are not discussing their relationship with *the* laboratory technician running seemingly endless genome scans, but with the group that controls this data, and with the senior bioscientists that hold authority within that groups. The creators of *primary inscriptions* are groups; the specific individuals doing the wet work are often afforded little status or power, but the primary inscriptions produced are scientific capital. It might also be noted that large-scale science produces a more instrumental division of labour than the cosy, collegial collaboration that is often seen as the scientific ideal (Yearly, 1988).

Because technical work is dependent on a body (or bodies, in the case of bioinformatics) of knowledge policed by autonomous scientific disciplines (or crafts and professions) the construction of the occupation of technicians is largely 'passive'. Professions, crafts, and academic disciplines are 'active' with regard to the boundaries of their occupations, set through struggles to control the system of knowledge, the processes of training and recruitment and so on. In contrast to this, the occupation of technicians is shaped by powerful actors from outside. In the case of bioinformatics, biologists have the power to build the bioinformatics that suits them, through control of departments, funding bodies, peer review, etc. As Keefe and Potosky (1997) write, "most technicians work in complex organizations where they neither set the entry and performance standards of their occupation, nor control the educational process through which new recruits are trained. Most cannot formally self-regulate or self-govern their work practices. Most important, technicians often operate within an established profession's field of knowledge and competence" (p. 54). Bioinformatics depends on the *mass inscriptions* of post-genomic science. As producers of *secondary inscriptions*, their work might be epistemologically central, but it is also institutionally dependent on (subordinate to, even) biology. The struggle for bioinformaticians, as they see it, is to ensure that they are able to build a disciplinary identity of their own, rather than an identity determined by biologists: the *Amalgamated Union of Bioinformaticians*, perhaps.

Building a disciplinary identity for an inter-discipline is not straightforward; bioinformaticians come from a number of disciplinary backgrounds, primarily from the life sciences and computer science. As described, this produces a situation in

which bioinformaticians conceive of their emerging discipline in quite different ways. Not just philosophically, but socially: which journals are read by their peers, which conferences are the focus of the community, what should be the curricula of undergraduate and postgraduate education, what sorts of activities are granted esteem? The uncomfortable positioning of bioinformaticians as service providers can, in part, be explained by the lack of consensus as to what bioinformatics *is*, but bioinformaticians see this as exacerbated by the ignorance of biologists of the practices of bioinformatics.

Furthermore, and of interest to an *Amalgamated Union of Bioinformaticians*, the black box pessimist is also concerned that the black boxing of knowledge has a deskilling effect, as the range of skills and knowledge required to do work narrows (McNally 2008, p223). Black boxed tools might even shield researchers from being aware that they lack knowledge of, and expertise in, the very processes that produce the knowledge that they publish. In the case of bioinformatics, the ability to perform bioinformatics at a distance – working on digitized data – means that biologists relying on bioinformaticians as service providers can also be shielded from the fact that their research is underwritten by the work of highly skilled interdisciplinary experts who are vital to the success of post-genomic projects (Lewis, 2010).

Of course, the possibilities are not limited to two polar opposites; service or independent discipline. Within the broad discipline of biology, there are a tremendous number of specialisms – divided by technique and subject matter – so, are the questions facing bioinformatics simply the same questions that have been faced by other biospecialities many times before? The key difference, we maintain, is that the majority of these specialisms produce *primary inscriptions*. In the case of bioinformatics though, bioinformaticians are reliant on these *primary inscriptions*, on the labour of biologists in the laboratory or the field, in order to produce *secondary inscriptions* that contribute to biological knowledge.

What then of the position of a discipline such as mathematics in biology, producing no inscriptions of its own? Comparison can be drawn. However, the difference here is one of institutional and cultural power. Mathematics is an established discipline with a solid identity before its tools and expertise are turned to biological problems. It is

secure – at least as much as is possible – with departments, courses, and prestigious chairs. Bioinformatics, on the other hand, arises out of the demands of biologists. And in the case of mathematics, the position in the hierarchy of the sciences (Storer, 1967; Cole, 1983) is clear; as the joke runs, “biologists answer only to chemists, chemists answer only to physicists, physicists answer only to mathematicians, and mathematicians answer only to God.” It seems that without the institutional and cultural power derived from centuries of departments, courses, prestigious chairs, and propagandists (comedic or otherwise), bioinformaticians must answer to biologists.

