

Social Studies of Science

<http://sss.sagepub.com>

Databases as Scientific Instruments and Their Role in the Ordering of Scientific Work

Christine Hine

Social Studies of Science 2006; 36; 269

DOI: 10.1177/0306312706054047

The online version of this article can be found at:
<http://sss.sagepub.com/cgi/content/abstract/36/2/269>

Published by:



<http://www.sagepublications.com>

Additional services and information for *Social Studies of Science* can be found at:

Email Alerts: <http://sss.sagepub.com/cgi/alerts>

Subscriptions: <http://sss.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.co.uk/journalsPermissions.nav>

Citations <http://sss.sagepub.com/cgi/content/refs/36/2/269>

ABSTRACT Speculation on the implications of increased use of information and communication technologies in scientific research suggests that use of databases may change the processes and the outcomes of knowledge production. Most attention focuses on databases as a large-scale means of communicating research, but they can also be used on a much smaller scale as research tools. This paper presents an ethnographic study of the development of a mouse genome mapping resource organized around a database. Through an examination of the natural, social and digital orderings that arise in the construction of the resource, it argues that the use of databases in science, at least in this kind of project, is unlikely to produce wholesale change. Such changes as do occur in work practices, communication regimes and knowledge outcomes are dependent on the orderings that each database embodies and is embedded within. Instead of imposing its own computer logic, the database provides a focus for specifying and tying together particular natural and social orderings. The database does not act as an independent agent of change, but is an emergent structure that needs to be embedded in an appropriate set of work practices.

Keywords collaboration, database, ethnography, genome, information, laboratory science

Databases as Scientific Instruments and Their Role in the Ordering of Scientific Work

Christine Hine

The use of information technology in scientific research has increased dramatically in recent years. Databases provide a means of gathering together vast amounts of data, and performing investigations on those data that would be inconceivable without the use of computers. Whole new fields of research, notably genomics, have formed around massive data collections, and the use of genomic and proteomic databases has become a routine part of the daily work of biologists (Brown, 2003). Not all uses of databases are on such a large scale, however. Many researchers use personal databases to store and organize their results. Databases can also be created as shared resources or research tools on a much smaller scale than the major genomic databases. This paper gives an account of the development of one such database as a scientific resource. The aim, in the first instance, is to provide an ethnographic account of an important aspect of contemporary scientific practice that is still under-explored. The further aim is to use this account to engage with debates about the significance of

Social Studies of Science 36/2 (April 2006) 269–298

© SSS and SAGE Publications (London, Thousand Oaks CA, New Delhi)

ISSN 0306-3127 DOI: 10.1177/0306312706054047

www.sagepublications.com

increasing use of information and communication technologies in knowledge production, specifically how far databases impose a particular form of computer logic on knowledge production, and to what extent they provide new communication regimes, new forms of collaboration and new spatial organizations for science.

I begin by reviewing perspectives on the use of information technology in science and on the organization of scientific communication and practice that suggest key issues for ethnographic enquiry. I then discuss an analytic framework that was chosen for its ability to explore significant aspects of scientific practice without assuming in advance what functions databases serve. This leads into discussion of a case study of the development of a mouse genetics database as a tool for genetic mapping, focusing on its orderings of the natural, social and digital worlds. The conclusion explores implications for researching the role of databases in science.

Databases in Science, and the Organization of Laboratory Work

Change in science has often been attributed to changes in technology, even when the relationship is less than straightforward (Galison, 1997b). The role of databases within science has lately become topical, particularly in biomedicine, which Lenoir (1998b) argues has increasingly aimed to become an information science. Such culture change in molecular biology is also claimed by biologists themselves. Jordan & Lynch (1998: 773) describe high-profile commentaries that characterize modern molecular biology as 'big science', dominated by information science and standardized procedures. Crucial in justifying the view of a unified biology as an information science is the rhetoric that 'there is no outside of the genetic text' (Doyle, 1998: 304), which suggests that knowing the genome is the key to the rest of biology. The goal of achieving this unified theoretical basis for the discipline has, Lenoir suggests, shaped the activities of many biologists and directed their attention towards projects involving information technology. Advances in genomic science have offered glimpses of new ways of theorizing. This, Lenoir claims, may have radical consequences for the future organization of biology:

Many molecular biologists who welcomed the Human Genome Initiative undoubtedly believed that when the genome was sequenced, everyone would return to the lab to conduct their experiments in a business-as-usual fashion, empowered with a richer set of fundamental data. The developments in automation, the resulting explosion of data, and the introduction of tools of information science to master this data have changed the playing field forever. There may be no 'lab' to return to. In its place is a workstation hooked to a massively parallel computer producing simulations by drawing on the data streams of the major databanks and carrying out 'experiments' *in silico* rather than *in vitro*. The results of biology's metamorphosis into an information science just may be the relocation of the lab to the industrial park and the dustbin of history. (Lenoir, 1998b)

In this paper I hope to show that this somewhat 'cyberbolic' (Woolgar, 2002) vision of the relationship between databases and science is not exhaustive. There will, I argue, prove to be other ways of organizing work and other ways of using databases, and not all will erase the laboratory in the way that Lenoir suggests.¹ This vision does, however, provide a useful introduction to some key issues raised by databases: changes to the frameworks for evaluating research; changes to the work practices of scientists; and changes to the spatial environments in which science gets done. In the remainder of this section I explore recent literature that sheds further light on these issues.

The first set of issues, concerning databases and changes in the criteria for judging scientific work, is addressed by Beaulieu (2001) in a study of the development of computerized atlases of the brain. She explores altered ideals of objectivity in brain science revolving around the new atlases. These developments in 'style of organization and validation of knowledge' (Beaulieu, 2001: 637) involve increasing use of computerized techniques to bring together multiple sources of data and carry out averaging and probabilistic analysis. These technological capacities provide for new ways of combining, juxtaposing and exploring brain scans that give rise, Beaulieu argues, to a new ideal of objectivity, which then provides a basis for criticizing earlier atlases. Accepting the new forms of atlas encourages replacement of the expertise of skilled human observers with automated judgement. Gathering the large samples that such an atlas requires also entails developing conventions, standards and automated procedures. Consequently, on a variety of levels, the goals, values and practices of brain science change in the face of the new technology's acceptance.

Lenoir and Beaulieu are unusual in that they focus on change in the processes and outcomes of research. More often, discussions of information technologies, particularly large-scale databases, focus on how they change the dissemination of research results and protocols. Nentwich (2003), in a comprehensive review of literature on computer culture and computers in science, combined with his own comparative study of the 'cyberness' of different disciplines, suggests a range of dimensions along which the advent of computer-mediated communication may impact academic research cultures and practices, including: spatial organization; distributions of roles; representation of knowledge; quality control and allocation of credit; and economic and legal aspects of publishing. However, Nentwich's focus on communication and explicit exclusion of research tools limits his treatment of databases to various forms of digital library and archive. While he discusses the importance of disciplinary databases as standard tools, and speculates on the possibility of new forms of collaboration via shared knowledge bases, his focus on communication omits some of the forms that databases can take.

The same is true of Hilgartner's (1995) discussion of databases in the context of the 'communication regimes' of science. Communication regimes are established networks for the dissemination of science, comprising actively constituted systems of technical and social arrangements.

While the journal is the iconic regime, Hilgartner suggests that databases provide the occasion for new regimes to develop. He discusses three different regimes, each of which involves a database but puts in place different arrangements for acquiring, validating and circulating data, for determining the extent to which the system is centralized, and for establishing the relationship with journals. Databases themselves are not new communication regimes, but they can provide the impetus for diverse new regimes to form. Thinking of databases in this way does, however, tend to direct attention away from smaller-scale uses and occasions on which they act as research tools or instruments. In this paper I focus on a database that does communicate data, but also acts as a research instrument, and is not therefore adequately captured by the concept of the communication regime. The approach I take is more in sympathy with the concept of the 'data stream' developed earlier by Hilgartner & Brandt-Rauf (1994), as I will discuss later in this introduction. First, however, the question of databases and forms of knowledge needs further examination. Whether as communication regimes or research tools, some of the interest in databases has been created by speculation that they might not simply portray knowledge, but shape what counts as knowledge.

The use of databases in science raises the prospect that knowledge itself might be altered by new representational technologies. For example, Manovich (2001) attributes the impact of computers on culture to a changed mode of thinking, in which the world comes to be seen as composed of data structures and algorithms, fundamentally changing our relationship to it. This is, for Manovich (2001: 223), 'the projection of the ontology of a computer onto culture itself'. However, social shaping perspectives (Bijker, 1997) caution against explanations based on overriding technological logics. This caution is taken on board by Nentwich (2003), who discusses how the 'shadow of the format' can shape knowledge representation, but more through the practices and expectations of users than through a direct effect of the technology itself. Lenoir (2000) suggests that computer-mediated communication should be investigated as a site of new practices rather than as an agent of change in its own right. Elsewhere he suggests the need to study:

... the historically situated character of scientific representation, its multivalent and contested nature and the investment of scientific argument in narrative structures, vocabularies, grammars, patterns of analogy, and metaphor, both internal and external to the scientific text. (Lenoir, 1998a: 16)

Applying this perspective to the use of information technology in the laboratory, we might expect to find a more varied picture than Manovich (2001) suggests.

