E1-30-03-05

THE NEW KNOWLEDGE ECONOMY AND SCIENCE AND TECHNOLOGY POLICY

Geoffrey C. Bowker
Department of Communication, University of California, San Diego

Contents

Glossary

ADA: Programming language (named after Ada Lovelace, arguably the world's first programmer).
CODATA: Committee on Data for Science and Technology was established in 1966 by the International Council of Scientific Unions in order to regulate scientific standards. (http://www.nrc.ca/codata/welcome.html)
DOS: Disk Operating System – Microsoft DOS was one of the early operating systems for computers.
GRID: The term refers to a 'grid' of high-performance research networks. The web-based grid portal helps computer scientists, scientists and engineers by simplifying and consolidating access to advanced computing systems. One of a number of initiatives allowing distributed scientific collaboration.
Human Genome Initiative: Project to map the human genome.
IUBS: International Union of Biological Sciences.
IUMS: International Union of Microbiological Societies.
Memex: Early electromechanical hypertext work environment envisioned by Vannevar Bush in the 1940s, but never built.
Metadata: Refers to all fields in a database that give contextual data about the information contained therein.
MIME protocol: Multipurpose Internet Mail Extensions protocol – used for handling attachments to email messages.
NASA: National Aeronautics and Space Administration, USA.
QWERTY keyboard: The standard Anglo-American keyboard (QWERTY are the first five letters from the left on the top row of letters).
SPARC: Space Physics and Aeronomics Collaboratory. (http://si.umich.edu/sparc/)
VHS: Vertical Helix Scan (Videotape technology).
Windows/NT: Two operating systems that have replaced DOS.
Z39.50: The standard specifies a client/server-based protocol for searching and retrieving information from remote databases.

Summary

The new knowledge economy has significant implications for the practice of scientific and technological development. There is the possibility in theory of producing a democratic global

scientific community, with open access to technoscientific knowledge and practices – from the production of very large shared databases, to the use of Internet tools by a distributed scientific community. There are also some very real difficulties with effecting this vision. Some of these difficulties are technical, some are social and political.

1. Introduction

For the past few hundred years, many books and articles have begun with a phrase such as: 'We are entering a period of rapid change unimagined by our ancestors'. The statement is both as true and as false now as it has been over the previous two centuries. It is true because we are as a society adjusting to a whole new communication medium (the Internet) and new ways of storing, manipulating and presenting information. We are, as Manuel Castells and others remind us, now in many ways an information economy, with many people tied to computers one way or another during our working day and in our leisure hours. It is false because we are faced with the same old problems – getting food, shelter and water to our human population; living in some kind of equilibrium with nature – as ever we were. How is the new knowledge economy impacting and potentially can impact science and technology policy concerned with sustainable life?

2. The New Technoscientific Information Infrastructure

2.1. What is Infrastructure?

Central to the new knowledge economy has been the development of a new information infrastructure. When we think of infrastructure in a common-sense way, we picture that which runs 'underneath' actual structures – railroad tracks, city plumbing and sewage, electricity, roads and highways, cable wires that connect to the broadcast grid and bring pictures to our TVs. It is that upon which something else rides, or works, a platform of sorts. This commonsense definition begins to unravel when we populate the picture, and begin to look at multiple, overlapping, and perhaps contradictory infrastructural arrangements. For the railroad engineer, the rails are only infrastructure when she is a passenger. Almost anyone can flip an electric switch, for a variety of purposes. When the switch fails, we are forced to look more deeply into the cause – first check the light bulb, then the other appliances on the same circuit, then look at the circuit breaker box, then look down the block to see if it is a power outage in the neighborhood or city, and finally, depending on one's home repair skills, consider calling an electrician. Finally, increasingly many of us are faced with infrastructures designed by one group, that may not work for us. For instance, someone in a wheelchair appreciates the tiny (and not so tiny) barriers that are considered 'wheelchair accessible' by the able-bodied. Four steps can be a mountain if the specific conditions of usability are overlooked.

Infrastructure is not absolute, but relative to working conditions. It never stands apart from the people who design, maintain and use it. Its designers try to make it as invisible as possible, while leaving 'pointers' to make it visible when it needs to be repaired or remapped. It is tricky to study for this reason.

We can begin with Star and Ruhleder's definition of the salient features of infrastructure in order to bound and clarify the term:

- *Embeddedness*. Infrastructure is sunk into, inside of, other structures, social arrangements and technologies;
- *Transparency*. Infrastructure is transparent to use, in the sense that it does not have to be reinvented each time or assembled for each task, but invisibly supports those tasks;

- *Reach or scope*. This may be either spatial or temporal – infrastructure has reach beyond a single event or one-site practice;
- *Learned as part of membership*. The taken-for-grantedness of artifacts and organizational arrangements is a *sine qua non* of membership in a community of practice. Strangers and outsiders encounter infrastructure as a target object to be learned about. New participants acquire a naturalized familiarity with its objects as they become members;
- *Links with conventions of practice*. Infrastructure both shapes and is shaped by the conventions of a community of practice, e.g. the ways that cycles of day-night work are affected by and affect electrical power rates and needs. Generations of typists have learned the QWERTY keyboard; its limitations are inherited by the computer keyboard and thence by the design of today's computer furniture.
- *Embodiment of standards*. Modified by scope and often by conflicting conventions, infrastructure takes on transparency by plugging into other infrastructures and tools in a standardized fashion.
- *Built on an installed base*. Infrastructure does not grow *de novo*; it wrestles with the inertia of the installed base and inherits strengths and limitations from that base. Optical fibers run along old railroad lines; new systems are designed for backward-compatibility; and failing to account for these constraints may be fatal or distorting to new development processes.
- *Becomes visible upon breakdown.* The normally invisible quality of working infrastructure becomes visible when it breaks: the server is down, the bridge washes out, there is a power blackout. Even when there are back-up mechanisms or procedures, their existence further highlights the now-visible infrastructure.

