Bidirectional Shaping and Spaces of Convergence

THIS IS AN ADVANCED DRAFT OF A PUBLISHED PAPER. REFERENCES AND QUOTATIONS SHOULD ALWAYS BE MADE TO THE PUBLISHED VERION, WHICH CAN BE FOUND AT:

García-Sancho M. (2012) "Bi-directional shaping and spaces of convergence: interactions between biology and computing from the first DNA sequencers to global genome databases", *Science, Technology and Human Values,* 37(1): 124-164.
URL: http://dx.doi.org/10.1177/0162243910397969

Bi-directional Shaping and Spaces of Convergence:
Interactions Between Biology and Computing From the First DNA Sequencers to Global Genome Databases[1]

Peter A. Chow-White
School of Communication
Simon Fraser University
petercw@sfu.ca

Miguel García-Sancho
Department of Science, Technology and Society
Spanish National Research Council (CSIC)
miguel.garciasancho@cchs.csic.es

Abstract: This paper proposes a new bi-directional way of understanding the convergence of biology and computing. It argues for a reciprocal interaction in which biology and computing have shaped and are currently reshaping each other. In so doing, we qualify both the view of a natural marriage and of a digital shaping of biology, which are common in the literature written by scientists, STS, and communication scholars. The DNA database is at the center of this interaction. We argue that DNA databases are *spaces of convergence* for computing and biology that change in form, meaning, and function from the 1960s to the 2000s. The first part of the paper shows how, in the 1980s, DNA sequencing shifted from passively incorporating computers to be increasingly modelled in digital coding and decoding. Information retrieval algorithms, reciprocally, were altered according to the peculiarities of DNA in the first sequence-storage databases. The second part of the paper investigates the impact of these reciprocal interactions and globalization on the organization of research centers, ways of conducting big science, and scientific values. Through convergence and new technologies such as data mining, biology and computing were transformed technologically, institutionally, and culturally into a new bio-data enterprise called genomics.

Keywords: Genomics, DNA, databases, data mining, sequencers, convergence, globalization, Internet, biology, computing

> Through the 1970s, a small group of individuals began to realize that computers and sequence information were a natural marriage. Bride and groom struggled to overcome vast cultural differences. Computer scientists and molecular biologists traced their lineage through different tribes, with vastly different norms, and only a few hardy souls could converse in both languages and command respect in both communities. The database that stored sequence data became their meeting ground.
> (Cook-Deegan 1994, 285)

> The fact that the development of computer technology, with its demands on information theory, has occurred contemporaneously with the growth of molecular biology has not merely provide the physical technology, in instrumentation and computing power, without which the dramatic advances of the decades since the 1960s would not have been possible. It has also given the organising metaphors within which the data was analysed and theories created.
> (Rose 1997, 120)

Scholars in STS, the sociology of information, and communication have suggested that there has been a co-evolution of genomics and information technologies over the past 30 years. Innovations in biomedicine between the 1970s and 1990s provided the technological foundation for the rise, in the subsequent decades, of genomics, the Human Genome Project (HGP) and other large-scale initiatives directed to the sequence of nucleotides in the DNA molecule of different organisms. At about the same time, developments in electronics that clustered around the personal computer resulted in a network of information and communication technologies that are epitomized by the Internet. Genomic and information technologies have become two of the most important instruments of the so-called information age (Burnett and Marshall 2003). Such are their interconnections that both genomic researchers and social sciences scholars claim that biology has become "an information science" (Hood 1992; Gilbert 1992; Castells 2000 [1996]; Zweiger 2001; Capra 2002; Marturano 2003; MacKenzie 2003).

This appreciation is shared by members of the International HapMap Project, a multi-nation collaboration that collected and sequenced genomes from individuals from four different global locations (Chow-White 2008).[2] When asked about the impact of information technologies on their research, the first words used to describe them by HapMap scientists were "central" (Interviews 1001, 1002), "essential" (1005, 1009, 1013, 1014, 1017), "paramount" (1010), "fundamental" (1016), and "foundational" (1005). "If it wasn't for technology", they claim, "it would be unfeasible to handle the large amount of data" required by genomic research (Interviews 1003 and 1005). The HapMap participants, however, are less unanimous when referring to the particular form and effects of the interactions between biology and computing.

There are also inconsistencies and disagreements in the STS literature about the nature of these technological interactions. Scholars who wrote shortly before and after the conclusion of the HGP tend to refer to the convergence of genomic and information technologies as a "natural marriage." This label implies a sense of inevitability, as if biology and computing were predestined to coalesce. By comparing the functioning of the DNA molecule with a computer program or a code, popular and some scholars portrayed it as particularly suitable to digital analysis, especially after the development of sequencing techniques and the personal computer in the late 1970s and 1980s (see

opening quote from Cook-Deegan; Moody 2004).[3] The natural marriage approach also suggests a purely instrumental association in which both biological and information technologies, once combined, have maintained their previous identity.

There is, however, evidence of the interactions with computing being more complex for biology and, more generally, scientific research (Agar 2006). By raising different case studies, Kling (2000), Bowker and Star (1999) and Boczkowski and Lievrouw (2007) have argued that technological innovation is a socio-technical process and information and communication technologies (ICTs) are socio-technological networks. This means that their introduction into scientific institutions affects their functioning and practices in all their dimensions, such as the way of conducting investigations, organization and interactions between researchers, values arising from their activity, and reception of their work by society.

As modern biology has increasingly relied on computer simulations, computational models and computational analyses of large data sets, scholars argue that this process has led to a theoretical convergence between genomics and information technologies (Gezelter 1999; Haraway 1997). Holdsworth suggests, ''it is not just that computing tools are rather convenient for doing genomics. Rather, [genomic projects] have re-organised themselves around the bioinformatics paradigm'' (Holdsworth 1999, 89). Marturano (2003), Burk (2002) and Lyon (2005) claim that genomic projects are not only biomedical enterprises they also bioinformatic ones, which makes genomic technologies currently inseparable from information technologies.[4]

An STS problem arises from the view that genomics and information technologies have converged. If they are not just instrumentally associated, but theoretically interdependent, what is the concrete nature of such interdependence and its effects for biomedical research? Lenoir gives a tentative answer to this question by arguing that the computer and the database have decisively shaped biomedical theory. Biomedical researchers have progressively reduced laboratory experimentation and aimed to draw their conclusions from electronic datasets (Lenoir 1999). In a more recent work, Lenoir and other scholars engage with the literature on the post-human body (Haraway 1997; Hayles 1999) and claim that the boundaries between the biological and the computational have become increasingly blurred, given the increasing presence of flesh-and-wire cyborgs in current societies (Lenoir 2002a, 2002b).

The image of biology as a *digitally shaped* and *data-bounded* science has inspired literature on the governance, expectations, public participation and emergence of bioinformatics and biomedical databases, written from the perspectives of bioethics, sociology, and philosophy of science (e.g. Bowker 2008; Fortun 2008; Tutton 2007; Gibbons et al. 2007; Holdsworth 1999). However, some scholars question the suitability of this perspective for capturing the development and current state of biomedicine. From an anthropological viewpoint, Fujimura and Fortun each show the propagandistic motivations in the labeling of genomics as an information science, as well as the opposition by some biomedical researchers. They argue that biomedical practices are still necessary to produce meaning from the stored data (Fujimura and Fortun 1996; Fujimura 1999). In a similar fashion, Hine claims that computing technologies alone are not sufficient to transform the biomedical sciences. They should, hence, be regarded in combination with other scientific and social factors involved in research (Hine 2006). These arguments link with historical narratives in which the connections between

biology, data and computation are traced back to the 1940s and the practices of biomedical research have decisively shaped the development and design of associated computing technologies (de Chadarevian 2002, ch. 4; November, 2004, 2006; Leonelli 2010; Suárez-Díaz. 2010; Suárez-Díaz and Anaya-Muñoz, 2008; Kay, 2000; Sarkar, 1996; García-Sancho 2007a; Lenoir 1999)..[5]

Our paper will build on these critical perspectives on the interactions between biology and computing. By drawing on a historical analysis of the first DNA sequencing technologies and on a series of interviews with participants in the International HapMap Project (see Appendix A) we will show that the effects of the convergence between genomics and information technologies cannot be limited to biomedical theory. Other factors such as the organization of genomic centers and values arising from this sort of research are also shaped by the progressive incorporation of computing. We will also argue that the interactions between biology and computing began as bi-directional, where the state of knowledge and progress in the life sciences also decisively impacted the development of information technologies. At the center of this interaction was and still is the DNA database. We argue that DNA databases are *spaces of convergence* for computing and biology that change in form, meaning, and function from the 1960s to the 2000s. Biology and computing converged over time and the lines between them are blurred into a new type of venture called genomics, in which the biological and the computational are currently indivisible.