To conclude, we have highlighted two main tensions within the fledging field of bioinformatics: that between conceptions of bioinformatics as a service or as a discipline, and that between bioinformaticians coming to the field from the radically different hinterlands of biological science and computer science. The consequences of these two tensions are inter-related. The positioning of bioinformatics as a service to biology rather than an autonomous academic discipline may deny bioinformaticians the opportunities required to make advances, both professionally (by consequence of their peripheral access to the traditional reward structures of science) and scientifically. The complaint that few people in the field have the ‘individual interdisciplinarity’ (a solid grounding in both computer science and biology), is a complaint about barriers to scientific advancement. Pushing the boundaries of bioinformatics is the privilege of those with some degree of scientific autonomy coupled with access to the requisite level of interdisciplinarity (individual, collaborative, or collective). Given that bioinformatics is central to the production of knowledge in post-genomic science, its positioning on the institutional edges of the academy could be seen to be a dis-service, to bioinformatics and, consequently, the life sciences in general.

References

- Abbate, J. 2001. *Inventing the Internet*. London: MIT Press.
- Abbott, A., 1988. *The System of Professions: An Essay on the Division of Expert Labor*. Chicago: University of Chicago Press.
- Abbott, A., 2001. *Chaos of Disciplines*. Chicago: Chicago University Press.

- Arribas-Ayllon, M. and Bartlett, A. *in press* Sociological Ambivalence and the Order of Scientific Knowledge. *Sociology*.
- Arribas-Ayllon, M., Bartlett, A. and Featherstone K., 2010. Complexity and Accountability: The Witches Brew of Psychiatric Genetics. *Social Studies of Science*, 40 (4), 499-524.
- Attwood, T.K. and Parry-Smith, D.J., 1999. *Introduction to Bioinformatics*. Harlow: Addison Wesley Longman Limited.
- Baker, K.S. and Millerand, F. 2010. Infrastructuring Ecology: Challenges in Achieving Data Sharing. *In*: J.N. Parker, N. Vermeulen and B. Penders, eds. *Collaboration in the New Life Sciences*. Surrey: Ashgate Publishers, 111-138.
- Bayat, A., 2002. Science, Medicine and the Future: Bioinformatics. *British Medical Journal*, 324, 1018-1022.
- Becher, T., 1990. The Counter-Culture of Specialization. *European Journal of Education*, 15 (1), 333-346.
- Bourdieu, P., 1988. *Homo Academicus*. Stanford, CA: Stanford University Press.
- Boguski, M.S. and McIntosh, M.W., 2003. Biomedical Informatics for Proteomics. *Nature*, 422, 233-237.
- Bruun, H., 2007. Genomics and the Transformation of Knowledge: The Bioinformatics Challenge. *In*: P. Glasner, P.A. Atkinson and H. Greenslade, eds. *New Genetics, New Social Formations*. London: Routledge, 187-204.
- Calvert, J., 2010. Systems Biology, Interdisciplinarity and Disciplinary Identity. *In*: J.N. Parker, N. Vermeulen and B. Penders, eds. *Collaboration in the New Life Sciences*. Farnham: Ashgate, 201-219.
- Calhoun, C., 1992. Sociology, Other Disciplines and the Project of a General Understanding of Social Life. *In*: T.C. Halliday and M. Janowitz, eds. *Sociology and Its Publics: The Forms and Fates of Disciplinary Organization*. Chicago, IL: Chicago University Press, 137-195.
- Canguilhem, G., 1988. *Ideology and Rationality in the History of the Life Sciences*, trans. A. Goldhammer. Cambridge: MIT Press.
- Castells, M., 2000. *The Rise of the Network Society*. Oxford: Blackwell Publishers.
- Cole, S., 1983. The Hierarchy of the Sciences. *American Journal of Sociology*, 89, 140-165.
- Cook-Deegan, R., 1994. *The Gene Wars: Science, Politics and the Human Genome*. London: W.W. Norton and Company.