The effects of information technology on biology may be less than universal. There is also a limit to their predictability. Unintended consequences can become irreversibly embedded when new infrastructures are developed (Bowker & Star, 1999). Bowker (2000) reviews developments in

the production of biodiversity archives, and suggests that combining databases involves political and ethical decisions as much as technical decisions. In creating a large-scale infrastructure, database designers may be shaping the future possibilities, not just of their databases, but also of the science that revolves around them and the conservation activities that may ultimately depend upon them. Beaulieu's (2001) review of the development of brain atlases shows how these tools are embedded into wider policy aims and have effects on knowledge production that go beyond straightforward improvements in efficiency.

The development and use of databases may provide sites at which scientific work practices and communication regimes are reconfigured, whether by design or by accident, along a single trajectory or in multiple different ways. As Beaulieu (2001: 636) suggests, 'cyberscience' (Wouters, 2000) may prove to represent 'a novel, technologically-supported organization of knowledge-production, in which the digital format and electronic transmission of data figure prominently'. The nature and implications of this novelty are subjects for research, rather than issues to be extrapolated from the nature of the technology. To that end I use Knorr-Cetina's (1999) vision of science as composed of distinctive social and natural orderings to consider how they may be altered or supplemented by the advent of 'digital orderings' in the construction of the database. In discussing this framework I bring in the third of the key themes from Lenoir's (1998b) vision of a transformed biology, the spatial organization of scientific work.

In *Epistemic Cultures*, Knorr-Cetina (1999) compares the social and natural orderings present in two laboratories, working within different scientific disciplines. She investigates a 'machinery of knowledge construction' made up of the empirical, the social and the technological: each plays its part in contributing to a unique epistemic culture, within which the scientist is produced as a particular kind of epistemic subject. In Knorr-Cetina's (1999) view, the laboratory is the site for assembling the machinery of knowledge construction. It is, however, more than a physical space. The laboratory produces and contains specific orderings of both natural and social worlds, and it is these different orderings that comprise a distinctive machinery of knowledge production. The natural orderings in the laboratory enable objects to be detached from their usual contexts, and made amenable to observation and intervention. The social ordering ties together laboratory staff by assigning particular sets of responsibilities, obligations and rights. These orderings manifest differently in the two disciplines that Knorr-Cetina considers. They also can change over time, and with the advent of new technologies and work practices. Knorr-Cetina's view of molecular biology is based on observations from 1984 until the early 1990s. This raises an empirical question: has molecular biology been transformed in its natural and social orderings since this period? How would a study conducted several years later compare with her observations? How might the introduction of databases have altered the situation?²

Lenoir, as noted earlier, suggests that informational perspectives will change the spatial organization of science. For Knorr-Cetina, like many other analysts, the spatial organization of science is crucial for the outcomes. While it is too simple to say that the physical place of the laboratory alone makes the activities within it scientific (Lynch, 1991), there is frequent suggestion that the boundedness and isolation of the laboratory enable particular epistemic practices both to be contained and to be held separate from the world at large (Ophir & Shapin, 1991; Livingstone, 1995; Shapin, 1995; Galison, 1999). While databases can be compiled by one person, on one computer, and within a single laboratory, far more often they are produced as collaborative projects across sites and used from remote locations. Spatial change therefore forms one of the themes to consider in the present case study: does the laboratory form a key site? What other locations are introduced in the course of the project? How are communication and collaboration across sites managed?

The coming of large-scale data collection has been experienced as problematic within the research cultures of biology. A recent Wellcome Trust report (2003) singles out community resource projects as important contributions that entail tensions over recognition for their creators. Beaulieu (2003) discusses a data-sharing initiative that met considerable difficulties in motivating data submissions and overcoming lack of trust from researchers. Hilgartner & Brandt-Rauf (1994) argue that data-sharing in science is an appropriate yet neglected focus for ethnographic enquiry, provided that we 'define data broadly, to include such entities as biological materials, reagents, novel instrumentation, and other scarce or limited resources' (1994: 356). The question of sharing access to data between scientists raises a spectrum of questions relating to the ways in which objects are constituted and boundaries defined. In Knorr-Cetina's terms, data-sharing becomes a question of achieving social and natural orderings, which decide both what is to be shared, and how to share it. To return to an earlier point, whether this looks like the creation of a communication regime or the development of a scientific instrument is an outcome of the process, rather than an issue to be decided in advance.

In summary, Knorr-Cetina's portrayal of science as diverse epistemic cultures, constituted through distinctive social/natural orderings, provides the lens through which the ethnographic case study is presented, and frames the following questions:

- How are natural and social orderings achieved in the development of a database resource?
- How do the social orderings produce collaboration, trust and data-sharing?
- Do novel 'digital orderings' (akin to Manovich's [2001] computer ontology) influence social and natural orderings? Do these diminish the importance of the spatial ordering of the laboratory?

- Does the advent of such projects involve a reconfiguring of the social and natural orderings of scientific practice, so as to represent a distinct epistemic culture?

These questions are explored through a multi-site ethnographic approach that follows a data resource through its construction, use and eventual placement within the scientific literature. Overall, the aim was to explore the many sites in which the data resource was enacted, in order to gain a rich perspective of the social and natural orderings in which it participated.

The Ethnography of a Mouse Mapping Resource

The resource being considered here was conceived as an aid to mapping the mouse genome. Mouse genomics are generally held to be relevant to human genomics, because the genetic makeup of the mouse is sufficiently similar to that of the human for it to be a 'model organism' for the study of human disease. It is much easier to perform selective breeding experiments and build up data from successive generations with mice than with human beings, and thus the mouse acquires the dubious status of 'honorary human'. The mouse has also been highly developed as a laboratory technology (Löwy & Gaudillière, 1998) in similar fashion to *Drosophila* (Kohler, 1994). While the lead laboratory in the production of the resource described here does work on mouse models for human disease, this particular resource will not in itself contribute directly to the understanding of human disease. The resource's main advantages are expected to be for mouse genomics, enabling quicker and more accurate localization of particular genes. Study of human disease through mouse models may benefit eventually, but only indirectly. A further dimension of the project is that it represents a collaboration between various prominent European laboratories, but not the leading competitors in the USA, and hence it is designed to act as a visible European achievement.

In technical terms, the resource is in the first instance an aid to mapping the mouse genome via genetic linkage analysis.³ Selective breeding of specific strains of mice produces a set of DNA, which is used as the input to a process of linkage mapping. This type of mapping depends on an understanding of the behaviour of chromosomes when cells divide. Individual chromosomes are not passed down between generations as intact units. At cell division cross-over between chromosomes can redistribute genes. The closer together two loci are on a chromosome, the less likely it is that a cross-over event will occur between them, and thus the greater the likelihood, broadly speaking, that the markers or genes at these loci will be found together in the next generation. Linkage mapping is thus based on the analysis of patterns of segregation of markers, and on predicting the most likely ordering of loci on chromosomes given those patterns.⁴ The resulting map is a statistically informed achievement based on the assumption of the least number of chromosome cross-overs necessary to produce the observed pattern.⁵

The initial phase of resource development aimed to provide a 'skeleton map' by placing a set of genetic markers systematically chosen to provide coverage across all chromosomes. This beginning would subsequently facilitate high-resolution mapping. The database would allow future researchers to locate the likely positions of new markers via a linkage analysis searching for patterns of segregation between the new marker and the skeleton map markers. Such mapping would be speeded up by use of the database to identify mice that are recombinant in particular chromosome regions of interest, and hence likely to be productive in mapping a new marker in that region. Programs associated with the database would calculate the likely ordering of markers on the map. The database containing the map information was to be made available on the World Wide Web. To use the map in order to place new markers it would, however, be necessary for researchers to have access to DNA samples from the original set of mice, in addition to having access to the database. In conjunction with the DNA samples, the database acts as a scientific instrument for chromosome mapping.

Beyond the provision of a genetic map based on linkages, the next goal was to use this information to help in production of a physical map. A genetic map gives the relative locations of genetic markers, while a physical map consists of an ordered set of segments of DNA (physical mapping being 'the process of characterizing the DNA so that it can be handled in less than genome-sized chunks' [Bishop, 1999: 6]). In this instance, the collaborators who produced the original genetic mapping resource also aimed to produce a physical map, the details of which were to be stored in database tables linked with the genetic map.⁶ During my stay in the mouse genetics laboratory this physical mapping segment of the database was undergoing development and testing. Also under development was a facility to link the database to a robot that would automatically select DNA for a given set of mice, eliminating user error from the supply process. Various additional programs allowed the selection of particular subsets of mice for detailed analysis, depending on the chromosome regions of interest for a particular study. The database was therefore designed to be, not only a record of work done, but also an integral part of ongoing research and a tool for calculating results and designing future work.