STS scholars Kleinman and Vallas (2001, 2006) define convergence as the trading of institutional norms, practices, knowledge, and technologies between the boundaries of academia and industry. Kleinman (2003) explores the effect of convergence on academic culture and, in particular, the effects of commercial cultural norms of competition and entrepreneurship. While Kleinman's theory and ethnography in an industry funded university lab shed light on the ways in which asymmetrical convergence influences science in the academy, his research does not say much about convergence between disciplines within the academy.

In the field of communication and media studies, scholars locate the origins of the concept of convergence with Pool's (1983) foundational text on changes in the media industries (Jenkins 2006). Pool explained that the lines between different communication industries such as news, television, film, and telephony were blurring and innovations and developments in information technology played a central role. Pool saw this as a prolonged transition marked by competition and collaboration between different media systems, rather than through a lens of inevitable technological progress marked by an information revolution. Jenkins argues that convergence is not simply the result of the implementation of a specific organization practice or technology from one institutional context into another, it "represents a paradigm shift" (Jenkins 2006, 15). With paradigm, he refers to the alteration of social relationships, relationships between institutional actors, related enterprises, and cultural logics. Most importantly, convergence is not an end product or the marriage or fixed relationship between two organizational parties: "It operates as a constant force for unification but always in dynamic tension with change" (Pool 1983, 53-54). Convergence is, thus, a process that unfolds over space and time.

We borrow the term *spaces* from Castells concept the *space of flows*, which he defines as the "material organization of time-sharing social practices that work through

flows, [which are] purposeful, repetitive, programmable sequences of exchange and interaction between physically disjointed positions held by social actors" (Castells 2000, 442). Castells argues that the dominant logic of spatial, cultural, and material organization in the information age is the network. Geographical locations do not become irrelevant, "but their logic and meaning become absorbed in the network" (Ibid, 443). For our purposes in this paper, there are two layers to our use of the term *spaces*. Spaces refer to the materiality of emerging spatially configured global hubs of expertise, public research institutes, biotechnology firms, and financiers. These hubs or strategic places (Sassen 1998) change over time. The developers of the early sequencing technologies and databases in the 1970s and 1980s were based mainly in the US and Europe. In the 21st century, other players have positioned themselves to be major nodes in the global genomics network. Spaces also refer to the digital communication networks and databases that provide a second type of material support for the global flow of genome information, stakeholders, and capital.

Jenkins and Pool are interested in the convergence of media and Castells is interested in the spaces of flows for reconfiguring of social, political, and cultural organization. We are interested in how these two concepts can help us view the development of a new research field in the late 20th and early 21st century around DNA databases. Spaces of convergence are technologically mediated processes of communication. They are the space of flows of people, disciplinary expertise, finance, cultural values, institutional ethics, technology, information, data, and code. At the core of the convergence of biology and computing are genome databases. They are currently connected in a global network between university labs, global genome projects, biotechnology companies, state sponsored research institutions, and public interest organizations. Through an interdisciplinary approach, combining history of science and communication, we undertake a historical and contemporary analysis of the techno-social shaping of genomics to fully understand the nature of the investigated interactions.

In the first part of the paper, we address the historical genesis of the interactions between biology and computing in the 1970s and 1980s during what we call the formative phase of DNA sequencing and databases. Like many other enterprises in society during those decades, biologists increasingly used computers for different tasks and created new computer-mediated techniques through routinization, experimentation, and automation (Zuboff 1988). Automation is not an entirely new phenomenon and largely characterized machine-assisted work in the industrial age. Similarly, business owners and managers deployed information technologies to speed up and routinize tasks as well as expand communication networks. However, the key shift Zuboff identifies is how computer mediated tasks,

> …simultaneously generate information about the underlying productive and administrative processes through which an organization accomplishes its work… when the technology also informates the processes to which it is applied, it increases the explicit information content of tasks and sets in motion a series of dynamics that will ultimately reconfigure the nature of work and the social relationships that organize productive activity. (Zuboff 1988, 10-11)

In the 1980s, Zuboff argued that computers were able to automate human activities as well as informate. The intersection between computing and biology, between computer

codes and genomic codes, is more than technical (Thacker 2004). The convergence of the two breaks down ontological distinctions and cultural distinctions. In short, the process of informatization brought these two fields together in the DNA database where "the biological "informs" the digital, just as the digital "corporealizes" the biological" (Thacker 2004, 7). During this formative phase, we show how DNA sequencing instruments and databases shifted from passively introducing the available computing technologies for automating sequencing to informating and modeling their operations on digital coding and decoding. Database algorithms were also adapted to the biological functioning of DNA sequences and laboratories incorporated the organization and managerial models of the recently consolidated information technology industry.

In the second part of the paper, we explore the consolidation and expansion phase of biology and computing from the 1990s to the early 2000s where DNA databases expanded and proliferated due to developments in communication networks and, especially, the innovation of data mining. The Internet, as an emerging space, becomes a key element in the convergence of genomic and information technologies triggered by the increasingly sophisticated databases. Databases built by scientists in the first phase tended to be locally based and limited in scope and content. Similar to other enterprises in business and science, the database as a space of convergence flourished in the virtual space of the Internet.

During the 1990s, scholars investigated worldwide social, political, and economic changes under the rubric of globalization (Held 2000; Waters 2001). The convergence of biology and computing was no exception to the speeding up and intensifying of international social relations and the way in which genomics developed during the 1990s and early 2000s prompted the suggestion that "genomics is globalization" (Thacker 2006, 47). The spaces of technological, political, and economic flows enabled scientists, data, DNA, and the human genome to become globalized. We investigate the effects of this globalization in the emergence of increasingly large, international and virtual spaces of collaboration organized around biomedical databases. The databases prompted the generalization of values and concepts such as discovery science and open access. Within them, biology and computing were transformed technologically, institutionally, globally, and culturally into a new bio-data enterprise called genomics.

Computing enters biology: from passive incorporation to active modeling.

The first researchers involved in sequencing biological molecules (proteins, RNA and DNA, between the 1950s and 1970s) simply introduced already existing computer applications into their techniques. Early sequencing methods and databases incorporated programming strategies previously used for the management of data produced by large administrative and corporate offices, the main clients of the then emergent computing and software industries. With the development and automation of these biological instruments in the 1980s, the operations for sequencing DNA were altered to be adaptable to the computer. Likewise, special software was designed to deduce features from the sequences after their database storage.

The Development of the Automatic Sequencer

Shortly after Fred Sanger and Walter Gilbert independently invented methods to deduce the sequence of nucleotides in the DNA molecule (1975-77), researchers began seeking ways to incorporate the computer into this process. The early DNA sequencing

software, developed between the late 1970s and mid 1980s in the US and Europe, allowed scientists to edit and store the sequences in files.[6] The first sequencing program used at Sanger's group was designed with the help of a professional computer programmer, who was the brother-in-law of a member of the laboratory (McCallum and Smith 1977; Sanger and Dowding 1996, 344-45). IBM, Bell Laboratories and other computer manufacturers were at that time producing word processing software to assist in the writing and revision of texts in government offices, banks, travel agencies, and other large private companies (Haigh 2006b). The designers of sequencing software acknowledged in later versions of their programs that they were conceiving the sequences as "words" and using the algorithms – programming orders – applied in searches within texts by the word processors (Staden 1982, 4743; Dumas and Ninio 1982, 197).

This early sequencing software required manual operation in all its steps. The user needed to input the sequence, edit it, and save the results on tape or early magnetic disks. Later program versions attempted to automate the introduction of the sequence through interfaces such as a digital stylus or a proto-scanner (Staden 1984; Sulston et al 1988). However, they faced a pervasive problem: the outcome the sequencing methods produced was difficult to process by the computer.

Both Gilbert and Sanger's methods – respectively developed at Harvard University and the Laboratory of Molecular Biology of Cambridge, UK (LMB) – yielded at the end of the sequencing process a picture called autoradiograph. On it, the DNA sequence was represented as a two-dimensional pattern of black spots. Each spot corresponded with a nucleotide in the DNA molecule and, from their position in the picture, the researcher could deduce the sequence (Sanger 1975, 1988; Sanger et al, 1977; Maxam and Gilbert 1977; Gilbert 1980). The pattern was easy to interpret with the eye, which could distinguish between the spots and slightly correct their position (García-Sancho, 2010). A computer, on the contrary, faced constant difficulties, since all the spots possessed the same shape and color, and their location was sometimes ambiguous within the two-dimensional pattern.