Something that was once an object of development and design becomes sunk into infrastructure over time. Therefore an historical, even archeological approach to the development of infrastructure needs to complement sociological, regulatory and technical studies.

2.2. Building an Infrastructure

Both standardization and classification are essential to the development of working infrastructures. Work done on standards committees and in setting up classification schemes is frequently overlooked in social and political analyses of technoscientific infrastructure, and yet it is of crucial importance.

There is no question that in the development of large scale information infrastructures, we need standards. In a sense this is a trivial observation – strings of bits traveling along wires are meaningless unless there a shared set of 'handshakes' among the various media they pass through. An email message is typically broken up into regular size chunks, and then wrapped in various envelopes, each of which represent a different layer of the infrastructure. A given message might be encoded using the MIME protocol – this will allow various other mail programs to read it. It might then be chunked into smaller parts, each of which is wrapped in an envelope designating its ultimate address and its order in the message. This envelope will be further wrapped in an envelope telling it how to enter the Internet. It will then quite possibly be wrapped in further envelopes telling it how to change from a configuration of electrons on a wire to a radio message beamed to a satellite and back down again. Each envelope is progressively opened at the end, and the original message reassembled from its contingent parts then appears 'transparently' on your desktop. It's the standards that let this happen.

One observation that we can make at once is that it's standards all the way down: each layer of infrastructure requires its own set of standards. We might also say that it's standards all the way up. There is no simple break point at which one can say that communication protocols stop and technical standards start. As a thought experiment, let us take the example of a scientific paper. I write a paper about palaeoecology for the journal *Science*. I know that my immediate audience will not be leading edge palaeoecologists, but a wider community who know only a little about Jurassic plant communities. So I wrap the kernel in an introduction and conclusion that discuss in general terms the nature and significance of my findings. A journalist might well write a more popular piece for the *Perspectives* section which will point to my paper. If this is my first paper for *Science* then I will probably go through quite a complex set of negotiations through the peer review process in order to ensure the correct 'handshake' between my message and its audience – is it written at the right level? Have I used the right tone of authority? – and so forth. Similar sets of standards/protocols arose with the development of office memoranda in the nineteenth century; they can also be seen with the development of email 'genres'. I then need to be sure that the right set of economic agreements have been put into place so that my paper can be found on the web by anyone at any time.

This common vision of disciplinary, economic and network protocols serves the purpose of highlighting a central fact about infrastructures. It is not just the bits and bytes that get hustled into standard form in order for the technical infrastructure to work. People's discursive and work practices get hustled into standard form as well. Working infrastructures standardize both people and machines. A further example will clarify this point. In order for the large scale states of the nineteenth century to operate efficiently and effectively, the new science of statistics (of the same etymological root as the word 'state') was developed. People were sorted into categories, and a series of information technologies were put into place to provide an infrastructure to government work (regular ten year censuses; special tables and printers; by the end of the nineteenth century punch-card machines for faster processing of results). These standardized categories (male or female; professional; nationality, etc.) thus spawned their own set of technical standards (80-column sheets – later transferred to 80-column punch cards and computer screens…). They also spawned their own set of standardized people. As Alain Desrosières and Laurent Thévenot note, different categories for professional works in the French, German and British censuses led to the creation of very different social structures and government programs around them. Early in the nineteenth centuries, the differences between professionals in one country or the other did not make so much difference: by the end of the century these differences had become entrenched and reified – people became more and more like their categories.

At both the technical and the social level, there is no guarantee that the best set of standards will win. The process of the creation of standards for infrastructures is a long, tortuous, contingent one. The best known stories here are the adoption of the QWERTY keyboard (good for its time in preventing keys jamming in manual typewriters; counterproductive now for most in that puts most of the work onto the left hand – a hardship for many, but one appreciated by the southpaws (left-handers) amongst us – but so entrenched that there is no end in sight for it); the victory of the VHS standard over the technically superior Betamax standard; the victory of DOS computing system and its successors over superior operating systems.

So why does the best standard not always win? There are two sets of reasons for this. First is that in an infrastructure you are never alone – no node is an island. Suppose there are 500 users of DOS to every one user of a Macintosh. If I am a software developer, why should I write software for Macintosh computers, when I have a much smaller potential user base? So the strong get stronger. Going down one level, if I have to choose between writing an API (to allow communication between one program and another) for interoperation between my program and an industry standard one or between my program and a rarely used one, it is clear which I will choose. More generally put, a new kind of economic logic has developed around the network infrastructures that have been created over the past 200 years – the logic of '*network externalities*'. The logic runs as follows. If I buy a telephone for $50 and there are only 5 people on the network, then it's not worth very much to me, unless I really like those five people. If 5000 more people buy telephones then I haven't had to outlay another cent, but my phone is suddenly much more valuable. This situation describes *positive externalities*. De facto standards (such as DOS, QWERTY, etc.) gain and hold on to their position largely through the development of positive externalities. I buy a personal computer because I know I will have access to the latest and best software. The second reason for the success of possibly inferior standards is that standards setting is a key site of political work. Arguably some of the most important decisions of the past fifty years have been taken by standards setting bodies: although one does not find a 'standards' section of the bookshop alongside of 'history' and 'new age'. Consider, for example the 'open source' movement. This movement has a long history, running deep into the origin of the Internet, proclaiming the democratic and liberatory value of freely sharing software code. The Internet, indeed, was cobbled together out of a set of freely distributed software standards. The open source movement has been seen as running counter to the dominance of large centralized industries – the argument goes that it puts power over the media back into the hands of the people in a way that might truly transform capitalist society. This promise of cyberdemocracy is integrally social, political and technical. While there is much talk of an 'information revolution', there is not enough of the ways in which people and communities are constituted by the new infrastructure.