This description, of course, is not a neutral portrayal of the resource. Rather, it depends heavily on the public version of the project's rationale and substantive aims as delivered by the project leader. This individual, the laboratory leader of a well-recognized mouse genetics laboratory in London, allowed me to spend 3 months in his laboratory.⁷ My analysis is based on interviews with the laboratory leader and other members of the laboratory, including PhD students and post-doctoral researchers, who carried out mapping work related to the database and worked on mapping of a particular mouse chromosome. I also attended a demonstration of the database to potential users, and had access to documentation including the manual describing its use. Previously, I spent 3 months carrying out ethnographic research and interviews with personnel at the research-

council-funded centre that provided the computing expertise to develop the database. This resource centre also provided biological service personnel to prepare and distribute the DNA and to carry out some of the mapping work, but I did not carry out extended observation with the biological services section. The project was explicitly a European collaboration, including a well-known French genetics laboratory that was often referred to by the London-based team, but which I did not visit. Also not visited were satellite laboratories using the resource and participating in the steering committee overseeing the resource. I therefore conducted a partial ethnography that focused on two key sites in the production of the database, but did not exhaust the possibilities.

The observations occurred at the time when the basic database structure was in place, but refinements to allow for robot selection of DNA and for physical mapping were being actively worked on by the computing personnel at the resource centre in consultation with the researchers at the London laboratory. Subsequently I tracked the resource's emergence into the scientific literature and on the World Wide Web and in newsgroups. At the time of writing it has been announced that the DNA supplies necessary for further mapping are all but exhausted, and while the database is still in principle available on the web, the software is obsolete and it is no longer supported as a usable resource. This paper, therefore, claims to present ethnographic moments across the full lifespan of the creation, use and decline into dormancy of the resource. My multi-sited and varied engagements were intended to be sufficient to provide a sense of the different social, natural and digital orderings in which the resource participated, to make an assessment of how successful the resource was in achieving its goals, and to permit a discussion of the significance of the resource in relation to laboratory work and laboratory space.

The Making of a Mouse Mapping Resource

Natural Ordering

In terms of natural ordering, the mouse mapping resource can be seen to characterize a set of objects and to fix (or attempt to fix) a set of relationships between them. These relationships include links between mice and humans, between genes and markers on chromosomes, and between database entries and DNA samples. Provided these relationships hang together, the resource will render the mouse genome into a manipulable object consisting of an ordered set of genetic markers. The database system makes the mouse genome into an epistemic thing, in Rheinberger's (1998)⁸ terms. It is particularly important from the perspective of the leader of the project that this epistemic thing be seen as having relevance, not just to researchers interested in the location of mouse genes, but also to a much broader and politically prominent set of researchers concerned with understanding human genetic disease. This aspect of the natural ordering is not reflected by the structure of the database itself. It is to be achieved, rather, in the work practices into which the resource is embedded

and the forms of scientific knowledge that it is used to generate. Other aspects of natural ordering are more directly attributable to the database. The database is a relational one, structured in terms of the entities it contains, the attributes they can bear and the relationships that can exist between them. It thus defines a quite explicit natural ordering.

The basic aim of the resource is to place genetic markers on chromosomes by means of linkage analysis. Reduced to its simplest level, the database can be thought of as a matrix that lists individual mice in one dimension, and chromosome markers in the other.⁹ In relational database terms, mice and chromosome markers are entities (with attributes such as mouse number and marker name), and the matrix establishes the relationships between them (for example, mouse 3 carries marker B). Each marker is characterized as either of two variants (relating to the strains of the parent mice originally crossed in production of the offspring used to generate the DNA samples). The matrix forms the basis for the calculation of the ordering of markers along chromosomes. At this point the database depends on the interpretation of established thinking in genetics. Implementing this thinking as a computer program does, however, require a very exact statement of the rules. The database developer pointed out that while he programmes the version that he is given, sometimes assumptions turn out to be misplaced:

I can only tell them that this is the best I can come up with at the moment and complying with the rules they've told me. If they find exceptions to the rules, like first of all, they told me that errm, that they couldn't put new genes above the top anchor because they thought they'd get that right at the centromere which was the zero but they couldn't do that so the algorithm didn't account for that.

The natural ordering depicted in the database thus involves provisional fixings of objects and their relationships. In order for the database to be designed, the developer presses the scientists to tell him what can and cannot exist, what forms of relationship are possible and how they might change. This involves some changes in the way geneticists think about their work:

Interviewer: Do you think it has changed the way you look at a problem? Being involved in the database?

Geneticist: I don't know that it has actually, because I mean I guess one well, it has from the point of view I think of going back and, for example the physical mapping. I think we are much more aware of what people who are designing new databases from computers require at the outset. I think that initially when we started this whole programme, you know, the computing people had to sort of stop us and say, Hey, we need to move back a bit. But errm we, you know, I think we tended to race ahead without getting everything very clearly established.

The scientists are pushed to provide a natural ordering of a kind more explicit than they might otherwise operate with, but all those concerned consider it the role of the database simply to depict a pre-existing natural ordering.

I stated earlier that the database made the mouse genome into an epistemic thing. This is not strictly true, in that the database alone cannot achieve this. The relationship between the resource and the DNA samples is one of mutual dependency. Without the database, the samples of DNA cannot be readily deployed in ongoing mapping work. In a similar way, the database without the DNA would be a record of work that others might consult in their own ongoing mapping efforts, but it would not be a tool to be embedded into work practices in the way that its producers intend. The mutually dependent relationship is reflected in the announcements of the end of the project on the website:

The current backcross DNA samples are a limited resource, and they are now exhausted or seriously depleted for many mice. We are not planning to re-establish the DNA panel. Similarly, the [name of project] database was built using a version of Sybase which is now outdated, and which is no longer supported by that company. However, the [name of project] data can be viewed in text form ...

In this statement it is clear that without a supply of DNA samples and a working database, the maps are reduced to a set of data to be viewed rather than a tool to be deployed.

The resource therefore encapsulates a natural ordering and ties together the database and the DNA samples in a form suitable to be embedded in work practices. The publicity for the database ties together mouse and human genomics. Each relationship within the natural ordering (between mice and humans, genes and ordered chromosome markers, database and DNA samples) is, however, potentially fragile. While the mouse is indeed generally established as a model organism for the understanding of human disease (for example in hearing defects [Steel, 1995]), there are circumstances in which it is argued that other models may be more easily researched or be closer models (for example, Whitfield, 2002). While the placing of markers is achieved by a set of generally accepted and publicly available algorithms (Weir, 1996), this placement is always a statistical procedure based on assumptions that could be brought into doubt. The natural ordering also holds only in so far as the data contained in the database are considered trustworthy; that is, we accept that the scientists involved carried out their tests accurately on the sets of markers and mice that they say they did, and that their data were correctly entered into the database. We must also accept the quality of the work of the computer professional who built the database, trusting that they have accurately programmed the correct algorithms and designed the appropriate data structure.

The natural ordering is sustained not only by the resource itself, but also by the accounts of the resource given in published papers, on websites and in the words of the collaborators involved, continually attempting to fix the resource as a coherent object. For this natural ordering to be created in the first place, and to hold together subsequently, however, appropriate social orderings are also needed. These social orderings are explored in the next section.

Social Ordering

As Knorr-Cetina (1999) suggests, the development of the mouse mapping resource turns out to involve a very specific set of social orderings. The orderings that manifest in this instance are, however, rather more complex than those prevailing in the molecular biology laboratory that Knorr-Cetina observed. Two themes emerged in the structuring of that laboratory: the individual researcher with a specific project; and the laboratory as a unit headed by a laboratory leader. There was a clear distinction between the laboratory leader and the less senior members of the laboratory. The laboratory leader represented the laboratory in external contacts and was the key figure who crossed the boundary between the individual laboratory and the field as a whole. Individual scientists, from this perspective, were seen as part of the resources of the laboratory, and were expected to divide their efforts between individual goals and service to the laboratory. To a large extent this picture held true in the mouse genetics laboratory that I observed. The development of the resource described in this paper, however, added new social dynamics that cut across the existing social ordering. In particular, the tension between the pursuit of individual scientific goals and investment of effort into provision of services to the laboratory, and to outsiders, was enacted rather differently. 'Service' proves not to be a straightforward concept. In one sense, the whole resource was presented as a service to the scientific community, and, as will be discussed later, involvement in this kind of work did create problems for some members of the laboratory as Knorr-Cetina suggests. However, during production of the resource, an additional level of separation between 'service' and 'scientific' activities was enacted. This helped to alleviate tensions by outsourcing the most obviously service-oriented work: database development and routine supply of DNA samples.

The development of the database was carried out not in the laboratory, but by computing specialists working for a resource centre funded by the UK's Medical Research Council. The centre was set up to provide both biological and computing facilities to the UK genome mapping community (Balmer, 1995; Glasner et al., 1998).¹⁰ Continued funding for the centre was seen to depend on being visibly useful for genome mapping and being seen to provide resources that the mapping community valued. As an ethnographer within the centre, I saw how the computing group felt under pressure to demonstrate that their systems were being used. They did so by storing and displaying usage statistics for their systems, writing reports, actively promoting use of services and recruiting allies. The pressure to demonstrate usefulness organized many of the centre's priorities. Involvement in the mouse mapping resource was seen as a good thing, because it provided a high-profile service and was an example of a product developed by the centre, rather than by a third party. Because the centre was an explicitly service-oriented organization, staff here did not experience pressures to gain recognition for input to scientific work or to make visible scientific progress. This division of labour between service providers and

scientific staff largely resolved any tension between service provision and scientific reward. The maintenance of the social boundary between the groups did, however, bring its own concerns, focused on communications between the groups, and particularly on how this boundary was used to allocate responsibility for problem identification and resolution.