A group led by Leroy Hood at the California Institute of Technology (Caltech) was seeking ways to automate the sequencing process since the late 1970s. Their first attempt was to eliminate human intervention in the processing of the autoradiographs, but given the problems the computer faced with the spot pattern, researchers found no success (Caltech 1980, 52). After this disappointment, Hood's team decided to "abandon the autoradiograph world" and to create "a new approach to sequencing" that differed in important ways from Sanger and Gilbert's methods (Interview with Smith; Interview with Hunkapiller).

In Sanger and Gilbert's techniques, the black spots on the film were a consequence of labeling the DNA molecule with a monochrome radioactive substance. Hood's team, instead, used fluorescent dyes of various colors and applied a different color depending on the DNA nucleotide to which each spot corresponded. This way, since the nucleotides were distinguishable, they could be aligned one-dimensionally on the film and be presented as a row in which, for instance, adenine was marked red, cytosine blue, etc. A computer could easily process these colored patterns (Smith et al 1986).

Figure 1 goes about here

The differences between the manual and automatic approaches to sequencing were largely motivated by the distinct strategies operating in the laboratories where the inventors were based. Gilbert and Sanger's groups, respectively at Harvard and the LMB, were pursuing basic biomedical research. The LMB had traditionally been funded through a block grant scheme by the Medical Research Council (MRC), a body of the British Government which predicted the financial necessities of the laboratory over long periods of time. This system, which resisted attempts of liberalization due to the support of scientists and MRC officers, allowed Sanger and other LMB members to work without the pressure of finding applications to their research (de Chadarevian 2002, Part III). Conversely, Hood's team belonged to Caltech, a technical school that depended on research contracts. These contracts could be signed with either the US Administration or private institutions, but were always oriented to particular research outcomes (Kay 1993).

These contrasting settings resulted in different attitudes towards sequencing and the pursuit of research more generally. Whereas for the LMB research always implied human involvement, at Hood's group some of the research activities could be perfectly automated. This led Sanger and his assistants to apply the manual sequencing techniques to various microorganisms since the late 1970s without devoting further efforts to their automation (e.g. Sanger et al 1977, 1982). The Caltech team, in contrast, started at the same time a program to automate sequencing, which was considered a repetitive and monotonous practice (Caltech 1980-1985). Since the early 1980s, Hood created a spin-off company, Applied Biosystems, to commercialize the resulting automatic sequencers. Neither Sanger nor Harvard had been especially proactive towards the then emergent biotechnology market (Kenney 1986; Bud 1993). When in the early 1980s, Gilbert attempted to create the biotechnology company Biogen, he faced firm opposition by Harvard academic authorities, who argued that this was not an appropriate activity for a University (Mendelsohn 1992, 17-19).

The different priorities at Sanger's and Hood's groups also fostered divergent attitudes towards computing. Sanger and his group only applied computers to sequencing in the late 1970s, once the DNA methods had been developed by exclusively biochemical means. This was done despite software applications for protein and RNA sequencing being available since the 1960s (e.g. Dayhoff and Ledley 1962; Needleman and Wunsch 1970; see note 7), and Sanger having been hitherto involved in both fields (García-Sancho 2010). Furthermore, instead of adapting this previous software, Sanger's group incorporated a researcher with computing expertise – Rodger Staden – who designed DNA sequencing programs always involving, to a certain degree, human intervention. This intervention was maintained in further versions of the software used at the LMB throughout the 1980s and presented as a necessary human check over the computer (Sulston et al 1988, 126).

Conversely, the Caltech team modeled the entire sequencing process on the computer. In the early 1980s, the members of Hood's group considered Sanger and Gilbert's techniques unsuitable for automation (Interview with Hunkapiller; Interview with Smith). They decided to modify the manual approach to sequencing and all the modifications they introduced sought to make the sequencing outcome easy to process for the computer.

First, the horizontal alignment of the nucleotides and their labeling with four different colors sought to transform the DNA sequence into a string. Lloyd Smith and Tim Hunkapiller, the main proponents of this strategy, had been developing software in their previous research and aiming to introduce computers into Hood's laboratory (Smith, 2008; Interview with Hunkapiller). During the early and mid 1980s, the processing of strings as one-dimensional linear arrangements of discrete data was becoming a main problem of computing. The text processing software then expanding at public and private offices incorporated algorithms directed to strings and these algorithms had been exported by biologists to early DNA sequencing programs (Staden 1982; Dumas and Ninio 1982).[7]

Second, the Caltech team used a laser to scan the color string in order to transform this information into sequence data and transfer it to the computer. Since the 1960s, the laser, a technology with a marked Cold War military origin, transitioned into civilian uses, including detecting tumors and other tissues or molecules of biomedical interest (Bromberg 1991, 208-219). Smith, who became the main responsible for the development of the sequencer, had been using laser technology for measuring lateral diffusion of lipids and proteins in cell membranes prior to his arrival at Hood's group (Smith et al, 1981).

Also, between the early and mid 1980s, the laser began to be used in processing digital information in computer discs (Guenther et al 1991). The disc manufacturers organized the data in one-dimensional and linear arrangements of spots, mirroring the strings in which the Caltech group was transforming the DNA sequences. Other domestic devices emerging at that time – such as the CD or the supermarket barcodes – incorporated the same technology (Campbell-Kelly and Aspray 1996, ch. 7). Surrounded by these developments, Smith referred to the laser as a "detector" of sequence "information" in a 1986 report written after he developed the first prototype automatic sequencer (Caltech 1986, 73-74).

Figure 2 goes about here

STS literature has generally considered the personal computer and first DNA sequencing software as the technologies, which broke the "barrier" between biology and informatics in the late 1970s (Moody 2004). With these technological achievements, biologists began using the computer to sequence DNA and the natural marriage between both fields started (Cook-Deegan 1994). This scholarship overlooks the necessity of adapting the biological processes to the computer for the technologies to converge. In the case of sequencing, the process was not fully automated until the DNA outcome was transformed into a color string to be processed by a laser, at that time beginning to be used as a digital decoding device.

Keating, Limoges and Cambrosio have used Caltech's effort as a case study to argue that successful automation does not imply the mimicking of the previous manual procedures. According them, Hood's team succeeded unlike rival automation attempts for substituting manual sequencing by another process which "redistributed" human and technological resources (Keating, Limoges and Cambrosio 1999, 127-32). This new process, we claim, derived from the modeling power of the computer and its penetration into pre-existing biological technologies, such as sequencing.

The modeling impact of the computer did not only affect the organization of the sequencing process. Applied Biosystems (ABI), the company that manufactured and commercialized the automatic sequencer from Smith's prototype, gradually adopted the structure and management models of the information technology industry. When creating the firm in 1981, Hood sought financial support in a series of venture capitalists based in San Francisco's Bay Area. The first chairmen selected by these investors were André Marion and Sam Eletr, former managers of Hewlett Packard (HP), located in nearby Silicon Valley. Marion and Eletr aimed from the beginning to implement "the same strategies" which had made HP a leading company in the computing industry. ABI, after all, would commercialize "biological instrumentation," not that far from the "electronic instrumentation" HP marketed since the 1940s (Interview with Marion).

Marion and Eletr's claim is far from rhetorical. ABI, in its first reports, presented itself as a pioneer firm in the manufacturing of instruments to deal with basic biological "information" (ABI 1983, 4; 1986, central triptych; 1987, 4). Its teams combined research and marketing staff in order to adapt the manufactured sequencer to the necessities and likes of the final users, i.e. the biologists. This strategy involved a rupture with previous traditions of "highly compartimentalized" firms which had proliferated within the chemical and pharmaceutical industries since the early 20th century (Interview with Marion; Chandler Jr. 2005). In the mid and long term, it was crucial for biologists used to the manual techniques accepting ABI's device, which was substantially redesigned according to their feedback during the second half of the 1980s (García-Sancho 2008, pp. 188 and ff.).

Lécuyer, a business historian, has shown that the electronic instrumentation companies of Silicon Valley –Fairchild Semiconductor, Intel and National Semiconductor – implemented this strategy of coupling "product development with market demands" between the 1950s and 1970s. He has also documented frequent transfers of expertise and capital between the information technology and emerging biotechnology industries of the Bay Area during the late 1970s and 1980s (Lécuyer 2006, 165 and 292-94). In the case of ABI, these transfers were materialized in the organization and managerial models of the electronic instrument company HP, whose incorporation was essential for the successful commercialization of the automatic DNA sequencer.