- Dayhoff, M.O., 1969. *Atlas of Protein Sequence and Structure*. (4th ed). Silver Springs: National Biomedical Research Foundation.
- Diamond, I. and Woodgate, D., 2005. Genomics research in the UK – the Social Science Agenda. *New Genetics and Society*, 24 (2), 239-252.
- Delamont, S., Atkinson, P.A. and Parry, O., 2000. *The Doctoral Experience: Success and Failure in Graduate School*. London: Falmer Press.
- European Molecular Biological Network (EMBnet). 2006. *What is EMBnet?* Available from: <http://www.embnet.org/node/73> > [Accessed 10 January 2006].
- Fenstermacher, D., 2004. Introduction to Bioinformatics. *Journal of the American Society for Information Science and Technology*, 56 (5), 440-446.
- Flecker, J. and Hofbauer, J., 1998. Capitalising on Subjectivity: The ‘New Model Worker’ and the Importance of Being Useful. In: P. Thompson and C. Warhurst, eds. *Workplaces of the Future*. London: Macmillan, 104-123.
- Garcia-Sancho, M., 2012. *Biology, Computing and the History of Molecular Sequencing: From Proteins to DNA, 1945-2000*. Basingstoke: Palgrave Macmillan.
- Gieryn, T.F., 1983. Boundary-work and the Demarcation of Science from Non-Science: Strains and Interests in Professional Interests of Scientists. *American Sociological Review*, 48, 781-795.
- Gieryn, T.F., 1999. *Cultural Boundaries of Science: Credibility on the Line*. Chicago: University Chicago Press.
- Harvey, M. and McMeekin, A., 2002. *UK Bioinformatics: Current Landscapes and Future Horizons*. London: Department of Trade and Industry.
- Heeney, C and Smart, A., 2012. Enacting Governance – the Case of Access. In: J. Kaye, ed. *Governing Biobanks: Understanding the Interplay Between Law and Practice*. Oxford: Hart Publishers, 232-258.
- Hine, C., 2006. Databases as Scientific Instruments and their role in the Ordering of Scientific Work. *Social Studies of Science*, 36, 269-298.
- Hood, L., 1990. Human Genome Project: Is Big Science Bad for Biology? No, and anyway the HGP isn’t Big Science. *The Scientist*, 4 (22), 13-15.
- Hood, L. 1992. Biology and Medicine in the Twenty-First Century. In: D.J. Kevles, and L. Hood, L, eds. *The Code of Codes: Scientific and Social Issues in the Human Genome Project*. Cambridge, MA: Harvard University Press, 136-163.
- Huerta, M., Haseltine, F., Liu, Y., Downing, G. and Seto, B., 2000. *NIH Working Definition of Bioinformatics and Computational Biology* [online]. National Institute

of Health. Available from: <http://www.bisti.nih.gov/docs/CompuBioDef.pdf> [Accessed 10 January 2006].

Iiffe, R., 2008. Technicians. *Notes and Records of The Royal Society*, 62: 3-26.

Kay, L.E., 2000. *Who Wrote the Book of Life? A History of the Genetic Code*. Stanford, CA: Stanford University Press.

Keefe, J. and Potosky, D., 1997. Technical Dissonance: Conflicting Portraits of Technicians. In: S.R. Barley and J.E. Orr, eds. *Between Craft and Science*. Ithaca, NY: Cornell University Press, 53-81.

Lemaine, G., Macleod, R., Mulkay, M., and Weingart, P., 1976. Problems in the Emergence of New Disciplines. In: G. Lemaine, R. Macleod, M. Mulkay and P. Weingart, eds. *Perspectives on the Emergence of Scientific Disciplines*. Chicago, IL: Mouton and Co, 1-23.

Latour, B. and Woolgar, S., 1986. *Laboratory Life: the Construction of Scientific Facts*. Princeton, NJ: Princeton University Press.

Leonelli, S., 2009. On the Locality of Data and Claims About Phenomena. *Philosophy of Science*, 76 (5), 737-749.