The database developer, initially trained as a mathematician, readily gave descriptions of his role in the project, the role of databases, and the way in which scientists operate:

[I]n terms of database design they don't need to know, they just need to know how to use it, and ... part of the objective of a program is to help the, to aid scientists to derive the results of their data or from their data.

The scientists make the discovery, we provide the service, I provide the tool. OK, that's my job, to provide a tool. My job is not to understand what the discovery will be.

This description places a very clear boundary between computer service specialists and scientists, and allocates responsibilities on either side of that boundary. This allocation was also invoked by the head of the genetics laboratory:

Basically he was put on the project in the early days and I mean I guess it was me sitting a lot of the time with him saying 'look', I mean he didn't know anything about genetics when he started but we basically looked at the basic problems – he learned what we were trying to do in genetic terms – I don't think I learned so much about computers!

This division of labour is conventional in development of information systems. The database developer is responsible for identifying 'user requirements', and is expected to get to know the users and find out what their needs are. The database developer, however, learns more than the language of genetics and the application of linkage mapping algorithms. He also learns to account for features of the database in terms of narratives of how scientific work proceeds:

So what happens if there's an experimental problem with that particular scoring type they can go to that particular mouse and say OK we'll check that out again and check the sample and that's the reason for that really.

Some of these understandings extend beyond issues of scientific practice to a more cultural understanding of scientific work:

Eventually the locus and probe forms ... will have additional information as to who, err who is responsible for deriving the data ... so lab XYZ, Mr X derives this data and he has got a reputation for having dodgy data, so if you see his name there he's going to get embarrassed sooner or later that's how scientists work, your reputation is kind of tarnished if you put wrong results.

Two different worlds are maintained, but in order for meaningful work to go on the two sides have to understand one another, not just in a technical

sense but also in a cultural sense. The balance of need to understand is, however, one-sided. The scientists never gave me the kind of rich narratives of what work in computing was like that the computing services personnel gave me about scientists. There is more to be said about why the database developer needed so much cultural understanding of scientists, and this is discussed below in the section on digital orderings. For present purposes, it is enough to note the persistence of a boundary between the two groups, and the different perceptions of the project that prevailed on either side.

The distinction between the worlds of computer service specialists and geneticists included different perceptions of the scale of problems. During the course of an interview with the database developer, two occasions were described on which the interpretation of the scale of a problem varied dramatically.

You see that, the red bits are the anchors, and the black bits are the new gene that they've mapped onto there, and what she wanted to know is can she actually have a condensed panel in which all these missing values are removed as compared to the others, so you get a condensed thing. This is what will appear, but she wanted me to kind of do something different with the data, so I said, oh well, that's pretty trivial, so I've already err, I've already actually, yeah, I've got it here.

Errm, there's extra bits such as the list of mice, that was only because some scientists said ah, we've been having difficulty scanning along here and looking at the mice to work out, can you just not give me the list of mice which are recombinants of this type and that means kind of, you know, two three weeks work to me, but to him it's a trivial thing.

In both instances it is notable that the database developer presents himself as willing to provide a service that meets user needs. The distance between the two groups is, however, sketched out dramatically in the distinction between tasks that are trivial and those that entail substantial work. Database adjustments that would, to the scientists, fit better with their work practices, may either require rewriting the entire database or simply altering a few parameters. The database developer's frustration was with conveying the difference between the two cases to scientists who do not, in turn, understand the practical constraints on his work.

The database developer also has a sense of where responsibility lies for the ultimate reliability of the resource. He is confident that the programming of the database is as correct as he can make it. Any problems are more likely to be concerned with the inapplicability of the rules he implemented. He admits that the database might contain programming errors, but he shows me sheaves of test printouts to demonstrate the effort he put into minimizing them. He expresses a duty to act upon what he feels are vaguely expressed wishes, and draws upon his own picture of how scientists operate, to translate those wishes into database features, algorithms and buttons to click on. The scientist, in his view, has responsibility for ensuring the accuracy of the data that they enter and for carrying out reasonable tasks with the database:

So what they will do now is, what they do is they type, she's only typed 48 mice, right, out of a thousand, which is a small sample, and they're saying use statistics to work out linkage, yeah, saves on the work basically, cos each takes about a day to do for each mice, or set of mice, so they'll say OK I've done 48, I'll just bung it in the database and get some data out, and the database gives you a prediction of that on the data that's supplied, and if that's incorrect that's her fault.

The approach to correctness of the database that he describes suggests that responsibilities are allocated on either side of a social boundary: while it is his job to see that the program is technically correct, it is outside his remit to tell scientists what to do. It also suggests a moral order in which *technical* problems must be solved, whereas *human* problems, particularly those caused by a different social group, are simply accepted as an inescapable, normal feature of the work.

When taken together, the two interviews portray a boundary between two social groups that are capable of achieving a certain amount of mutual understanding. No doubt this is partly based, as Galison (1997b) suggests, on a certain amount of stereotyping of self and other. However, the two groups also represent themselves as having learned from each other and having come to understand each other's needs, and both are part of the project team as a whole. Neither side portrays the relationship as being problematic, or as having implications for the database or the research. They largely treat the work practices of the other group as a given; an occasional source of difficulty or frustration that simply has to be endured. Each appears to be comfortable that appropriate rewards stem from their activities. To this extent, then, the problems identified in the Wellcome Trust report (2003) discussed earlier appear not to arise within this social ordering.

The geneticists in the laboratory are dependent on the expertise of the computing staff at the resource centre for the development of the database. Indeed, the realization of the resource depends on the interaction between groups, since neither can be said to have had a clear idea, prior to the project, of what the resource was or what it would do. This parallels the development of many forms of scientific instrument besides databases. Instrument-making has long been a skilled pursuit that requires scientists to work alongside other specialists (Schaffer, 1992). In this case, the fact that the skilled computer professionals work in a different institutional context can simplify matters, because they are not concerned with gaining scientific recognition. The computing staff has a different set of organizational allegiances and pressures to consider, in which continued funding stems from demonstrating usefulness in the right degree to the right audiences.

A further service, the provision of DNA samples, was necessary for the database to be fully integrated into a mapping resource. This service was also provided by the resource centre. A largely separate part of the organization performed much of the routine work to produce the skeleton map, and supplied the DNA samples to users of the resource. After having

spent so much time with the computing services side of the organization, I realized that I had largely adopted their view of the organization and treated the biological services section as a distant place (although only one floor up in the same building!). Therefore, I cannot give an equivalent account of the biological services point of view. The biological services section was staffed by biologists, who were more likely to have their own scientific aspirations and to share a conception of problems with the genetics researchers. Still, the matter-of-fact way that their involvement as service providers was treated by both computer scientists and laboratory geneticists implies a stable social ordering. This arrangement sustained them as a separate autonomous group and protected the laboratory geneticists from routine service tasks.

Thus far I have used the laboratory leader to stand in for the concerns of the geneticists in the development of the resource. It is worth noting, however, that one feature of the social ordering of the laboratory identified by Knorr-Cetina continues to prevail. That is the tension between the role of laboratory leader as the public face of the laboratory, and the individual scientists who have their own projects and goals for career advancement. This tension was manifest in the feelings of some laboratory members towards the way the database would make the data they had produced available to collaborators. For PhD students, and to some extent post-doctoral researchers, having data of one's own is vitally important. All the researchers in the laboratory that I spoke to, from research assistants to senior post-doctoral researchers, were able to discuss their individual projects with a sense of ownership, whilst also seeing themselves as working in teams towards larger goals. The common goals were realized in most instances through personal ambitions to gather sufficient data to write a thesis and gain a doctorate, or to get one's name on a publication.

For PhD students, the need to protect their data was a priority and the database was a potential threat. For example, one student worried that contributing data to the database would mean losing control over the data, leaving it open for someone else to use or delete accidentally. Much of the demonstration session in which the database developer explained the use of the resource was taken up with concerns about who could see and alter data, and how to set appropriate controls. For this PhD student, the resource was a potential threat to personal ambitions for gaining recognition. While the laboratory leader favoured inputting the data for the benefit of the resource and the collaboration, the PhD student saw little gain and much risk. The timescale of PhD study¹¹ meant that the student was unlikely to benefit directly from future uses of the resource to facilitate mapping. In this case, the PhD student delayed entry of data into the database as long as possible, and deployed all of the available options to restrict access to the data once it had been entered. Thus, whilst the separation of database development and DNA provision from laboratory work resolved some tensions, some still remained to be dealt with at the level of day-to-day work in the laboratory.