The automatic sequencer, therefore, was modeled on the computer from its conception to its manufacturing and final marketing to the users. The shaping power of information technologies did not only affect the operation of the sequencer, but also the structure of its production and commercialization. This shaping power was not, yet, one-directional. The development of another technology, the database to store the DNA sequences, shows that biological processes also informed the computer instruments which were incorporated to them.

The DNA Sequence Database

The spread and automation of sequencing triggered the development of a related technology, the database, to store the growing DNA sequences. The first DNA sequence databases emerged in the early 1980s, before the automation of the techniques and through large-scale national and international initiatives in Europe, the United

States, and Japan (Smith 1990; Strasser 2008; Cook-Deegan 1994, ch. 15). These DNA banks, though, were not the first incorporation of database technology into biology, for there had been computer-based repositories in the life sciences since the 1960s (Strasser, in press). Nevertheless, the DNA banks were the first in fully adapting the available database applications to the peculiarities of the stored sequences.

The database as a computer-based technology originated in World War II, when it was used in military operations. During the 1950s and 1960s, its use spread to public administration and the private corporate office. These databases were designed by large computer manufacturers such as IBM and an incipient software industry. Their main property was to allow users to gather information from multiple sources and to draw "vital intelligence" through its comparative analysis (Haigh 2001, 16, 2006a; Kline 2006; Campbell-Kelly 2003). An anti-aircraft fire system, for instance, could predict the position of enemy planes by combining data about their speed, trajectory and weather conditions. Equally, an insurance company or library could know the employees close to retirement or the overdue loans.

Biologists began using computerized databases in the mid 1960s within the fields of biophysics, biochemistry and human genetics. Olga Kennard, Margaret Dayhoff and Victor McKusick published, between 1965 and 1966, compilations of X-ray analyses of molecules, protein sequences and hereditary diseases respectively. All of them used computers in the form of punched card machines which permitted to store the gathered data in entries, retrieve such entries, and derive new knowledge through their comparison. Dayhoff attempted to reconstruct the evolutionary pathways between species from their protein sequences, while Kennard deduced the three-dimensional structure of molecules from their X-ray coordinates (Strasser 2006, in press; Kennard 1997).

A key difference between these early 1960s databases and the ones devoted to store DNA sequences was that the latter were run by computing and information management experts with little biological expertise. With the progressive reduction of size and price of computers since the 1960s, the large and external mainframes were substituted by microcomputers, workstations, and personal computers that would become permanent fixtures in laboratories (Ceruzzi 1998). This resulted in experts in the new technologies entering biological and other scientific institutions (November 2004; 2006). In the early 1980s, at the time of the international DNA sequence database projects, the computer and specialized staff moved into the laboratory and no longer received the punched cards prepared by biologists in remote offices.

The first information technology staff to develop a DNA sequence database was incorporated to the European Molecular Biology Laboratory (EMBL) in 1980. This institution issued a job vacancy announcement in which the key requirements for the position were having a background in "mathematics, physics or computer science", together with "numerical and statistical analysis" and the "development of computer programs". Since the position was at the level of "research assistant" or "manager", holding a PhD was considered a merit, but not a compulsory requirement for the job. The same applied to biological expertise.

Figure 3 goes about here

None of the professionals hired by the EMBL fulfilled those additional biological merits. Greg Hamm, the first database staff member, studied a combined degree in biology and engineering, but after graduation acquired a considerable work experience in the US computer industry designing military software. Graham Cameron, appointed in 1982, had abandoned an undergraduate degree in psychology and worked in the development of a university database with household information. Both of them acknowledge that biological expertise was not essential during their early work. The crucial skill was "understanding information", i.e. the workings of the data they stored. Hamm and Cameron did not consider themselves biologists, but "information engineers" (Interview with Cameron; Interview with Hamm). This term, together with "systems men", arose between the 1950s and 1960s, in the context of the use of databases in wartime operations, public administration and offices. It had, consequently, developed away from biology and the academic world (Haigh 2001; Mindell 2002).

One of the first conclusions of the new information engineers at the EMBL was that the available database technology was not appropriate for their task. The database structures that had been developed by the early 1980s represented a "table view of the world" with which DNA sequences did not square (Interview with Hamm). IBM and other producers had designed, between the 1960s and 1970s, a number of database models adapted to the necessities of their costumers. Their structures and ways of managing entries were, therefore, prepared to deal with independent data about prices, products, citizens, ages and properties, the main variables with which government departments or travel agencies operated (Date 1981 [1975]). The DNA molecule, however, worked as a string of interconnected units. Its constituent nucleotides were assembled by chemical bonds and could not be processed as discrete or independent data.

Hamm and Cameron, consequently, started a systematic review of the professional literature on computing in search for instruments to transform the database models. One of the main concerns of computer scientists in the early 1980s was the development of tools to handle the emerging word processing software (Haigh 2006b). This had led to the proliferation of algorithms – programming commands – such as 'check', 'find' or 'format', which allowed detection of patterns in a written text.

The first entries of the EMBL database were edited in an early text processor. This allowed Hamm and Cameron to realize the potential of this technology for managing the DNA sequence data. By adapting text processing algorithms designed to check the spelling or search for words, they could automatically detect errors and deduce properties from the stored sequences. However, Hamm and Cameron previously needed to transform the algorithms and use, as a reference to search and check the sequences, the rules governing the functioning of the DNA molecule rather than those of English orthography and grammar. For instance, biologists knew that genes were always surrounded by two specific sequences – initiation and termination codons – and, using this knowledge, Hamm and Cameron could program the database to automatically find genes within the stored sequences (García-Sancho in press).

The first two releases of the European database (1982 and 86) stored the position of the genes and other deduced sequence features in a specific section of the entries called *Feature Table* (Hamm and Stübert 1982; Hamm and Cameron 1986). This and

the nature of the stored information made the EMBL database different from the ones designed the decades before by the computer and software industries. Hamm and Cameron's entries were no longer regular tables with qualitative and quantitative tags attached to each category (e.g. name "John Smith"; age "23"; years in the company "13"; holidays "10th-23rd August"). They were dominated by a long string of interconnected characters – the DNA nucleotide sequence – from which other details were extracted.

Figure 4 goes about here

Hamm and Cameron combined biology and computing in a different way from other researchers developing sequencing technologies. They did not import information technologies from outside biology, as previous biological databases and the developers of the first sequencing software had done. They also did not adapt biological processes to available computing technologies, as in the case of the automatic sequencer. Given their condition of insiders in information management and computer science, Hamm and Cameron inverted the logic of adaptation and transformed the available computer algorithms according to the specifics of the DNA sequences.

This ability for reciprocal adaptation was crucial in the success of both the automatic sequencer and the database. The emergence of genomics as the "new discipline" of DNA mapping and sequencing in the late 1980s (McKusick and Ruddle 1987; Powell et al 2007) led both technologies to increase their funding dramatically and to have a prominent role in the further Human Genome Project. The practices of data gathering and analysis, considered as repetitive and marginal to biology at Kennard and Dayhoff's time (Strasser 2006; in press), were only 20 years later raised to the category of priority by both biological funding agencies and working biologists (García-Sancho, 2009; 2007b; 2007a, pp.27 and ff.).[8]

The human and other large-scale sequencing initiatives accentuated the interactions between biological and information technologies. Genomics as a field was and still is the result of the convergence of those two technologies. In the next section, we utilize interviews with members of a genomics initiative – the International HapMap Project – to explore the role of the database as a space of convergence and its connection with other spaces triggered by the increasing globalization of genomics, such as the spaces of collaboration and the virtual space of the Internet. We also analyze the impact of these converging spaces on concepts and values characteristic of genomics, such as discovery science or the proposal of open access to the DNA sequence data.

The Consolidation and Expansion of Biology and Computing and the Globalizing of Genomics

The next phase of interactions between biology and computing we call the consolidation and expansion phase. In the 1990s and 2000s, the nature of convergence between biology and computing changed and the interaction between the two disciplines cannot be characterized solely by the notion of bi-directionality. We argue that biology and computing are intimately intertwined in genomics, producing new data practices and a new scientific approach to understanding code and the body.

Globalization became a critical part of the shaping of the science of genomics as new constituencies entered the field and genomics expanded internationally. Scholars argue that biology became an information science, as much about if not more about computation than the wet biology of the 20th century (Castells 2001; Capra 2002; Lenoir 1999; Moody 2004; Zweiger 2001). We do not argue that this perspective is entirely wrong, as it has been a productive avenue of inquiry about institutional shifts in biology in the information age. However, this perspective misses any bi-directional influences between the two disciplines and the transformation of a globalizing science. More importantly, the informational turn approach fails to account for how genomics can be understood as a space of convergence between the two fields.