There are many models for information infrastructures. The Internet itself can be cut up conceptually a number of different ways. There is over time and between models a distribution of properties between hardware, software and people. Thus one can get two computers 'talking' to each other by running a dedicated line between them or by preempting a given physical circuit (hardware solutions) or by creating a 'virtual circuit' (software solution) which runs over multiple different physical circuits. Or finally you can still (and this is the fastest way of getting terabits of data between two cities) put a hard disk in a truck and drive it down… Each kind of circuit is made up of a different stable configuration of wires, bits and people; but they are all (as far as the infrastructure itself is concerned) interchangeable.

One can think of standards in the infrastructure as the tools for stabilizing these configurations. There is a continuum of strategies for standards setting. At the one end of the spectrum is the strategy of one standard fits all. This can be imposed by government fiat (for example the United States Navy's failed attempt to impose ADA as the sole programming language for their applications) or can take the form of an emergent monopoly (for example, Microsoft Windows/NT). At the other end of the spectrum is the 'let a thousand standards bloom' model. Here we enter the world of APIs (Application Program Interfaces – which permit two applications to share data) and such standards as the ANSI/NISO Z39.50 information retrieval protocol. Z39.50 is a standard that has been developed for being able to make a single enquiry over multiple databases: it has been very widely adopted in the library world. You can use whatever database program you wish to make your bibliographic database with, provided that the program itself subscribes to the standard. This means in essence that certain key fields like 'author', 'title' and 'keyword' will be defined in the database. Now, instead of having to load up multiple different database programs in order to search over many databases (a challenge of significance well beyond the library community as large scale heterogeneous datasets are coming into center stage in a number of scientific and cultural fields) one can frame a single query using the Z39.50 standard and have your results returned seamlessly from many different sources. One language that one hears about these two extremes is that the former is the 'colonial' model of infrastructure development where the latter is the 'democratic' model. There is some truth to the implied claim that one's political philosophy will determine one's choice; and there is some solace for democrats in the observation that with the development of the internet the latter has almost invariably won out – most Internet standards have been cobbled together in such a way that they permit maximal flexibility and heterogeneity. Thus, for example, if one looks at the emergence and deployment of collaborative computing, one finds that programs which try to do it all in one package (such as Notes) have been much less successful than less ambitious programs which integrate well with people's established computing environment: interoperability is the key.

Infrastructural development and maintenance requires work, a relatively stable technology and communication. The work side is frequently overlooked. Consider the claim in the 1920s that with the advent of microfiche, the end of the book was nigh – everyone would have their own personal libraries; we would no longer need to waste vast amounts of natural resources on producing paper; all the largest library would need would be a few rooms and a set of microfiche readers. A possible vision – and one that we should not discount just because it did not happen (anyone who has used a microfiche reader will attest that it's a most uncomfortable experience – whereas if the same resources had gone into the failure as into the successful technology, it probably could have been). However, the microfiche dream, like the universal digital library, runs up against the problem that someone has to sit there and do the necessary photography/scanning; and this takes a huge amount of time and resources. It is easy enough to develop a potentially revolutionary technology; it is extremely hard to implement it – and even harder to maintain it.

Further, one needs a relatively stable technology. If one thinks of some of the great infrastructural technologies of the past (gas, electric, sewage and so forth) one can see that once the infrastructure is put into place it tends to have a long life. Electrical wiring from before World War II is still in use in many households; in major cities there is frequently no good map of sewage pipes, since their origin goes too far back in time. The Internet is only virtually stable – through the mediation of a set of relatively stable protocols (for this reason, it can be called an internetwork technology; rather than a network technology). However, there is nothing to guarantee the stability of vast datasets. At the turn of the twentieth century, Paul Otlet developed a scheme for a universal library which would work by providing automatic electro-mechanical access to extremely well catalogued microfiches, which could link between each other. All the world's knowledge would be put onto these fiches – his vision was a precursor to today's hypertext. He made huge strides in developing this system (though he only had the person power to achieve a miniscule fraction of his goal). However, within forty years, with the development of computer memory, his whole enterprise was effectively doomed to languish as it does today in boxes in a basement – why retrieve information electromechanically using levers and gears when you can call it up at the speed of light from a computer? Much the same can be said of Vannevar Bush's never realized but inspirational vision of the Memex – an electromechanical precursor to the computer workstation. Large databases from the early days of the computer revolution are now completely lost. Who now reads punch cards – a technology whose first major use in the United States was to deal with the massive datasets of the 1890 census, and which dominated information storage and handling for some seventy years? More close to home, the 'electronic medical record' has been announced every few years since the 1960s – and yet it has not been globally attained. Changes in database architecture, storage capabilities of computers and ingrained organizational practices have rendered it a chimera. The development of stable standards together with due attention being paid to backwards compatibility provide an in principle fix to these problems. It can all unravel very easily, though. The bottom line is that no storage medium is permanent (CDs will not last anyway near as long as books printed on acid free paper) – so that our emergent information infrastructure will require a continued maintenance effort to keep data accessible and usable as it passes from one storage medium to another and is analyzed by one generation of database technology to the next.

In order to really build an information infrastructure, you need to pay close attention to issues of communication. We can parse this partly as the problem of reliable metadata. Metadata ('data about data') is the technical term for all the information that a single piece of data out there on the internet can carry with it in order to provide sufficient context for another user to be able to first locate it and then use it. The most widespread metadata standards are the Dublin Core, developed for library applications; the Federal Geographic Data Committee (FGDC – http://www.fgdc.gov/) standard developed for Geographical Information Systems. If everyone can agree to standard names for certain kinds of data, then one can easily search for, say, authors over multiple databases. We have already seen this.