A final aspect of the social ordering was the explicit presentation of the project as a collaboration between laboratories. The project was initiated as a collective effort in which different laboratories and institutions, rather than competing, would pool their efforts. They would be able to claim ownership of a resource, which they hoped would become widely used and known, to further their own research goals by using the database to understand particular genes and diseases and to produce collaborative publications reporting on their mapping efforts. Laboratories that in many circumstances would see themselves as competitors in this instance found themselves contributing to a joint effort. The existence of the collaboration is partly attributable to funding regimes, and the availability of funding for explicitly European endeavours. To a large extent the collaboration was carried out by means of the database: contributing laboratories divided up the work between them, performed it separately and then placed the resulting data in the database. This is to neglect, however, the considerable extent to which designing the project, and designing the database as part of it, were collaborative endeavours. The project was collaborative, but for the most part the actual work was solitary. This proved to be an effective means for organizing the work, with tensions largely being experienced at the level of individuals, as discussed above, and not permitted to threaten the success of the collaboration as a whole. As the next section discusses, safeguards within the database were designed to protect the confidentiality of particularly sensitive data and allow laboratories to maintain their competitive advantage in specified areas of the genome.

In summary, the creation of the resource entailed the enactment of a variety of different forms of social ordering, within the laboratory, between the laboratory and the resource centre providing services, and between the laboratory and its key competitors/collaborators. Most significantly, the mouse mapping resource could not have been produced without the combined efforts and expertise of mouse geneticists and computer experts. Instruments, as Golinski says, link 'individual settings to widely distributed networks of production and circulation of artefacts' (Golinski, 1998: 138). Bringing together the understanding of the two groups, mouse geneticists and computer experts, was essential for the resource to tie together the material DNA with information on the ordering of markers. The two groups worked together in order to understand one another's concerns in a way that allowed each to meet its own needs. This is similar to the 'trading zones' (Galison, 1997a) that allow coordination of work between different subcultures in physics, in the construction of instruments and the conduct of experiments. It is important to remember, however, that neither geneticists nor computer experts are homogeneous groups. Tensions, differing perspectives and divisions of labour existed within the groups, most notably between the laboratory leader's desire for high-level collaboration and the PhD students' needs for distinctive contributions. The notion of the trading zone should not be taken to imply a coming together of homogeneous groups. Rather, it comprises a zone of common concern that brings together many and varied interests, including different professional

groups and different hierarchical levels. Databases prove not to be particularly distinctive in this respect: any project that enrolled enough interest could have brought together different groups in this way.

In this project there was a strong division between scientific and service functions, and a similarly strong division between computing and biological services. Each took some reward from collaborating, different in each instance and based on their specific understandings of the policy and funding climate and of the importance of visibility in their relevant field of operation. In carrying out this collaborative effort much mutual learning occurred, but without a breaking down of professional or institutional boundaries. Within this over-arching social ordering, the coexistence within the project of different frames of relevance led to varying sets of tensions and anxieties, maps of responsibilities, and understandings of problems and their solutions. The social orderings described here, however, only go a certain way towards furnishing the natural orderings described in the previous section. The structure of the database itself tied together the natural and social orderings in the digital domain.

Digital Ordering

Manovich (2001), as mentioned earlier, claims that the computer projects are an ontology onto the world. We can now ask whether, and to what extent, this idea applies to the creation of this database. Was it experienced as enforcing or suggesting particular ways of seeing the world? Did it produce a specifically digital form of ordering that was projected onto, and over-rode, other forms of ordering and ways of seeing? It turns out that this is an over-simplified way of viewing the process of database construction. Rather than an over-riding digital ordering, what we see is an incorporation of natural and social orderings into the database. Where there is an appeal to digital ordering, it tends to be treated as a resource for enforcing or encouraging effects that are desired for natural or social ordering; for introducing quality controls on data or producing appropriate relations between data users and contributors, and the team producing the resource.

It was argued earlier that the database developer was required to develop a cultural understanding of how scientists work, as well as a technical understanding of genetic linkage analysis. An important aspect of the database design was to incorporate appropriate safeguards to ensure that those contributing data would feel that they were being adequately protected against misuse of their data and were receiving credit for their work. At the same time, safeguards needed to be put in place to increase trust among data users. As with the technical understanding of linkage analysis, the geneticists did not simply bring an existing world-view to the database developer and ask for it to be incorporated into the resource. Rather, the necessary components were worked out in discussion, through trial and error. Requirements were mapped out in advance as far as possible, and then refined through trial sessions and demonstrations of the

database to potential users. One such occasion was the training session I attended at the computer services centre with a PhD student and post-doctoral researcher from the genetics laboratory. They were to use the finished database, and were ostensibly present for training. The session, however, rapidly became a discussion of the desirability and necessity of particular features, not just for representing data, but also for safeguarding access and attributing credit.

The database represented not only the mouse genome as a manipulable object, but also a culture of trust (and mistrust), reward and error: in Collins' (1998) terms the database portrayed a particular evidential culture. This included the provision of different levels of data access for different users. The default position was that one should not necessarily have access to all the data. Rather, reasons had to be demonstrated for needing access to different regions of the mouse genome, and while in some cases a researcher might be permitted to alter data, in other instances they might only merit permission to read the data. Some data might be so sensitive that one might wish to keep them confidential. Although the whole rationale for the database was to create a resource through data pooling, researchers recognized that when the data were the product of a race to characterize a particular gene, and when another laboratory was also close to a solution, they might wish to enter these data in the database but hide them from others.

While data were pooled for purposes of map order calculation, the database also preserved information on their origins. At the heart of the database was a belief in scientific reputation and ability attached to the individual and not to teams or laboratories. Individual scientists would both take credit for good work and be blamed for faulty work. The programs that calculated linkage and map positions were, as the database developer stated, as correct as he could make them. They also depended on accepted algorithms. The results might, however, be erroneous for a variety of reasons. The original typing of mice could be at fault, due to human error, or use of the wrong DNA sample. In addition, linkage-mapping is a statistical process, and there is no guarantee that the results calculated, even with flawless data, will represent the actual ordering of markers on chromosomes. For these various reasons, all deductions about the ordering of markers required the approval of accredited scientists before being added to the database.

In the case of approving mapping results, the solution was to introduce a human step to over-ride the automated result produced by the application of the algorithm. The scientist was required to check the result and make sure it made sense. A certain amount of automated error checking was, however, also incorporated into the database. This required that any data input must be biologically sensible and devoid of obvious data entry errors (for example, only allowing entries of a specified format or within a pre-defined range). Human error was also the target of the whole project to automate DNA selection. Humans, it was believed, could pick the wrong DNA sample and threaten the integrity of the typing data used for

mapping. Unlike some data-sharing arrangements, in this resource there was no requirement that data be peer reviewed or formally published before being put into the database. The probability thresholds used for calculating map position were expected to tolerate a few wrongly typed mice, and wildly wrong results would be screened out during the process of mapping, but there were no guarantees.

The digital ordering was, then, designed to minimize error, but whether this entailed introducing human judgement, eliminating human involvement or using probability thresholds to tolerate error was a situated choice. The aim of such judgements on the social character of scientific research was to preserve the integrity of the natural ordering incorporated into the database. A final component of the database was intended to ensure that data arrived in the first place; that is, it was designed to motivate scientists to contribute. While the original collaborators had allocations of skeleton markers to map, subsequent users of the resource did not have the same incentives to lay claim to their own data. The credit accruing from seeing their names in the database might not, it was feared, overcome paranoia about premature exposure of personal results or laziness in learning to use the database. Incentive to contribute data was instead subtly woven into the procedure for using the resource. DNA samples would be routinely supplied to users without the necessary codes to make sense of the data. Only when users returned the new information on typing of markers that they had generated would they be given the codes to enable mapping of those markers. The database and the DNA samples were linked together in the work practices of building and using the resource in such a way as to maximize data submission. This solution had the potential to be seen as a fair exchange, with users contributing their data in order to receive the results that they needed for their own research projects. This seems to have been a less controversial mechanism for encouraging data-sharing than the one described by Beaulieu (2003), where contribution of data was required as a prerequisite for publishing research in a particular journal.

The 'digital ordering' represented by the database is thus highly contingent, representing the upshot of lengthy negotiations between the collaborators over the nature of the natural objects involved and of the scientific culture in the workplace. There does not appear to be a dominant computer culture reshaping all that it touches. To the extent that there is a digital ordering, it is in the use of the database to fix elements of natural and social ordering, with the aim of stabilizing the resource. Just as the laboratory, for Knorr-Cetina (1999), was the site at which specific orderings of the natural and social world were assembled, so also does the database provide a focus for the assembly of natural and social orderings. Not every database will be used in this way, just as every building with doors, windows and even fume cupboards does not function as a science laboratory. Even the robustness of digital orderings is open to some question. Access to the database is controlled by password, and it is via this password that an individual scientist is identified to the system. The rules

on access to data, confidentiality, credit and blame operate at this level of the identified individual. Within the world of the database-developer it was self-evident that people would want and should use their own passwords to access computer systems. Within the laboratory, this understanding was not so clear-cut. Sharing of passwords and multiple use of accounts was rife, undermining the assumption that each user identity corresponded to a single scientist. While the version of scientific culture embedded in the database seemed rigid, it was still open to undermining when the assumptions of disciplined computer use were not adhered to.