In the second part of the paper, we draw on the example of the International HapMap Project to explore how biology and computing consolidated in genomics and the field expanded into a global enterprise in the 1990s and 2000s. We explore the consolidation and expansion of genomics by focusing on technological innovations, the politics of stakeholders, and the politics of genomic information. Technologically, scientists and entrepreneurs increasingly turned to the new types of online databases that could be accessed due to the proliferation and increased capacity of the Internet, linking labs, researchers, and companies. The networking of labs and databases also enabled the sharing of scientific knowledge globally. The number of databases in existence either in the academy, publically funded institutes, or private biotechnology companies grew enormously. The trickle of digital genetic data from the human genome projects became a flood during the first decade of the 21st century within databases based in the West and new ones across an unevenly distributed international network.

Politically, the central players in genomics expanded from the US and UK, who largely defined the formative phase. For example, the HapMap Project Consortium was composed of a transnational set of stakeholders from public institutes, university labs, and biotechnology companies from Africa, Europe, Asia, and North America. Further, new types of private biotechnology ventures started up in the 1990s, such as Celera Genomics, and in the 2000s, such as direct to consumer genomics companies 23andMe, which is a partnership between the largest global biotechnology company Genentech (recently purchased by Roche) and the largest information company Google. While the stakeholders became more diverse and the technological capacity expanded, a new politics of the database also emerged as scientists and entrepreneurs struggled over the meaning of genomic information and who would have access to it. Many scientists argued genome information should be a public good while entrepreneurs operated from proprietary business models.

The Database as a Space of Convergence: Data Mining and the Shift from Single to Multiple Sequence Analysis.

The major challenges for biologists and computer scientists shifted from sequence storage and single sequences in the formative phase to cross-indexing of multiple sequences using data mining technologies in the consolidation and expansion phase. This shift in focus, in part, characterizes the development of genomics, in which scientific achievements have been increasingly measured in terms of making sense of the sequences through the comparison of data rather than accumulating them in databases. Two technological developments in particular were commonly highlighted among HapMap participants: databases/data mining and the Internet. Biologists,

computer scientists, bioinformaticians, and engineers designed new databases to store, analyze, and distribute the data and findings. The hypertext model of the Internet is used to create methods for annotating genomes.

In 1982, the NIH launched the Genbank database that, similarly to the EMBL one, aimed to collect and annotate all publicly available DNA sequences (Strasser 2008). This particular database has become important in biomedical research not only as a resource but also as a way of encouraging scientists to make their data public. Many journals require submission to a database before authors can submit articles. Sequencing of DNA was slow in the early days. After four years, Genbank had less than 700,000 base pairs (Moody 2004, 26). Developments in sequencing technologies and communication technologies have rapidly sped up the collection of DNA data over the past two decades. DNA collection has globalized as international collaborations have increased through projects such as HapMap and the 1000 Genomes Project, and networked databases such as Genbank, which increased to over eighty-five billion base pairs in 2008 (Genbank 2009). Globally, genome databases exist in Iceland, the UK, Switzerland, Japan, and both Latvia and Estonia have their own genome projects (Kaiser 2002; Fortun 2008). To manage, sort, classify, and analyze digital DNA information, computer scientists, biologists and geneticists worked with computer scientists and bioinformaticians to innovate and develop data mining technologies in genomics. Data mining is a technique for searching and creating knowledge out of digital databases. Derived from the computer sciences, data mining multiplies the possibilities of discovering knowledge in data. Compared to early search programs in the 1980s, data mining software searches databases with more speed, capacity, and complexity. Data mining techniques are made up of more refined algorithms, neural networks, and artificial intelligence models. Information analysts can program data mining software to work from pre-determined sets of categorical variables or they can go beyond what a user knows to request and "discover" unseen patterns, facts, relationships between the data (Chow-White 2008; Danna and Gandy 2002).

During the 1990's, the most important event that consolidated the interactions between biology and computing was the public and private human genome projects through the combining of biological approaches such as the shotgun method for sequencing and computing science innovations such as data mining of databases and the Internet. Data mining techniques and "large, easily-accessible databases that would allow the extraction and comparison of data were absolutely essential for being able to put together any kind of sequence database" (Interview 1014). Data mining has become a central technology in genomics, where computer science theory meets biological theory and practice. Lenoir (1999) describes computational biology and bioinformatics as the theoretical and instrumental/experimental components of genomics. The database is where they converge to construct scientific knowledge through the core technique of data mining. The development of this relationship was key to the success of the private Human Genome Project. Craig Venter's team at Celera developed the biological and informational Shotgun sequencing method that combines Polymerase Chain Reaction (PCR) and data mining. The Shotgun method reduces a strand of DNA into random cloned sections and puts them back together again as a genome. A computer takes a number of these partial sections, rearranges them and stitches them together to form a complete sequence. Since the biologist works at an abstract level, much like software developers, the time-consuming work of mapping is eliminated. Venter's contribution to the sequencing of the human genome sped up the actual process of completing a

working sequence, and compelled Celera's public counterpart in the HGP, led by Francis Collins, to accelerate their approach. The two HGPs also sped up the methodological, theoretical, and technical convergence between the fields of genomics and information technologies.

The genome posed enormous biological challenges in sequencing and mapping, as well as bioinformatic challenges for data generation, storage, and analysis. According to Leroy Hood, "a completely new approach" to biology was needed, which he called "discovery science" (Hood 2001, online). Much like the basic assumptions of data mining described above, discovery science is "the idea that you take an object and you define all its elements and you create a database of information quite independent of the more conventional hypothesis-driven view" (Ibid.). In the traditional scientific method, scientists start with a theoretically sound hypothesis and then collect and analyze data. With discovery science, scientists tend to collect first and ask questions later. It has also been largely credited with the success of the HGP. This new scientific approach and the techniques of data mining are affecting the organizational structures of genomic research teams.

The politics of stakeholders in genomics: Interdisciplinary Teams and Discovery Science

A second major part of the consolidation and expansion of genomics was a cultural shift in science from the individual, disciplinary oriented scientist and lab group to interdisciplinary, team-based approaches. Traditionally, biology was not a quantitative discipline and researchers were generally unaccustomed to working with such large data sets. The "cultural differences" between life sciences researchers and computational staff at biological centers stated by Cook-Deegan that shaped the early development of the EMBL database (Cook-Deegan 1994, 285; see note 9) continued and the different type of personnel and expertise expanded. A HapMap biologist described a slow but increasing migration of computationally sophisticated people into the field of biology during the 1990s and into the 2000s. While the numbers of such people have swelled, the integration of the two cultures was slow. A biologist observed that this convergence required "not just an intellectual shift, but also a real cultural shift because biologists are used to… the limiting step being their ability to collect data with their hands" (Interview 1001). In the age of Google and massive, flexible data sets, informational thinking has clearly taken biologists and computer scientists out of their 'comfort zone' in order to tackle the deluge of data generated in genome research over the last decade. In these spaces of scientific innovation and convergence, biologists, computer scientists, and engineers work side by side, borrowing methodologically, theoretically, and culturally. In the process, biology becomes bound up in data. On another part of the university campus, computing science departments regularly offer courses on genomics, bioinformatics, and computing theory based on molecular biological systems.

This shift from observation to a "data-bound science" (Lenoir 1999, 35) is at the core of the transformation of biology. The days of the individual scientists working in isolation in her lab, scribing notes and models in a notebook, have transformed into multi-disciplinary teams of researchers. For example, the International HapMap Project is made up of biologists, geneticists, statistical geneticists, doctors, legal scholars, bioethicists, bioinformaticians, anthropologists, and sociologists. A broad range of knowledge is needed to conduct large-scale genome research. As the inclusion of social

scientists in the list of personnel shows, a significant part of the research team belongs to the Ethical, Legal, and Social Implications (ELSI) committees. Team-based projects are sometimes focused on a particular problem and located in a particular lab or located across a number of different centers, often in different countries, sharing information in a common, digital database.[9] The database itself may be maintained in a university lab, accessible only to the project group, or in a centralized location, such as the National Institutes of Health or European Bioinformatics Institute – the new institutional setting of the database division of the EMBL – open to the public, and accessible to various research teams around the world. A biologist suggested that there may be a new generation of scientists who will have the breadth of biological and computational skills and knowledge to master all the aspects of genomic research, but for now that does not exist. He also suggested "it may be that no one ever does know all these things, because there are too many things to know" (Interview 1001). For now, big science requires interdisciplinary teams.