Philosophically, however, metadata opens up into some far deeper problems. Take the example of biodiversity science. It is generally agreed that if we want to preserve a decent proportion of animal and floral life from the current great extinction event (one of six since the origin of life on this planet) then policymakers need the best possible information about the current species size and distribution, as well as the ability to model the effect of potential and current environmental changes. In order to do this, you need to be able to bring together data from many different scientific disciplines using many different information technologies (from paper to supercomputer). Now imagine that I am measuring lake water acidity in the Midwest in order to see if there are any effects from recent acid rain. I might be lucky and come across a very large dataset going back eighty or a hundred years and giving lake water acidity levels at one year intervals. However, this might well not be enough for me to actually use the data. It makes quite a difference if the measurement was taken immediately at the lake or later, when the samples were brought back to the laboratory – there will be different amounts of dissolved carbon dioxide in the sample. And as it happens, it makes a difference which technology of measurement I use – a new technology can provide a consistent jump in values for the same sample. Now to make matters worse, I as a scientist am no expert in measuring lake water: I am an expert modeler, and I just want some data to plug into my model. But the point is that there is no such thing as pure data. You always have to know some context. And as you develop metadata standards you always need to think about how much information you need to give in order to make your information maximally useful over time. And here we circle back to the first difficulty with developing an information infrastructure – the more information that you provide in order to make the data useful to the widest community and over the longest time, the more work that you have to do. Yet empirical studies have shown time and again that people will not see it as a good use of their time to preserve information about their data beyond what is necessary to guarantee its immediate usefulness – thus the medical culture of producing quick and easy diagnoses on death certificates in order to meet the administrative need and free time to get onto the next (live) patient.

So standards are necessary – from social protocols down to wiring size. Indeed, when you foreground the infrastructure of your lives, you will see that you encounter thousands of standards in a single day. However, the development and maintenance of standards is a complex political and philosophical problem; and their implementation requires a vast amount of resources. Standards undergird our potential for action in the world, both political and scientific; they make the infrastructure possible.

2.3. Ownership of Scientific and Technological Ideas and Data

It has often been asserted that science is a public good: meaning that scientific work does not fit into the globally dominant market economy. In the new knowledge economy, however, we are increasingly seeing the penetration of the market right down to the molecular level, right down to the stuff of scientific inquiry. Thus it is possible to patent genes, genetically modified plants, animals and so forth. In this process, there has developed. Taking a fairly wide definition of ownership, we can see three main sets of issues arising from the implementation of this knowledge/information market: control of knowledge; privacy; and patterns of ownership.

By control of knowledge, I refer to the question of who has the right to speak in the name of the science. Since the mid-nineteenth century this has been a fairly simple question to answer: only professionally trained scientists and doctors can speak for science and medicine in turn. Only they had access to the resources that were needed in order to speak authoritatively about a given subject – they had the journals, the libraries, the professional experience. Within the new information economy this is not the case. For example, many patient groups now are being formed on the Internet. These groups often know more about a rare condition (for example, renal cell carcinoma) than a local doctor does – they can share information twenty four hours a day, and can bring together patients from all over the world. This flattening out of knowledge hierarchies can be a very powerful social force. It carries along with it, though, the need to educate the enfranchised public about critical readership of the web. There are many websites which look official and authoritative but in fact only push the hobby-horse of a particular individual. We have through our schools and universities good training in reading and criticizing print media; but we have little expertise as a culture in dealing with highly distributed information sources.

Privacy concerns are a significant dimension of science and technology policy in the new economy. It is now technically possible to generate and search very large databases, and to use these to integrate data from a whole series of domains. As this happens, the potentialities for data abuse are increasing exponentially. Much has been written, for example, about data mining of the Icelandic population. After much public debate, citizens of Iceland agreed to sell medical and genealogy records of the country's 275 000 citizens to a private medical research company. There were two central reasons for choosing Iceland: it has a population that has a relatively restricted gene pool; and it has excellent medical records dating back some thousand years. While the science may prove useful (the question is open); it certainly opens the specter of genetic screening of prospective employees by a given company. It is extremely difficult to keep records private over the new information infrastructure – many third party companies, for example, compile together data from a variety of different agencies in order to generate a new, marketable form of knowledge. There is no point in trying to adhere to the old canons of privacy; however open public debate and education about the possibilities of the new infrastructure are essential.

Thirdly, we will look at patterns of ownership of information/knowledge. Science has frequently been analyzed as a 'public good'. According to this line of argument, it is in the interests of the state to fund technoscientific research since there will be a payoff for society as a whole in terms of infrastructural development. With the increasing privatization of knowledge (as we turn into a knowledge-based economy), it is unclear to what extent the vaunted openness of the scientific community will last. Many refer back to a 'golden age' when universities were separate from industry in a way that they are not today. While a lot of this talk is highly exaggerated (science has always been an eminently practical pursuit) it remains the case that we are in the process of building new understandings of scientific knowledge.

A key question internationally has been that of who owns what knowledge. This is coming out in fields like biodiversity prospecting, where international agreements are in place to reimburse 'locals' for bringing in biologically active plants and so forth. However, the ownership patterns of knowledge of this sort are very difficult to adjudicate in Western terms. For example, consider a Mexican herbalist selling a biologically active plant in a market in Tijuana. He owns the plant, but is not the source of knowledge about biologically active plants. This knowledge does not go back to a single discoverer (as is needed in many Western courts of law adjudicating matters of ownership of intellectual property) but to a tradition held, often, by the women of a collectivity. The herbalist may well not be able to trace back the chain of ownership that goes back to the original harvesting of the specific he or she is selling. Similarly, Australian aborigines or the Native Americans had very different concepts of land ownership from the white settlers; leading to complex negotiations that continue today about the protection of natural resources. We need anthropological/sociological studies of local knowledge (to the extent to which this is being mined by scientists) again in order to help design just frameworks and studies of issues of data ownership in different countries. There is a danger when we talk of the explosion of information in the new knowledge economy that we forget the role of traditional knowledge in the development of sustainable policies for a region. Thus research has shown that management of some parks in the Himalayas has relied on models brought in from the outside and taught to villagers through the distribution of television programs – while at the same time ignoring centuries of local ecological knowledge because it is practice based, and has its own intricate weaving of knowledge about the environment, religious belief and mythological expression and cannot be easily conjured into a form that can be held on a computer.