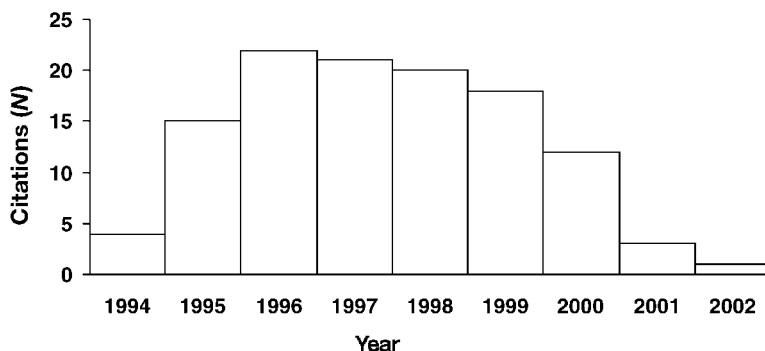
A Successful Resource?

The resource could be considered successful in so far as the natural, social and digital orderings that comprise it hold together. To the extent that the resource did form a basis for publications, it achieved what it was intended to do. A publication in *Genome Research* proudly announced that the collaboration had achieved its goal of a high-resolution genetic map for the mouse. Still, however, this only tells us that the group achieved success in its own eyes and those of a limited set of peer reviewers. One way of assessing the wider success of the resource is to track its place in the literature. A simple citation analysis is a very partial way to assess the visibility of the resource to outsiders; still, a totally unsuccessful resource would be unlikely to gain any significant visibility in the literature.

The most cited publication describing the resource is a 1994 publication in *Human Molecular Genetics*, which sets out the aims and method of construction. This paper is suited for citation in the methods sections of papers making use of the resource. Figure 1 shows the distribution of citations to this publication, using data from ISI Web of Science. The 1994 publication had received 118 citations by the end of 2002. As can be seen from the chart, citations peaked in 1996 and then slowly declined. The actual use of the resource within laboratories takes a somewhat different course, but we can see from the chart that the resource did acquire

FIGURE 1

Citations in the published journal literature of the 1994 publication describing the mouse mapping resource (source: Institute for Scientific Information Web of Science).



visibility in the literature. Crucially for the rhetorical success of the project, the citations include uses of the resource that link mouse genetics to human disease genetics, and that include groups not involved in the initial collaboration. To this extent at least, the database was a success, as its initial natural, social and digital orderings were extended to new contexts.

Beyond raw citation data, a further indicator of success is to note the form in which the resource is cited in the methods sections of papers. It tends to be cited without qualification, justification or extended discussion, as an aid to mapping. These factual citations without modality (Latour & Woolgar, 1986: 77) indicate the successful accomplishment of the resource as an aid to mapping. In this sense, the resource is treated as a scientific instrument that gives transparent access to the facts of mouse genome mapping. The circumstances of its production and the sources of the data are treated as irrelevant when the resource travels to new contexts. The resource also gained considerable visibility on the World Wide Web in guides to genome computing and lists of mapping resources. These links largely reproduced the descriptions given by the producers of the resource without comment or qualification.

In its own terms, then, the resource can be considered a success. In the descriptions of its producers, in its appearances in the academic literature and by becoming an acknowledged entity on the World Wide Web, it met its goals and achieved a level of unproblematic acceptance. It may never have been hailed as revolutionary, and it was a relatively limited resource in the overall project of human genome mapping, but it effectively achieved what it set out to do, without attracting undue attention. We might, therefore, conclude that the social, natural and digital orderings held together in achieving a robust resource that made its way into the relevant discipline. That it did so provides testimony to the fact that building a resource of this kind need not involve radical adjustments to standard practices. The prior existence of a group whose funding situation predisposed them to take responsibility for service functions in the project eased the situation considerably.

Conclusion

The database described in this paper operated as both a scientific instrument and a means of communicating results. It was a highly contingent and contextual achievement, produced through collaboration among groups of biologists and computer scientists. The database was enacted differently in different circumstances, sometimes figuring in discussion as an object in its own right and sometimes only making sense as part of a complex of material artefacts and work practices.¹² Production and use of the database were so closely tied to particular circumstances and specific social networks that generalizing about effects is problematic. While it is interesting to speculate about the transformation of science through the use of new information and communication technologies, the case study in

this paper suggests that it would be misguided to expect to find evenly distributed and homogeneous effects. New technologies, such as databases, may provide occasions for developing new work practices in science, and may lead to the exploration of new areas of knowledge, but consequences do not flow directly or predictably from the technology. In the present case, the database was deployed not as a radical transformation of scientific practice, but as a relatively small-scale intervention in work practices. It did not give rise to a distinctive epistemic culture, but was an additional ordering resource within existing research cultures.

In particular, the resource described in this paper was as much a research tool or scientific instrument as it was a medium of publication. Researchers were concerned that placing their data on the database would cede control over its use. To this extent, the database was seen as a communication medium. However, by rolling up data from the skeleton mapping exercise with the DNA samples and algorithms necessary to generate new mapping data, it became a research tool that was supposed to be deployed in ongoing mapping projects. The database acted as a resource to specify, emphasize, and in some cases enforce particular social and natural orderings. Like spatial orderings, digital orderings are flexible resources: people constitute them, interpret them and make them meaningful, rather than simply being subject to them.

The resource described here was recognized and applied both within and beyond the mouse genetics community. It was, then, a transportable technology for doing science. It was not, however, a broadly generalizable technology such as those described by Joerges & Shinn (2000, 2001) and Shinn & Joerges (2002). Their 'research technologies' were far more generalizable in their use, and applicable across a broad range of disciplines and forms of scientific work. Similarly, the standardized packages described by Fujimura (1987, 1988, 1996) are far broader in the disciplinary contexts into which they are inserted and the potential range of problems to which they are applied. Paradoxically, by presenting the resource described in this paper as a product with circumscribed uses, the collaborators may have both promoted its initial success by making sure it would fit current work practices, and ensured that it would be short-lived. Mapping technologies and areas of interest in mouse genomics moved on, as did software versions, and it proved not to be a priority for anyone to update the software or extend the availability of the DNA samples. For everyone concerned, efforts to develop the resource needed to be repaid through visible use in the lead collaborators' laboratories and elsewhere. When this happened less often, the resource lapsed into dormancy.

Thinking about the generalizability of the resource casts a new light on the trend towards recycling of research data identified by Schröder (2003): a thrifty science system is enabled by databases to get more 'wear' out of its data. In the mouse mapping resource, subsequent uses were carefully circumscribed by the design of the resource, such that a package of natural, social and digital orderings were presented to users. Getting more 'wear' out of these data entails passing on a perspective about work to be done

with the data and how to do it. Recycling of mapping data is also promoted by the highly standardized conditions of their production and interpretation: linkage mapping is already a well-established procedure and it is accepted practice to use published mapping data. Ideas about the reusability of data differ between contexts. In the case described by Beaulieu (2003), the re-use of data proved more problematic, since conventions were less established and researchers still felt it was vital to understand the specific experimental conditions that generated the data. By contrast, large-scale DNA and protein databases occasion many more unforeseen and alternative uses. Like linkage mapping data, sequence data are thought to be reusable without the necessity to know the specific circumstances of their production. A large database may, however, offer up more possibility for users to integrate them into diverse projects, because of their more open approach to specifying contexts of use (Fujimura & Fortun, 1996). On their own, then, databases do not promote data re-use: the way that the package is presented and researchers' expectations about the interpretation of data influence both whether and how data can be taken up and incorporated into new projects.

I began this paper by posing the question of whether the application of information and communication technologies changes scientific knowledge production. The example of the mouse mapping resource suggests that each instance of database development is likely to bring together unique sets of natural, social and digital orderings. In this project no prevailing digital logic over-rides other ways of understanding project goals and social situations, although the decision to see a project in terms of data structures and algorithms does play a part in shaping the visions of possibility and necessity: the 'shadow of the format' (Nentwich, 2003). The digital perspective, therefore, may shape goals to a certain extent, but it does not determine outcomes. Consequently, while practices and outcomes of knowledge production may change with increasing use of information and communication technologies, such changes do not do away with existing frameworks or necessarily produce new epistemic cultures.

It may seem that the significance of the laboratory would diminish as work practices increasingly dealt with distributed data, and as scientific phenomena more often took the form of digital representations rather than wet material artefacts. It can be seen from the example of the mouse mapping resource that the relationship between mice and their digital representations was more complex than the straightforward replacement of the one with the other. It can also be seen that the laboratory and its boundaries remain a highly relevant source of ordering, albeit coexistent with other sources of ordering. In Lynch's (1991) terms, the laboratory remains a place where different topical contextures are enacted, but it is cross-cut by the topical contexture of the database. Lynch describes in ethnomethodological terms how both spaces and instruments (like the mouse mapping tool) can act as fields that configure what counts as work and its site:

Scientific laboratories, observatories, linear accelerators, mainframe computers, and other equipmental complexes can be treated similarly as matrices for human conduct that do not simply provide places where human beings work but instead provide distinctive phenomenal fields in which organizations of 'work' are established and exhibited. (Lynch, 1993: 132)

The database and the laboratory can therefore co-exist as different frameworks for organizing action, without one necessarily threatening the other.