The interdisciplinarity and large size of the teams combining computing and biology in genomics has led this field to be considered "big science" (Galison and Hevly 1992). Since its inception in 1990, scientists and scholars defined the HGP as the "Manhattan Project" of biology, given the amount of people involved, their diverse background, the strong technological component of the enterprise, and its unprecedented funding by both public and private institutions (Lenoir and Hays 2000). There is currently evidence of the whole field of biology becoming big science, such as the standardization and routinization of particular practices (Jordan and Lynch 1998; Lynch et al 2008). Some biologists view the farming out of "cookbook techniques," that become repetitive practices and procedures, as a sign of a successful science (Gilbert 1992, 93). As mentioned above, biotechnology companies specialize in practices such as sequencing and provide outsourcing for techniques that used to be performed by skilled researchers. For example, Illumina, Sequenom, and ParAllele served as the genotyping and sequencing centers for the HapMap Project.

The Politics of the Database for "Democratizing the Data" on a Global Scale: The Internet, Distributed Networks, and Open Access

Many of the interviewees stated that the Internet transformed scientific research through the networking of scientists and information. They cited information sharing, online communication, and online collaboration as the main areas where the Internet has impacted their work and made possible massive genome projects such as HapMap. The Internet enables genome projects to move data between global locations and labs in the same building as well as provide open access from anyone interested in using the information (Hwang 2008). The genetic information collected from Nigeria, China, Japan, and Utah is sent to the participating labs located in six different countries. When the sequencing has been performed, that data is uploaded to the central HapMap database in Bethesda, Maryland, where it is maintained, checked for quality control, and stored. Since the project follows an open access model, the data can be downloaded freely by anyone who has broadband Internet. Scientists can feed their own annotations for publication back into the data in the HapMap database. As a HapMap geneticist and medical doctor commented, "I think it is fair to say that the entire concept of genomics, which is really one of data rich studies in biology where you have archival quality data that is comprehensive and is shared freely, is as much about, if not more about, computers and the Internet as it is about DNA technology" (Interview 1001).

The Internet and digital media are not only necessary for moving the information around, but also for structuring the data itself (Bowker and Star 2000; Bowker 2008). Where the databases developed by Hamm and Cameron in the 1980s aimed at storing and analyzing stretches of DNA, the genome database technology of the 1990s and 2000s would build on these innovations by becoming not only spaces of convergence, but also of collaboration and customization. Digital technologies allow for constant updating, cleaning, and translation and enable a "networked multilogue" (Loro 1995, 55) between scientists through the process of sorting and storing data, networking information, and constructing knowledge. Normalizing the data in digital code helps overcome the limitations of "whatever media you're stuck putting the content on…[as]… every time you'd have to move data from one media to another there's an opportunity for error" (Interview 1016).

The Internet and digital media allows for the embedding and hyperlinking of numerous supporting sorts of information. A biostatistician working on the HapMap described how the linking of different type of detailed and annotated sequence information in "a very easily queryable set of databases is an incredible advantage over linear, analog forms of data" (Interview 1002).[10] One of the more powerful features of an electronic database is the cross referencing of information, which resides in different repositories, simultaneously enabling very complex searches to be done on huge amounts of information and complex analysis algorithms to be run easily (Interview 1009). Further, "the data is changing on a daily basis. It is being added to, it is being refined, it's being developed, the interpretations, the mistakes are being corrected and so on" (Interview 1007). Genome databases have an inherent anti-narrative logic to them as the customization and feedback loop features can tell different stories depending on the needs of the individual scientists. In turn, the outcome of a scientist's work on a particular chapter of the 'book of life' can be uploaded back into the database, thus changing the detail and scope of the original in real time. As a result, the same underlying data can have many different representations, based on a hypertext architecture.

While the development of digital technologies has played an important role in the articulation of genomics and information technologies, genome databases and the Internet have also been a space of convergence for political, social, and ethical values of biologists and computer scientists. For example, the value of freely circulating data, which possesses a long tradition in 20th century biomedical sciences, has been enhanced by the Internet and has reciprocally offered new contents to the virtual space. One of the key mandates of the project is that the information contained in the HapMap SNP database will be "freely available in the public domain, at no cost to users".[11] The open access model, a key development in computing science and the "hacker ethic" (Himanen 2001; See also Marturano 2003), represents a movement in the academic and public sector scientific community, particularly in genomics and biomedicine (Kaye et al 2009; Heeney et al in press).

A number of HapMap interviewees referred to opening access as "democratizing the data" and genome data as a public good. Some members see the data coming out of HapMap as being able to overcome global disparities in science and technology: "I think it's an opportunity for the West and the industrialized economies to efficiently transfer the intellectual benefits of wealth and investment and this technology to the

developing world." (Interview 1016). The Internet plays a critical role in providing the communication infrastructure to carry out the mandate of open access. Like other battles over intellectual property and copyright, such as the downloading of music and the Napster case, DNA sequences, particularly the one resulting from the HGP, have been the object of intense debates over open access and data sharing (Sulston and Ferry 2002; Cook-Deegan 1994; See also Goven 2006; Lassen and Jamison 2006; Salter and Salter 2007; Tutton 2007; Zanestoki et al 2006). HapMap members are particularly enthusiastic and principled about this practice (Interview 1016): "…in terms of making science really international and making science open in the humanistic, old sense of science, open as in belonging to the public, I think it's been absolutely tremendous" (Interview 1008).

In the early 1990s, Gilbert warned that the proliferation of databases would create a digital divide: "The next tenfold increase in the amount of information in the databases will divide the world into haves and have-nots, unless each of us connects to that information and learns how to sift through it for the parts we need" (Gilbert, quoted in Lenoir 1999, 18). Gilbert's warnings have partially come true. The HGP and HapMap are selectively global as the two projects included "only developed nations with the technological and economic infrastructure to support bioscience research" (Thacker 2006, 18). Thacker argues "genomics is a selectively global industry, creating a specific map determined by Western science, technology, and government and economic interest" (2006, 18). Even though the Internet provides the flow of data to the public, the digital divide remains a global issue in terms of access to the Internet as well as the quality and capacity of the digital pipes. Globalization scholars argue that everyone does not share technological advancements uniformly. Doreen Massey suggests that we need to pay attention to the "power geometry" of uneven distribution of resources and social inequality that is not ameliorated by globalization, but exacerbated. As the dominant organizing principle in the information age is the network (Castells 2000 [1996]; Newman, Barbási and Watts 2006), power operates through the space of flows. Only certain countries with the technological, scientific, and economic capabilities could become members of HapMap. While HapMap aims to map the molecular level, it is also a map of geo-political relations. For example, Japan and China represent Asia rather than Thailand or Vietnam.

One interview respondent felt that, in the long run, this approach would have a "great and profound impact on the way biomedical science is being done because it's a very infectious idea and it's not an idea that existed in biomedicine before" (Interview 1001). Democratizing the data depends on the network capacity of databases and the Internet as well as a social movement from within the biomedical sciences. It appears to directly confront private models of the biotechnology industry where the keeping of trade secrets in closed labs is considered crucial to competing in the marketplace. Marturano (2003) suggests that scientists should adopt the open source philosophy followed by many computer hackers where the source codes are shared, modified, and redistributed. This could strengthen the scientific community and refocus the emerging patent-and-perish culture to a gift economy where status among peers comes from the sharing of knowledge, which is already part of the practice of scientists. Ultimately, an open source philosophy seeks to protect genomic data as a public good, rather than something that can be owned by a corporation or individual scientist. The immediate release into the public domain approach could have far reaching effects in terms of the divide between information haves and have nots.

Conclusion

This paper has shown that the past and current interactions between biology and information technologies are better understood from the perspective of a mutual interdependence. In so doing, we attempted to engage with the claims of a natural marriage and of biology having become an information science present in some STS literature and accounts by genomic researchers and suggest novel theoretical and empirical directions for STS and communication. Automatic sequencers and especially computer-assisted databases to store DNA sequences have acted as spaces of convergence in which biology and information technologies interact and shape each other. Genomic research is the indivisible result of these interactions, which have operated bi-directionally: sequencing and other biological processes have been modeled on the computer and, at the same time, altered traditional programming algorithms such as those used in text processing.