2.4. Sharing Data

The form of scientific work which has been most studied by sociologists of science is that which leads from the laboratory to the scientific paper by means of the creation of ever more abstract and manipulable forms of data, which Bruno Latour has dubbed 'immutable mobiles'. In this process, there is no need to hold onto data after it has been enshrined in a scientific paper: the paper forms the 'archive' of scientific knowledge (frequently adopting names redolent of this storage ambition, such as the *Archives for Meteorology, Geophysics and Bioclimatology*). The scientific paper, which is the end result of science, contains an argument about an hypothesis (which is proved or disproved) and a set of supporting data which is, saving a controversy, taken on faith by the scientific community. The archive of scientific papers can then be indexed both in terms of arguments made and information stored.

However, over the past twenty years we have seen in a number of new and of formerly canonical sciences a partial disarticulation of these two features of scientific work. Increasingly, the database itself (the information stored) is seen as an end in itself. The ideal database should according to most practitioners be theory neutral, but should serve as a common basis for a number of scientific disciplines to progress. Thus one might cite the human genome initiative and other molecular biological projects as archetypical of a new kind of science in which the database is an end in itself. The human genome databank will in theory be used to construct arguments about genetic causation of disease, about migration patterns of early humans, about the evolutionary history of our species; but the process of producing causation is distinct from the process of 'mapping' the genome – the communities, techniques and aims are separate.

This disarticulation, which operates in the context of producing a working archive of knowledge, is not in itself new. To limit ourselves arbitrarily to the past two hundred years, a significant percentage of scientific work has been involved with creating such an archive. Napoleon's trip to Egypt included a boatload of geologists, surveyors and natural historians, and reflected a close connection between the ends of empire and the collection of scientific knowledge. Thus also Smith's geological survey of Britain or Cook's travels to Australia. Richard's *The Imperial Archive,* published in 1996, presents some wonderful analysis of the imperial drive to archive information in order to exercise control (a theme familiar to readers of Latour). The working archive is a management tool. What is new and interesting is that the working archive is expanding in scale and scope. As French philosopher Michel Serres points out we are now as a species taking on the role of managing the planet as a whole – its ecosystems and energy flows. We now see nature as essentially only possible through human mediation. We are building working archives from the submicroscopic level of genes up through the diversity of viral and bacterial species to large scale floral and faunal communities and the mapping of atmospheric patterns and the health of the ozone layer. There is an articulation here between information and theory, but the stronger connection is between information and action – with competing models based on the same data producing policy recommendations. In this new and expanded process of scientific archiving, data must be reusable by scientists. It is not possible to simply enshrine one's results in a paper, the scientist must lodge her data in a database which can be easily manipulated by other scientists.

In the relatively new science of biodiversity, this data collection drive is achieving its apogee. There are programs afoot to map all floral and faunal species on the face of the earth. In principle, each of these maps should contain economic information about how groups of animals or plants fend for themselves in the web of life (http://curator.org/WebOfLife/weboflife.htm) and genetic information (about how they reproduce). In order to truly understand biodiversity, the maps should not only extend out in space but back in time (so that we can predict how a given factor – like a 3 degree increase in world temperature – might effect species distribution). Very large scale databases are being developed for a diverse array of animal and plant groups and the SPECIES 2000 program of IUBS, CODATA and IUMS has proposed might eventually be merged into a single vast database of all the worlds organisms. NASA's Mission to Earth program is trying to 'document the physical, chemical, and biological processes responsible for the evolution of Earth on all time scales'. The UK Systematics Forum publication The Web of Life quotes E.O. Wilson's invocation: 'Now it is time to expand laterally to get on with the great Linnaean enterprise and finish mapping the biosphere' and speaks of the need to 'discover and describe the Earth's species, to complete the framework of classification around which biology is organized, and to use information technology to make this knowledge available around the world'. These panoptical dreams weave together work from the very small scale molecular biological to the large-scale geological and temporally from the attempt to represent the present to a description of the history of all life on earth. They constitute a relatively direct continuation of the drive for the imperial archive, where the notional imperial archive sought to catalog completely the far-flung social and political empire in order to better govern it, biodiversity panopticons seek to catalog completely the natural empire, for much the same reason. Although they cover only a thin slice of species and environments, they are created to be, and are manipulated as if they were, panopticons.

The information collection effort that is being mounted worldwide is indeed heroic. Databases from far flung government agencies, scientific expeditions, amateur collectors are being integrated more or less successfully into very large scale searchable databases. Science and technology policy analysts have a significant contribution to make to the process of federating databases in order to create tools for planetary management. We can produce means to engage the complexity and historicity of data within the sciences so that social, political and organizational context is interwoven with statistics, classification systems and observational results in a generative fashion. We need to historicize our data and its organization in order to create flexible databases that are as rich ontologically as the social and natural worlds they map and so which might really help us gain long term purchase on questions of planetary management.