The laboratory is in no danger of disappearing as a key location as long as it forms the organizational nexus for a strong laboratory leader and a team of goal-oriented researchers, as in this instance. The networks of connection, collaboration and communication in which the laboratory is embedded at all levels of the work are, however, more complex than those that Knorr-Cetina (1999) ascribed to the molecular biology laboratories she studied. She described individuated laboratories, and a laboratory leader acting as a bridge between the work of the laboratory and the outside world; but the creation and use of the mouse mapping resource involved far more, and far more complex, communications with the outside world across all levels of the laboratory hierarchy: for PhD students, post-docs and the laboratory leader. All of these researchers are involved, as Hilgartner & Brandt-Rauf (1994) describe, in an ever-present dilemma fraught with contradictory pressures over which data, materials, instruments and publications to make available beyond the laboratory. The result is a wide diversity of forms of communication media and research tools connecting researchers in a laboratory together with others across geographic locations. The patterns of connection and collaboration in scientific knowledge production involving databases can thus become both spatially and socially complex, building on existing networks but adding additional density, bandwidth and new tensions.

The different forms of connection produce as their cumulative effect highly complex and dynamic networks across time and space. Further research on such patterns of scientific work will require creative methodological approaches: a straightforward laboratory ethnography is unlikely to be sufficient for revealing the intricate patterns of connection that contemporary science can entail. In this paper I focused on the development of one resource. Obviously there are many different options for framing an ethnographic approach, especially to explore evolving epistemic cultures in depth. The words of Czarniawska on the study of organizations are relevant here:

Modern organizing ... takes place in a net of fragmented, multiple contexts, through multitudes of kaleidoscopic movements. Organizing happens in many places at once, and organizers move around quickly and frequently. (Czarniawska, 2004: 786)

For 'organizing' we could as easily read 'scientific research'. The work of science goes on in many places at once and moves faster than an ethnographer can keep up with. Novel ethnographic strategies for pursuing both

the locations of scientific work and the connections between them are likely to be increasingly important in future. The ethnography of contemporary science can still usefully focus on 'laboratory life', but needs to be able to take other forms of ordering, including digital ordering, into account.

Notes

I would like to thank Michael Lynch, Malcolm Ashmore and the anonymous reviewers for their concerted efforts to make me say what I meant and mean what I said. Much gratitude is owed to the two institutions that hosted me, and all the people who agreed to be interviewed and shadowed. This research was begun as a project funded by the UK's Economic and Social Research Council (ESRC) at Brunel University under the invaluable guidance of Steve Woolgar (award no. R000290001). Subsequent writing has been funded by another ESRC grant (award no. R000271262).

1. Lenoir himself is much more measured on this issue in other writings, as discussed later in the introduction.
2. There are, of course, some problems in posing the question in this way, as a straightforward comparison of then with now. To do so implies a monolithic molecular biology epistemic culture, as if the only distinctions between the situation Knorr-Cetina studied and the one described in this paper were time and a new technology. This is to transfer the mistake of thinking that there is a single scientific culture down to the level of the discipline. Rather than a controlled comparison, then, it is probably more useful to think of Knorr-Cetina's observations as an interrogative device. Her observations of one molecular biology culture provide a source of questions to ask of the case study described here.
3. Linkage mapping is described in a textbook on genetic databases thus:

Linkage analysis requires two or more variants of each of two or more observable characteristics of an organism. Some characteristics are inherited together by more than the chance amount (50:50), while others may show varying degrees of co-inheritance. By suitably designed breeding experiments, the degree of this linkage of characteristics may be measured. . . . The characters studied in linkage analysis need not be morphological. In recent years detailed linkage maps of both human and mouse have been built up using a variety of highly polymorphic markers. The human map has limited resolution for practical reasons that include the generation time and ethical considerations. The mouse map can be refined to a resolution of a few hundred kilobases (kb), which is enough to connect a phenotype to a single gene in many cases. It thus provides a remarkably powerful resource for mammalian biology. (Bishop, 1999: 5)

We can see clearly here the positioning of mouse mapping as a tool to understand the human genome, rather than as an end in itself.

4. Kohler (1994) discusses the development of genetic mapping using linkage in *Drosophila*, and explores its deployment both as a tool in its own right and as an aid to physical mapping.
5. Weir (1996) discusses algorithms for calculating linkage, and points out that while these may be used to estimate the distance in terms of the number of DNA base pairs between markers, this calculation is based on assumptions that may not hold true. In particular, the assumption that recombination is equally likely throughout the genome may not be valid.
6. Primrose & Twyman (2003: 47) describe the connection between genetic and physical maps as follows:

Today physical maps have assumed an importance that far exceeds their utility in facilitating whole genome sequencing. Perhaps the easiest way of understanding their importance is by analogy. Suppose we decide to drive

the 3500 miles from Los Angeles to New York. If the only map we have says that we will have to drive through Albuquerque and St Louis, then we have very limited information. Such a map is analogous to a genetic map of a chromosome: the mapped genes are not close enough to be of much utility. On the other hand, suppose that every mile between Los Angeles and New York there is a milepost giving the exact distance to both places. In this situation we will know exactly where we are at any time but the information is of no other interest. This is analogous to knowing the exact base sequence of a chromosome or genome. What we really want for our journey is a map that has different features of interest (towns, rivers, national parks, museums) evenly spread along our route that will enable us to relate observations to locations and position to geographical features. A physical map of a chromosome is exactly the same in that different kinds of markers are identified by physical means and mapped along the chromosome. Genetic markers can be located with reference to the physical markers and vice versa. Although the exact DNA sequence of the chromosome enables the exact location of physical features to be determined, the physical map is more useful for day to day use.

7. I have reached the conclusion that it is not necessary to identify the particular laboratory discussed here, nor the individuals concerned. Insiders will no doubt already know, and a small amount of Internet searching would probably provide the answer for a sufficiently interested outsider. That those concerned could be identified does not seem particularly threatening to their careers or peace of mind since nothing described here is particularly controversial among scientists, and everyone spoke on the record. At the same time, I am reluctant to give names where it is not absolutely necessary so to do, and even identifying the resource by name would inevitably lead to the individuals concerned. As in a previous instance (Hine, 2002), I leave the judgement to the reader whether to dig deeper in identifying the research site, and whether this makes a difference to the status of the analysis.
8. The mouse mapping resource does not, however, fully meet Rheinberger's (1998) notion of an experimental system, since he sees experimental systems as necessarily involving 'differential reproduction' or the capacity to generate surprises. The mouse mapping resource was intended to accumulate results in a more predictable fashion. In this sense, the mouse mapping resource was only a part of other experimental systems looking for disease genes that would have the capacity to find the unexpected.
9. This matrix is, from another perspective, not a starting point but the endpoint of a sequence of transformations of representations. The process of 'typing' mice, or identifying which markers each mouse carries, involves processes of selectively digesting DNA, performing electrophoresis to separate out DNA fragments, probing with known markers and interpreting the results. The most common forms of representation of this process found in the laboratory were first, photographs of the electrophoretic gels and second, diagrams drawn up on the word processor depicting mouse typings by black and white squares for the two different parental types. The database stood in for these prior representations.
10. There are some similarities with the European Bioinformatics Institute, as described by Wouters & Reddy (2003).
11. In the UK, a PhD student will typically be funded for 3 years.
12. As evidenced by a certain slippage, both in the descriptions of participants and in this paper, between discussions of 'the database', 'the resource' and 'the project'.

References

- Balmer, Brian (1995) 'Transitional Science and the Human Genome Mapping Project Resource Centre', *The Genetic Engineer and Biotechnologist* 15(2/3): 89–98.