The interactions between information technologies and the life sciences not only affect biological processes or computing instruments. Values and models such as discovery science, open access or large-interdisciplinary research groups are the results and materializations of this mutual interdependence. Genomics may, thus, be defined as the result of the consolidation and expansion of these bi-directional interactions and a field in which biology and computing are currently indistinguishable. The Internet, as a technology of data sharing and exchange, has also contributed towards this definition and been essential for the development of initiatives such as the Human Genome Project and the HapMap Project. This model of reciprocal interactions between information technologies and biomedical enterprises constitutes a suitable framework for the growing STS scholarship on organization and governance of genomic centers (Balmer 1996; Hilgartner 2004; Gibbons et al 2007; Ramillon 2007). Genomics may, therefore, be considered an information science not just because it incorporates information technologies. In this incorporation, the biomedical areas under research, the research activity itself, and the incorporated technologies are deeply transformed. Because of these transformations genomics is different from other biological uses of computers and databases, thus deserving an independent framework for STS analysis.

This independent framework should take into account all the dimensions of the interactions between biology and computing, and not merely label them as a natural marriage. It requires an interdisciplinary approach, and the historical and socio-cultural reappraisal presented in this paper constitutes an initial step towards this end. Our perspective also has implications for more general STS scholarship on technoscience and society. The bi-directional model of interaction we have described challenges, at the same time, the arguments for a "social shaping" of technology and for a "digital shaping" of biology. Both arguments have a strong implementation in STS literature (e.g. MacKenzie and Wajcman 1999; Moody 2004; Zweiger 2001) and show different forms of determinism in the understanding of the connections between science, technology and society. Our interdisciplinary analysis and proposed model of convergence links with the "co-production of knowledge" that Jasanoff has proposed as a general STS framework (Jasanoff 2004). Biology, computing and social orders interact and are reciprocally shaped around spaces of convergence, but none of them fully determines the sequencer, the database, or other genomic technologies.

References

ABI. 1983-1992. *Annual Reports.* Foster City: Applied Biosystems.

Agar, J. 2006. What difference did computers make? *Social Studies of Science* 36 (6): 869-907.

Balmer, B. 1996. Managing mapping in the Human Genome Project. *Social Studies of Science,* 26: 531-573.

Boczkowski, P., and L. A. Lievrouw. 2007. Bridging STS and Communication Studies Scholarship on Media and Information Technologies. In *The Handbook of Science and Technology Studies*, edited by E. J. Hackett, O. Amsterdamska, M. Lynch and J. Wajcman. Cambridge, MA: MIT Press.

Bowker, G. 2005. *Memory Practices in the Sciences.* Cambridge: MIT.

Bowker, G., and S.L. Star. 1999. *Sorting Things Out: Classification and its Consequences*. Cambridge, MA: MIT Press.

Bromberg, J.L. 1991. *The Laser in America, 1950-1970.* Cambridge: MIT Press.

Bud, R. 1993. *The Uses of Life: A History of Biotechnology.* Cambridge: Cambridge University Press.

Burk, D. L. 2002. Lex genetica: The law and ethics of programming biological code. *Ethics and Information Technology, 4*: 109-121.

Burnett, R., and P.D. Marshall. 2003. *Web Theory*. London: Routledge.

Caltech (1975-1988) *Biology Division Annual Reports.* Pasadena: California Institute of Technology.

Campbell-Kelly, M. 2003. *From Airline Reservations to Sonic the Hedgehog: A History of the Software Industry.* Cambridge: MIT.

Campbell-Kelly, M., and W. Aspray. 1996. *Computer: A History of the Government Machine.* New York: Harper & Collins.

Capra, F. 2002. *The Hidden Connections: Integrating The Biological, Cognitive, And Social Dimensions Of Life Into A Science Of Sustainability*. New York: Doubleday Books.

Castells, M. 2000 [1996]. *The Information Age: Economy Society and Culture.* Malden and Oxford: Blackwell Publishing.

Castells, M. 2001. Informationalism and the network society. In *The Hacker Ethic and the Spirit of the Information Age*, edited by P. Himanen, 155-178. London: Vintage.

Ceruzzi, P. 1998. *A History of Modern Computing.* Cambridge: MIT.

Chow-White, P. A. 2008. The informationalization of race: Communication technologies and the human genome in the digital age. *International Journal of Communication, 2*: 1168-1194.

Collins, F. 2006. The heritage of humanity. *Nature Human Genome,* S1: 9-12.

Cook-Deegan, R. 1994. *The Gene Wars: Science, Politics and the Human Genome.* London and New York: W.W. Norton and Company.

Danna, A., & O.H. Gandy 2002. All that glitters is not gold: Digging beneath the surface of data mining. *Journal of Business Ethics, 40*(4): 373-386.

Date, C.J. 1981 [1975]. *An Introduction to Database Systems.* London: Addison Wesley).

Dayhoff M., and R. Ledley 1962. Comprotein: a computer program to aid primary protein structure determination. *Proceedings in the Fall Joint Computer Conference.* Santa Monica: American Federation of Information Processing Societies.

de Chadarevian, S. 2002. *Designs for Life: Molecular Biology after World War II.* Cambridge: Cambridge University Press.

Dumas J.P. and J. Ninio 1982. Efficient algorithms for folding and comparing nucleic acid sequences. *Nucleic Acids Research*, 10(1).

Fortun, M. 2008. *Promising Genomics: Iceland and deCODE Genetics in a World of Speculation.* Berkeley: University of California Press.
Fujimura, J. 1999. The practices of producing meaning in bioinformatics. In *The Practices of Human Genetics*. Edited by M. Fortun and E. Mendelsohn: 49-87. London: Kluwert.
Fujimura, J., and M. Fortun. 1996. Constructing knowledge across social worlds: the case of DNA sequence databases in molecular biology. In *Naked Science: Anthropological Inquiry into Boundaries, Power, and Knowledge.* Edited by L. Nader: 160-173. London and New York: Routledge.
Galison P., and B. Hevly (eds.). 1992. *Big Science.* Stanford: Stanford University Press.
García-Sancho M. 2007a. The rise and fall of the idea of genetic information (1948-2006). *Genomics, Society and Policy,* 2(3): 125-132. Available open access at http://www.hss.ed.ac.uk/genomics/vol2no3/Garcia-sanchoabstract.htm
----. 2007b. Mapping and sequencing information: the social context for the genomics revolution. *Endeavour,* 31(1): 18-23. Available at http://dx.doi.org/10.1016/j.endeavour.2007.01.006
----. 2008. *Sequencing as a Way of Work: A History of its Emergence and Mechanisation–From Proteins to DNA, 1945-2000.* PhD thesis, Centre for the History of Science, Imperial College, London.
----. 2009. The perception of an information society and the emergence of the first computerized biological databases, 1948-1992. In *Human Genome: Features, Variations and Genetic Disorders.* edited by A. Matsumoto and M. Nakano. New York: Nova Publishers: 257-276.
----. 2010. A new insight into Sanger's development of sequencing: from proteins to DNA, 1943-1977: 265-323
Available at http://dx.doi.org/10.1007/s10739-009-9184-1
----. (in press) "From metaphor to practices: the introduction of 'information engineers' into the first DNA sequence database" in *History and Philosophy of the Life Sciences.*
----. Forthcoming. *A History of Protein and DNA Sequencing. Biology and Computing at the Boundaries (1945-2000).* Basingstoke: Palgrave-Macmillan.
Genbank. 2006. International sequence databases exceed 100 gigabases. Retrieved June 15, 2006, from http://www.ncbi.nlm.nih.gov/Genbank/
Genbank. 2009. Genbank Overview. http://www.ncbi.nlm.nih.gov/Genbank/
Gezelter, D. 1999. Catalyzing Open Source Development in Science: The OpenScience Project. Retrieved March 28, 2005 http://www.openscience.org/talks/bnl/img0.htm
Gibbons, S., J. Kaye, A. Smart, C. Heeney, and M. Parker. 2007. Governing Genetic Databases: Challenges facing Research Regulation and Practice. *Journal of Law & Society*, 34(2): 163-89.
Gilbert, W. 1980. DNA sequencing and gene structure, *Nobel Lecture*, available at www.nobel.se/chemistry/laureates/1980/gilbert-lecture.html
----. 1992. A vision of the grail. In *The Code of Codes: Scientific and Social Issues in the Human Genome Project.* edited by D.J. Kelves and L. Hood. Cambridge: Harvard University Press.
Goven, J. 2006. Processes of inclusion, cultures of calculation, structures of power: Scientific citizenship and the Royal Commission on Genetic Modification *Science, Technology, & Human Values, 31*(5): 565-598.
Guenther, A. H., H. R. Kressel, and W. F. Krupke. 1991. The laser now and in the future. In *The Laser in America, 1950-1970.* edited by J. L. Bromberg. Cambridge: MIT Press.