Even if we can name everything consistently, there are the problems of how to deal with old data and how to ensure that one's data doesn't rot away in some information silo for want of providing enough context. The problem with much environmental data – is that the standard scientific model of doing a study doesn't work well enough. In the standard model, one collects data, publishes a paper or papers and then gradually loses the original dataset. A current locally generated database, for example, might stay on one's hard drive for a while then make it to a zip disk, then when zip technology is superseded it will probably become for all intents and purposes unreadable until one changes jobs or retires and throws away the disk. There are a thousand variations of this story being repeated worldwide – more generally along the trajectory of notebooks to shelves to boxes to dumpsters.

When it could be argued that precisely the role of scientific theory as produced in journals was to order information – to act as a form of memory bank – this loss of the original data was not too much of a problem. The data was rolled into a theory which not only remembered all its own data (in the sense of accounting for it and rendering it freely reproducible) but potentially remembered data which had not yet been collected. By this reading, what theory did was produce readings of the world that were ultimately data independent – if one wanted to descend into data at any point all one had to do was design an experiment to test the theory and the results would follow.

However, two things render this reading of the data/theory relationship untenable. First, it has been shown repeatedly in the science studies literature that scientific papers do not in general offer enough information to allow an experiment or procedure to be repeated. This entails that in a field where old results are continually being reworked, there is a need to preserve the original data in as good a form as possible. Secondly, in the biological sciences in general – and the environmental sciences in particular, the distributed database is becoming a new model form of scientific publication in its own right. The Human Genome Initiative is resulting in the production of a very large collaborative database, for example. In the environmental sciences, where the unit of time for observing changes can be anything from the day to the millennium, there is a great value in having long, continuous data sets. The problem of what data to retain in order to keep a data set live is a metadata problem; and as Ingersoll, Seastedt et al. note in their 'Model Information Management System for Ecological Research': 'the quality of metadata is probably the single most important factor that determines the longevity of environmental data'.

Science is an eminently bureaucratic practice deeply concerned with record-keeping, as Latour reminds us in *Science in Action*. Disciplines do mixed jobs of keeping track of their own results over time – indeed a key finding of science studies has been that using 'theory' as a way of storing old, and accounting for potential data, can be highly problematic since replacement theories do not automatically account for all the data held in the outgoing one (the *locus classicus* is Kuhn's book on *The Structure of Scientific Revolutions* first published in 1962). The difficulties become apparent when you move beyond the arrangement and archiving of data within a given science to look at what happens in the efforts of a vast number of sciences (working from the scale of molecular biology on up to that of biogeography or even cosmology) to coordinate data between themselves within the field of biodiversity. In practice, the sciences use many differing 'filing systems' and philosophies of archival practice. There is no automatic update from one field to a cognate one, such that the latest classification system or dating system from the one spreads to the other. Further it is often a judgment call whether one needs to adopt the latest geological timeline, say, when storing information about ecological communities over time; particularly if one's paper or electronic database is structured in such a way that adopting the new system will be expensive and difficult. Such decisions, however, have continuing effects on the interpretation and use of the resultant data stores.

There has been relatively little work dealing with the organizational, political and scientific layering of data structures. It is clear, though, that the assignation of an attribute to the world of discourse or of materiality is shifting, post hoc. Information infrastructures such as databases should be read both discursively and materially; they are a site of political and ethical as well as technical work; and that there can be no a priori attribution of a given question to the technical or the political realms.

Practically, this means that it is a policy priority to pay attention to the work of building very large scale databases, or developing large-scale simulations. It is no longer the case that knowledge held in a particular discipline is enough to carry out scientific work. From the 1940s on (with the Manhattan project) one might say that large scale technoscience is inherently massively multidisciplinary. However, scientists are not trained to share information across disciplinary divides. And computer scientists cannot do the work of translating between disciplines. Indeed, one of the major difficulties with developing new scientific infrastructures using computers is that the work that is interesting for the computer scientist is often very high-end: involving, say, the latest object-oriented programming and visualization techniques. However, the work that is important for the scientist might be theoretically uninteresting for the computer scientist: for example, producing good ways of updating obsolete databases. There are two sides to the solution here. One the one hand, career paths must be developed which are more in tune with the needs of technoscience. This has worked successfully with the training of a cadre of bioinformaticians with the human genome project at the University of Washington. This cadre knows both molecular biology and computer science – and has a possible career path outside of the confines of the traditional disciplinary structure. On the other hand, we need to put the maintenance of the information infrastructure high on the agenda. Many scientists will go for grants to get the latest equipment; few will concern themselves with upgrading old databases. Huge amounts of data are being lost this way – data about the effects of human activity on this planet which is essential if we are to build a workable future. Just as software re-use has become the clarion call of the latest revolution in programming techniques, so should data-re-use become a clarion call within technoscience.

3. Working Collaboratively

3.1. International Technoscience

There has been much hope expressed that in the developing world, the new information infrastructure will provide the potential for a narrowing of the knowledge gaps between countries. Thus an effective global digital library would allow third world researchers access to the latest journals. Distributed computing environments (such as the GRID, being developed in the United States) would permit supercomputer grade access to computing to scientists throughout the world. The example of the use of cell-phone technology to provide a jump in technology in countries without landlines has opened the possibility of great leaps being made into the information future. As powerful as these visions are, they need to be tempered with some real concerns. The first is that an information infrastructure like the Internet functions like a Greek democracy of old – everyone who has access may be an equal citizen, but those without access are left further and further out of the picture. Further, access is never really equal – the fastest connections and computers (needed for running the latest software) tend to be concentrated in the first world. This point is frequently forgotten by those who hail the end of the 'digital divide' – they forget that this divide is in itself a moving target. Thirdly, governments in the developing world have indicated real doubts about the usefulness of opening their data resources out onto the Internet. Just as in the nineteenth century, the laissez-faire economics of free trade was advocated by developed countries with most to gain (because they had organizations in place ready to take advantage of emerging possibilities) so in our age, the greatest advocates of the free and open exchange of information are developed countries with robust computing infrastructures. Some in developing countries see this as a second wave of colonialism – the first pillaged material resources and the second will pillage information. All of these concerns can be met through the development of careful information policies. There is a continuing urgent need to develop such policies.