Leonelli, S., 2010. Documenting the Emergence of Bio-ontologies: or why Researching Bioinformatics Requires HPSSB. *History and Philosophy of the Life Sciences*, 32 (1), 105-126.

Lewis, J., 2010. Matchmaking Mechanisms: Collaborative Arrangements in Proteomics and Bioinformatics. In: J.N. Parker, N. Vermeulen and B. Penders, eds. *Collaboration in the New Life Sciences*. Surrey: Ashgate Publishers, 179-200

Marijuán, P.C., 2002. Bioinformatician: Untangling the Networks of Life. *Biosystems*, 64, 111-118.

McNally, R., 2008. Sociomics: CESAGen Multidisciplinary Workshop on the Transformation of Knowledge Production in the Biosciences, and its Consequences. *Proteomics*, 8, 222-224.

Merton, R.K., 1970. *Science, Technology and Society in Seventeenth Century England*. New York, NY: Howard Fertig.

Moody, G., 2004. *Digital Code of Life: How Bioinformatics is Revolutionizing Science, Medicine and Business*. New Jersey, NJ: John Wiley and Sons.

National Center for Biotechnology Information (NCBI). 2005. *Home Web Page* [online]. Available from: <http://www.ncbi.nlm.nih.gov/> [Accessed 22 October 2005].

- Nelsen, B.J., 1997. Work as a Moral Act: how Emergency Medical Technicians Understand their Work. *In: S. Barley and J. Orr, eds. Between Craft and Science: Technical work in US Settings.* Ithaca, NY: Cornell University Press, 154-184.
- Parker, J.N., Vermeulen, N and Penders, B., 2010. *Collaboration in the New Life Sciences.* Farnham: Ashgate.
- Penders, B., Horstman, K. and Vos, R., 2008. Walking the Line Between Lab and Computation: The Moist Zone. *Bioscience*, 58 (8), 747-755.
- Persidis, A. 1999. Bioinformatics. *Nature Biotechnology*, 17, 828-830.
- Shapin, S., 1989. The Invisible Technician. *American Scientist*, 77, 554-563.
- Shrager, J., 2010. From Wizards to Trading Zones: Crossing the Chasm of Computers in Scientific Collaboration. *In: M.E. Gorman, ed. Trading Zones and Interactional Expertise: Creating New Kinds of Collaboration.* Cambridge, MA: The MIT Press, 107-124.
- Smith, T.F., 1990. The History of Genetic Sequence Databases. *Genomics*, 6, 701-707.
- Storer, N.W., 1967. The Hard Sciences and the Soft: Some Sociological Observations. *Bulletin of the Medical Library Association*, 55 (1), 75-84.
- Strasser, B., 2010. Collecting, Comparing and Computing Sequences: The Making of Margaret O. Dayhoff's Atlas of Protein Sequence and Structure, 1954-1965. *Journal of the History of Biology*, 43, 623-660.
- Suarez-Diaz, E., 2010. Making Room for New Faces: Evolution, Genomics and the growth of Bioinformatics. *History and Philosophy of the Life Sciences*, 32, 65-90.
- Turner, S., 2000. What Are Disciplines? And How Is Interdisciplinarity Different? *In: P. Weingart and N. Stehr, eds. Practising Interdisciplinarity.* Toronto: University of Toronto Press, 46-65.
- Weber, R., 1990. *Basic Content Analysis.* London: Sage.
- Weingart, P. and Stehr, N., 2000. *Practising Interdisciplinarity.* Toronto: University of Toronto Press.
- Wouters, P and Beaulieu, A., 2006. Imagining E-science Beyond Computation. *In: C. Hine, ed. New Infrastructures for Knowledge Production: Understanding E-Science.* London: Information Science Publishing, 48-70.
- Yearly, S. 1988. *Science, Technology and Social Change.* London: Unwin Hyman.
- Zweiger, G., 2001. *Transducing the Genome: Information, Anarchy and Revolution in the Biomedical Sciences.* New York, NY: McGraw Hill.

Acknowledgement

The support of the ESRC is gratefully acknowledged. This work is part of the research programme of the ESRC Genomics Network at Cesagen.