- Beaulieu, Anne (2001) 'Voxels in the Brain: Neuroscience, Informatics and Changing Notions of Objectivity', *Social Studies of Science* 31(5): 635–80.
- Beaulieu, Anne (2003) 'Research Woes and New Data Flows: a Case Study of Data Sharing at the fMRI Data Centre, Dartmouth College, USA', in Paul Wouters & Peter Schröder (eds), *Promise and Practice in Data Sharing* (Amsterdam: NIWI-KNAW): 65–88 < www.niwi.knaw.nl/en/nerdi2/group_members/paul/publ/datasharing/toonplaatje > .
- Bijker, Wiebe E. (1997) *Of Bicycles, Bakelite and Bulbs: Towards a Theory of Socio-Technical Change* (Cambridge, MA: MIT Press).
- Bishop, Martin J. (1999) 'Introduction', in Martin J. Bishop (ed.), *Genetics Databases* (London: Academic Press): 1–9.
- Bowker, Geoffrey (2000) 'Biodiversity Datadiversity', *Social Studies of Science* 30(5): 643–84.
- Bowker, Geoffrey C. & Susan L. Star (1999) *Sorting Things Out: Classification and its Consequences* (Cambridge, MA: MIT Press).
- Brown, Cecelia M. (2003) 'The Changing Face of Scientific Discourse: Analysis of Genomic and Proteomic Database Usage and Acceptance', *Journal of the American Society for Information Science and Technology* 54(10): 926–38.
- Collins, H.M. (1998) 'The Meaning of Data: Open and Closed Evidential Cultures in the Search for Gravitational Waves', *American Journal of Sociology* 104(2): 293–338.
- Czarniawska, Barbara (2004) 'On Time Space, and Action Nets', *Organization* 11(6): 773–91.
- Doyle, Richard M. (1998) 'Emergent Power: Vitality and Theology in Artificial Life', in Timothy Lenoir (ed.), *Inscribing Science: Scientific Texts and the Materiality of Communication* (Stanford, CA: Stanford University Press): 304–27.
- Fujimura, Joan H. (1987) 'Constructing Do-Able Problems in Cancer-Research: Articulating Alignment', *Social Studies of Science* 17(2): 257–93.
- Fujimura, Joan H. (1988) 'The Molecular Biological Bandwagon in Cancer-Research: Where Social Worlds Meet', *Social Problems* 35(3): 261–83.
- Fujimura, Joan H. (1996) *Crafting Science: A Sociohistory of the Quest for the Genetics of Cancer* (Cambridge, MA: Harvard University Press).
- Fujimura, Joan H. & Mike Fortun (1996) 'Constructing Knowledge across Social Worlds: The Case of DNA Sequence Databases in Molecular Biology', in Laura Nader (ed.), *Naked Science: Anthropological Inquiry into Boundaries, Power and Knowledge* (New York: Routledge): 160–73.
- Galison, Peter (1997a) *Image and Logic: Material Culture of Microphysics* (Chicago, IL: University of Chicago Press).
- Galison, Peter (1997b) 'Three Laboratories', *Social Research* 64(3): 1127–55.
- Galison, Peter (1999) 'Buildings and the Subject of Science', in Peter Galison & Emily Thompson (eds), *The Architecture of Science* (Cambridge, MA: MIT Press): 1–15.
- Glasner, Peter, Harry Rothman & Wan Ching Lee (1998) 'The UK Human Genome Mapping Resource Centre: A User Analysis', in Peter Wheale, Rene von Shomberg & Peter Glasner (eds), *Social Management of Genetic Engineering* (Aldershot, Hants.: Gower): 63–75.
- Golinski, Jan (1998) *Making Natural Knowledge: Constructivism and the History of Science* (Cambridge: Cambridge University Press).
- Hilgartner, Stephen (1995) 'Biomolecular Databases: New Communication Regimes for Biology', *Science Communication* 17(2): 240–63.
- Hilgartner, Stephen & Sherry I. Brandt-Rauf (1994) 'Data Access, Ownership, and Control: Toward Empirical – Studies of Access Practices', *Knowledge* 15(4): 355–72.
- Hine, Christine (2002) 'Cyberscience and Social Boundaries: the Implications of Laboratory Talk on the Internet', *Sociological Research Online* 7(2) < www.socresonline.org.uk/7/2/hine.html > .
- Joerges, Bernward & Terry Shinn (2000) 'Research-Technology: Instrumentation between Science State and Industry', in *Schriftenreihe der Forschungsgruppe 'Metropolenforschung'*

- des Forschungsschwerpunkts Technik – Arbeit – Umwelt am Wissenschaftszentrum Berlin für Sozialforschung, FS II 000-503.
- Joerges, Bernhard & Terry Shinn (eds) (2001) *Instrumentation Between Science State and Industry. Sociology of the Sciences Yearbook* (Dordrecht: Kluwer Academic Publishers).
- Jordan, Kathleen & Michael Lynch (1998) 'The Dissemination, Standardization and Routinization of a Molecular Biological Technique', *Social Studies of Science* 28(5/6): 773–800.
- Knorr-Cetina, Karin (1999) *Epistemic Cultures: How the Sciences Make Knowledge* (Cambridge, MA: Harvard University Press).
- Kohler, Robert E. (1994) *Lords of the Fly: Drosophila Genetics and the Experimental Life* (Chicago, IL: University of Chicago Press).
- Latour, Bruno & Steve Woolgar (1986) *Laboratory Life: The Construction of Scientific Facts* (Princeton, NJ: Princeton University Press).
- Lenoir, Timothy (ed.) (1998a) *Inscribing Science: Scientific Texts and the Materiality of Communication* (Stanford, CA: Stanford University Press).
- Lenoir, Timothy (1998b) 'Shaping Biomedicine as an Information Science', in Mary Ellen Bowden, Trudi Bellardo Hahn & Robert V. Williams (eds), *Conference on the History and Heritage of Science Information Systems* (Information Today, Inc): 27–45 < www.stanford.edu/dept/HPS/TimLenoir/shapingbiomedicine.html > .
- Lenoir, Timothy (2000) 'Science and the Academy of the 21st Century: Does Their Past Have a Future in an Age of Computer-Mediated Networks?', in Wilhelm Voßkamp (ed.), *Ideale Akademie: Vergangene Zukunft oder konkrete Utopie?* 12 May 2000 (Berlin: Berlin Akademie der Wissenschaften): 113–29 < www.stanford.edu/dept/HPS/TimLenoir/IdealAcademy.htm > .
- Livingstone, David N. (1995) 'The Spaces of Knowledge: Contributions Towards a Historical Geography of Science', *Environment and Planning D: Society and Space* 13(1): 5–34.
- Löwy, Ilana & Jean-Pierre Gaudillière (1998) 'Disciplining Cancer: Mice and the Practice of Genetic Purity', in Jean-Pierre Gaudillière & Ilana Löwy (eds), *The Invisible Industrialist: Manufactures and the Production of Scientific Knowledge* (New York: Palgrave Macmillan): 209–49.
- Lynch, Michael (1991) 'Laboratory Space and the Technological Complex: An Investigation of Topical Contextures', *Science in Context* 4(1): 51–78.
- Lynch, Michael (1993) *Scientific Practice and Ordinary Action: Ethnomethodology and Social Studies of Science* (Cambridge: Cambridge University Press).
- Manovich, Lev (2001) *The Language of New Media* (Cambridge, MA: MIT Press).
- Nentwich, Michael (2003) *Cyberscience: Research in the Age of the Internet* (Vienna: Austrian Academy of Sciences Press).
- Ophir, Adi & Steven Shapin (1991) 'The Place of Knowledge: a Methodological Survey', *Science in Context* 4(1): 3–21.
- Primrose, Sandy B. & Richard M. Twyman (2003) *Principles of Genome Analysis and Genomics* (Oxford: Blackwell).
- Rheinberger, Hans-Jörg (1998) 'Experimental Systems, Graphematic Spaces', in Timothy Lenoir (ed.), *Inscribing Science: Scientific Texts and the Materiality of Communication* (Stanford, CA: Stanford University Press): 285–303.
- Schaffer, Simon (1992) 'Late Victorian Metrology and its Instrumentation: A Manufactory of Ohms', in Robert Bud & Susan E. Cozzens (eds), *Invisible Connections: Instruments, Institutions and Science* (Bellingham, WA: SPIE Optical Engineering Press): 23–56.
- Schröder, Peter (2003) 'Digital Research Data as Floating Capital of the Global Science System', in Paul Wouters & Peter Schröder (eds), *Promise and Practice in Data Sharing* (Amsterdam: NIWI-KNAW): 7–11 < www.niwi.know.nl/en/nerdi2/group_members/paul/publ/datasharing/toonplaatje > .
- Shapin, Steven (1995) 'Here and Everywhere: Sociology of Scientific Knowledge', *Annual Review of Sociology* 21: 289–321.

- Shinn, Terry & Bernward Joerges (2002) 'The Transverse Science and Technology Culture: Dynamics and Roles of Research-Technology', *Social Science Information* 41(2): 207–51.
- Steel, Karen P. (1995) 'Inherited Hearing Defects in Mice', *Annual Review of Genetics* 29: 675–701.
- Weir, Bruce S. (1996) *Genetic Data Analysis II: Methods for Discrete Population Genetic Data* (Sunderland, MA: Sinauer Associates).
- Wellcome Trust (2003) *Sharing Data from Large-Scale Biological Research Projects: A System of Tripartite Responsibility*. Report of a meeting organized by the Wellcome Trust and held on 14–15 January 2003 at Fort Lauderdale, USA (Wellcome Trust) < www.genome.gov/Pages/Research/WellcomeReport0303.pdf > .
- Whitfield, Tanya T. (2002) 'Zebrafish as a Model for Hearing and Deafness', *Journal of Neurobiology* 53(2): 153–71.
- Woolgar, Steve (ed.) (2002) *Virtual Society? Technology, Cyberbole, Reality* (Oxford: Oxford University Press).
- Wouters, Paul (2000) 'The Impact of the Internet on the Sciences', paper presented at 4S/EASST conference 'Worlds in Transition: Technoscience, Citizenship and Culture in the 21st Century', University of Vienna, Austria.
- Wouters, Paul & Colin Reddy (2003) 'Big Science Data Policies', in Paul Wouters & Peter Schröder (eds), *Promise and Practice in Data Sharing* (Amsterdam: NIWI-KNAW): 13–40 < www.niwi.knaw.nl/en/nerdi2/group_members/paul/publ/datasharing/toonplaatje > .

Christine Hine is a senior lecturer in the Department of Sociology, University of Surrey. Her research involves ethnographic approaches to understanding information and communication technologies. While much of her work has focused on science, she has also explored the ethnography of media events on the Internet, and the emergence of beliefs about website design. She currently holds a fellowship funded by the UK's Economic and Social Research Council researching 'The Ethnography of Cyberscience: Information in Contemporary Science'.

Address: Department of Sociology, University of Surrey, Guildford, Surrey GU2 7XH, UK; email: c.hine@surrey.ac.uk