International electronic communication holds out the apparent promise of breaking down a first world/third world divide in science. With developments like the remote manipulation of scientific equipment (see the SPARC project, where scientists on the internet can manipulate devices in the Arctic Circle without having to go there. The possibility of attending international conferences virtually is also being held out. And if universities succeed in wresting control over scientific publications from the huge publishing houses (a very open question) then easy/cheap access to the latest scientific articles becomes possible for a researcher in outback Australia... . At the same time, there are a number of forces working to reinforce the traditional center/periphery divide in science internationally. Even with the move to open up access to scientific publications and equipment, there is no guarantee that the 'invisible colleges' – which operate informally and determine who gets invited to which conference and so forth – will change: indeed the evidence seems to be to the contrary. Further, at the current state of technological development there is a significant gap between information access in different regions of any given country, or different parts of the world. Consider the analogy of the telephone. In principle anyone can phone anywhere in the world; in practice some regions have more or less reliable phone services, which may or may not include access to digital resources over phone lines. In fact, half of the world's population does not have telephone access.

We can go beyond the continuing digital divide, however, to consider the possibility of mounting very large scale scientific data collection efforts. Such efforts are central to the social sciences, and to the sciences of ecology and biodiversity. With the development of handheld computing devices, it is becoming possible for a semi-skilled scientific worker with a minimum of training to go into the field and bring back significant results. Thus in Costa Rica, the ongoing attempt to catalog botanical species richness is being carried out largely by 'parataxonomists' who are provided with enough skills in using interactive keys (which help in plant recognition) to carry out their work almost as effectively as a fully trained systematist. Computer-assisted workers together with the deployment of remote sensing devices whose inputs can be treated automatically hold out the possibility of scaling up the processes of scientific research so that they are truly global in scale and scope.

3.2. Distributed Collective Work

Collaborative work is central to the new knowledge economy. Traditionally, scientific breakthroughs have been associated with particular laboratories – the Cavendish laboratory in Cambridge, England for example. Such laboratories have always been a site for collaboration – for example, the exchange of ideas with visiting scholars, the holding of conferences and the training of graduate students. However, it is impossible nowadays to imagine managing a large scale scientific project without including far-flung collaborators – particularly if one is seeking to develop a truly global scientific culture.

Although technoscientific work is inherently collaborative, management structures in universities and industry still tend to support the heroic myth of the individual researcher. Many scientists turn away from collaborative, interdisciplinary work – precisely the kind of work which is most needed in order to develop policies for sustainable life – because they are risking their careers if they publish outside of their own field. There is significant institutional inertia, whereby an old model of science is being applied to a brave new world.

This is coming out most clearly in the area of scientific publications. First in the field of physics and then in a number of other scientific disciplines we are witnessing the spread of electronic preprints, and electronic journals. Traditional academic journals run by huge publishing conglomerates just cannot turn around papers quickly enough to meet the needs of scientists working in cutting-edge fields. Throughout the past two centuries, there has been a relatively stable configuration whereby journal articles have become the central medium for the dissemination and exchange of scientific ideas. Now there is no in principle reason why a scientist should not publish their findings directly to the web. As in many sectors of the new information economy, the development of the new publication medium is leading to a reconsideration of just what kind of value the large publishing houses add to journal production. As more and more journals go online, they are being forced to go beyond their traditional service of providing distribution networks and find ways of bringing their material onto the web. It seems likely that all major scientific journals will soon be accessible on the web; even though the economics of such distribution is not yet fully worked out – possibilities include paying for each paper downloaded, or buying an institutional subscription for a whole journal. A more far reaching implication is that the journal article may no longer function as the unit of currency within the research community. Very large scale collaborative databases, for example, such as the human genome databank, are a new kind of product that made possible by the development of the Web.

The policy implications are clear. Great attention must be paid to the social and organizational setting of technoscientific work in order to take full advantage of the possibilities for faster research and publication cycles. There is a well known paradox about the development of computing – the productivity paradox – according to which the introduction of computers into a workplace tends to lead to a lowering of productivity in the short term (about 20 to 20 years). Paul David and others have argued that what is happening here is that we are still using the old ways of working, and trying to adapt them to the production of electronic text. A new academic field, social informatics, has grown up precisely with the goal of exploring social and organizational aspects of the new knowledge economy (http://www.slis.indiana.edu/SI/).

4. Conclusion

The choices that we are making now about the new information infrastructure are *irreversible*. The infrastructure is *performative* (in that it shapes the forms that technoscience will take) and it is *diffuse* (there is no central control). There is currently widespread belief in technical fixes for inherently social, organizational and philosophical problems – such as curing the ills of incompatible datasets through developing metadata standards. Further, there is a disjunct between the policy and the informatics discourses. Major works on the politics of environmental discourse do not mention environmental informatics at any point: they write as if there is no layer at all between science and politics.

We are globally faced with problems which cannot be solved by the generation of knowledge alone. Producing massive lists of flora and fauna will not of itself lead to a way to preserve biodiversity. However, the information economy does promise some major new tools. The real key to developing technoscientific policy in the context of the new knowledge economy is operating a deep understanding of the nature of information infrastructures. For many, the development of science policy evokes ethical and political questions such as genetic screening, cloning and so forth. This is important work; but much more important is the work of monitoring the standards and classification systems that are getting layered into our models and simulations; and of changing our institutions so that they can take maximum advantage of the new collaborative and information-sharing possibilities.

The central issues for science and technology in the context of the new knowledge economy are the development of flexible, stable data standards; the generation of protocols (both social, in the form of international agreements about data exchange, and technical, in the sense of metadata standards) for data sharing; and the restructuring of scientific careers so that the building of very large scale scientific infrastructures is as attractive a route as the performance of high profile theoretical work. The necessary tools are being created – but they are widely scattered, and are often lost. The new information infrastructure for technoscience will be extremely powerful; with good bricolage we can make it just and effective as well.

Bibliography

Abbate,, J.anet. (1999). *Inventing the Internet*,. Cambridge, Mass: MIT Press. [Standard text on the history of the internet - – good coverage of the technical development.]
Abbate, J., and B. Kahin, B., eds. (1995). *Standards Policy for Information Infrastructure*,. Cambridge, MA: MIT Press. [Useful collection of papers about information standards.].

Bowker, G.eoffrey C., and Susan Leigh Star,. S.usan L.eigh (1999). *Sorting Things Out: Classification and its Consequences*,. Cambridge, MA: MIT Press. [This books examines how classifications are created. Chapter 1 examines why information infrastructures such as databases of political, ethical as well as technical objects.]
Bush, V.annevar, James M. Nyce, J.ames M., and Paul Kahn, P.aul. (1991). *From Memex to hHypertext: Vannevar Bush and the Mmind''s Mmachine*. Boston: Academic Press. [Collection of papers about Bush''s Memex.]
Callon, M.ichel. (1994). "Is Science a Public Good?" *Science, Technology and Human Values,* **19,**: pp. 395–-424. [Article on the argument for the public funding of scientific research.]
Castells, M.anuel. (1996). *The Rrise of the Nnetwork Ssociety*,. Cambridge, Mass.: Blackwell Publishers. [Classic text on the socio-political implications of the information society.]
Connelly, James, and Graham Smith,. Graham (1999). *Politics and the Eenvironment: Ffrom m Ttheory to Ppractice*. London;/ New York: Routledge. [Central text on the politics of environmental science; but without reference to information issues.]
David, P.aul. (1985). "Clio and the Economics of QWERTY",. *American Economic Review, Papers and Proceedings* **75**: pp. 332–-337. [Describes the economics of networks.]
Hajer, M.aarten A. (1995). *The Politics Oof Environmental Discourse: Ecological Modernization aAnd tThe Policy Process*. Oxford: Oxford University Press. [Important texta about environmental policy; does not discuss information issues.].
Ingersoll, R.ick C., Tim R. Seastedt, Ti. m R., and Michael Hartman, M.ichael. (1997). "A model information management system for ecological research,." *BioScience,* **47**,: pp. 310–-316. [Article indicative of recent work to develop information policy for enivronmental science.].
Kuhn, Thomas. S. (1962). *The Sstructure of sScientific Rrevolutions*. Chicago: University of Chicago Press. [Classic text about how scientific change occurs.].
Latour, B.runo. (1987). *Science in Action: How to Follow Scientists and Engineers Throughthrough Society*. Milton Keynes, UK: Open University Press. [Important book about the way that scientists work.]
Michener, W.illiam K., James W. Brunt, J.ames W., John J. H.eelly, J. ohn J., Thomas B. K.irchirchner, T.homas B., and Susan G. Stafford, S.usan G.. (1997). "Nongeospatial mMetadata for the eEcological sSciences,." *Ecological Applications,* **7,:** pp. 330–-342. [Article which helped move the concept of metadata out of the field of Geographic Information Systems.]
Richards, T.homas. (1996) . *The Imperial Archive: Knowledge and the Fantasy of Empire*. London: Verso. [Book which describes the centrality of information for the modern state in the past 200 years.]
Serres, M.ichel. (1990). *Le Contrat Naturel*,. Paris: F. Bourin. [Philosophical enquiryinquiry into the relationship between humans and the planet we live on.]
Star, S.usan L.eigh, and Karen Ruhleder, K.aren. (1996). "Steps toward an Ecology of Infrastructure: Design and Access for Large Information Spaces.," *Information Systems Research,* **7**,: pp. 111–-134. [Analysis of difficulties of creating working information infrastructures for scientists.]
UK Systematics Forum. (1998). *The Web of Life: A Strategy for Systematic Biology in the United Kingdom*. London: UK Systematics Forum. [One of a number of ambitious plans to integrate biodiversity data from many sources.]
Yates, J.oAnne. (1989). *Control through Communication: The Rise of System in American Management.* Baltimore: Johns Hopkins University Press. [Book about the importance of information technology for management.]

Biographical Sketch

Geoffrey C. Bowker is Professor in the Department of Communication, University of California, San Diego. His PhD is in History and Philosophy of Science at Melbourne University. He studies social and organizational aspects of the development of very large scale information infrastructures in scientific and technological work. His first book (*Science on the Run*, Cambridge, MA: MIT Press) discussed the development of information practices in the oil industry. He has written with Susan Leigh Star a book on the history and sociology of medical classifications (*Sorting Things Out:*

Classification and Practice – published by MIT Press in September 1999) and has co-edited a volume on Computer Support Cooperative Work (*Social Science, Technical Systems and Cooperative Work: Beyond the Great Divide*, LEA Press, 1997). He is currently working on a project to analyze the mobility of knowledge in distributed scientific collaborations using high end collaborative software. He is also writing a book about archival practices in the sciences over the past 200 years. His main interest in these areas is the development of biodiversity science as a massively distributed scientific endeavor (covering many scientific disciplines and being carried out all over the world) which has implications for major government policy questions (how to save the environment?) and local political struggles (how to represent knowledge about flora and fauna in such a way that it respects local ways of thinking and is usable by local communities?). More information can be found at his website: http://weber.ucsd.edu/~gbowker.