Haigh, T. 2001. Inventing information systems: the systems men and the computer, 1950-1968. *Business History Review,* 75(1).
----. 2006a. A veritable bucket of facts: origins of the data base management system. *SIGMOD Record,* 35(2).
----. 2006b. Remembering the office of the future: the origins of word processing and office automation. *Annals of the History of Computing,* 28(4).
Hamm, G., and G. Cameron. 1986. The EMBL data library. *Nucleic Acids Research,* 14(1).
Hamm, G., and K. Stübert. 1982. EMBL Nucleotide Sequence Data Library. *Nucleotide Sequence Data Library News,* 1.
Haraway, D. 1997. *Modest_Witness@Second_Millennium.FemaleMan©_Meets_ OncoMouse™.* New York and London: Routledge.
Hayles, N. K. 1999. *How We Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics.* Chicago and London: Chicago University Press.
Heeney, C.; N. Hawkins; J. de Vries; P. Boddington and J. Kaye. In press. Assessing the privacy risks of data sharing in genomics. *Public Health Genomics.* Available at http://dx.doi.org/10.1159/000294150
Held, D., & A. G. McGrew (Eds.). 2000. *The Global Transformations Reader: An Introduction to the Globalization Debate*. Malden, MA: Polity Press.
Hilgartner, S. 2004. Making maps and making social order: governing American genome centers, 1988-93. In *From Molecular Genetics to Genomics: The Mapping Cultures of Twentieth Century Genetics.* edited by J.P. Gaudillière and H.J. Rheinberger: 113-128. London and New York: Routledge.
Himanen, P. (ed.). 2001. *The Hacker Ethic and the Spirit of the Information Age*. London: Vintage.
Hine, C. 2006. Databases as scientific instruments and their role in the ordering of scientific work. *Social Studies of Science,* 36(2): 269-298.
Holdsworth, D. 1999. The ethics of the 21st century bioinformatics: Ethical implications of the vanishing distinction between biological information and other information. In *Genetic Information. Acquisition, Access and Control*. edited by A. K. Thompson & R. Chadwick. New York: Kluwer/Plenum.
Hood, L. 1992. Biology and medicine in the twenty first century. In *The Code of Codes: Scientific and Social Issues in the Human Genome Project.* edited by D. J. Kevles and L. Hood. Cambridge: Harvard University Press.
----. 2001. Under Biology's Hood. *Technology Review.* September.
Hwang, K. 2008. International collaboration in multilayered center-periphery in the globalization of science and technology. *Science, Technology, & Human Values, 33*(1): 101-133.
Jasanoff, S. (ed.). 2004. *States of Knowledge: The Co-Production of Science and Social Order.* London and New York: Routledge.
Jenkins, H. (2006). *Convergence Culture: Where Old and New Media Collide*. New York: NYU Press.
Jordan, K., & Lynch, M. 1998. The Dissemination, Standardization and Routinization of a Molecular Biological Technique. *Social Studies of Science, 28*(5/6), 773-800.
----. 2000. *Who Wrote the Book of Life: A History of the Genetic Code.* Stanford: Stanford University Press.
Kaye, J.; C. Heeney, N. Hawkins, J. de Vries, P. Boddington. 2009. Data Sharing in Genomics - Re-shaping Scientific Practice. *Nature Review Genetics,* 10(5): 331-335.
Keating, P., C. Limoges, and A. Cambrosio. 1999. The automated laboratory: the

generation and replication of work in molecular genetics. In *The Practices of Human Genetics.* edited by M. Fortun and E. Mendelsohn New York: Kluwer.
Kennard, O. 1997. From private data to public knowledge. In *The Impact of Electronic Publishing on the Academic Community,* International Workshop Organised by the Academia Europaea and the Wenner-Gren Foundation. Available on-line at www.portlandpress.com/pp/books/online/tiepac/session6/ch2.htm
Kenney, M. 1986. *Biotechnology: The University-Industrial Complex.* New Haven: Yale University Press.
Kleinman, D. L. 2003. *Impure Cultures: University Biology and the World of Commerce*. Madison, WI: University of Wisconsin Press.
Kleinman, D. L., & Vallas, S. P. (2001). Science, capitalism, and the rise of the "knowledge worker": The changing structure of knowledge production in the United States. *Theory and Society, 30*, 451-492.
Kleinman, D. L., & Vallas, S. P. (2006). Contradiction in convergence: Universities and industry in the biotechnology field. In S. Frickel & K. Moore (Eds.), *New Political Sociology of Science: Institutions, Networks, and Power*. Madison, WI: University of Wisconsin Press.
Kline, R. 2006. Cybernetics, management science and technology policy. *Technology and Culture,* 47.
----. 2000. Learning about information technologies and social change: the contribution of social informatics. *The Information Society,* 16(3): 217-232.
Lassen, J., & Jamison, A. 2006. Genetic Technologies Meet the Public: The Discourses of Concern. *Science Technology Human Values, 31*(1): 8-28.
Lécuyer, C. 2006. *Making Silicon Valley: Innovation and the Growth of High Tech, 1930-1970.* Cambridge: MIT Press.
Lenoir, T. 1999. Shaping biomedicine as an information science. In *Proceedings of the 1998 Conference on the History and Heritage of the Science Information Systems.*edited by M. E. Bowden, T. B. Hahn, and R.V. Williams: 27-45. Medford: ASIS.
----. (ed.) 2002a. *Makeover: writing the body into the posthuman technoscape. Part one: Embracing the posthuman*. Special issue of *Configurations,* 10(2).
----. (ed.) 2002b. *Makeover: writing the body into the posthuman technoscape. Part two: Corporeal axiomatics*. Special issue of *Configurations,* 10(3).
Lenoir, T. and Hays, M. 2000. The Manhattan Project for biomedicine. In P.R. Sloan (ed.) *Controlling our Destinies: Historical, Philosophical, Ethical and Theological Perspectives on the Human Genome Project.* Indiana: University of Notre Dame.
Leonelli (2010) "Documenting the emergence of bio-ontologies: or, why researching bioinformatics requires HPSSB" in *History and Philosophy of the Life Sciences,* 32(1): 105-125.
Loro, L. 1995. Everyone's talkin' in the 'multilogue'. *Advertising Age, 66*(35): 28.
Lynch M., Cole S.A., McNally R. and Jordan K. 2008. *Truth Machine: The Contentious History of DNA Fingerprinting.* Chicago: Chicago University Press: ch.3.
Lyon, D. 2005. The sociology of information. In *The Sage Handbook of Sociology*. Edited by C. Calhoun, C. Rojec, and B. Turner London: Sage.
Mackenzie, A. 2003. Bringing sequences to life: How bioinformatics corporealizes sequence data. *New Genetics and Society, 22*(3): 315-332.
Marturano, A. 2003. Molecular biologists as hackers of human data: Rethinking IRP for bioinformatics research. *Journal of Information, Communication and ethics in society, 1*(4): 207-216.

Maxam A., and W. Gilbert. 1977. A new method for sequencing DNA. *Proceedings of the National Academy of Sciences*, 74.
McCallum and Smith 1977. Computer processing of DNA sequence data. *Journal of Molecular Biology*, 116: 29-30.
McKenzie, D., and J. Wajcman (eds.). 1999 *The Social Shaping of Technology.* London: Open University Press.
McKusick, V., and F. Ruddle. 1987. Editorial: a new discipline, a new name, a new journal. *Genomics*, 1.
Mendelsohn. E. 1992. The social locus of scientific instruments. In *Invisible Connections: Instruments, Institutions and Science.* Edited by R. Bud et al. London: SPIE Optical Engineering Press.
Mindell, D. 2002. *Between Human and Machine: Feedback, Control and Computing before Cybernetics.* Baltimore: Johns Hopkins University Press.
Moody, G. 2004. *Digital Code of Life: How Bioinformatics is Revolutionizing, Science, Medicine, and Business*. Hoboken, NJ: John Wiley & Sons, Inc.
Needleman, S., and D. Wunsch. 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology*, 443-53.
Newman, M., A. L. Barbási, and D. Watts. 2006. *The Structure and Dynamics of Networks.* New Jersey: Princeton.
November J. (2004) "LINC: biology's revolutionary little computer" in *Endeavour,* 28(3): 125-31.
November J. 2006. *Digitizing Life: The Introduction of Computers to Biology and Medicine.* PhD dissertation, Department of History of Science, Princeton University.
Penders, B., K. Horstman, and R. Vos. 2008. Walking the line between lab and computation: the 'moist' zone". *BioScience,* 58(8).
Pool, I. d. S. (1983). *Technologies of Freedom*. Cambridge, MA: Belknap Press.
Powell, A., M. O'Malley, S. Müller Wille, J. Calvert, and J. Dupré. 2007. Disciplinary baptisms: a comparison of the naming stories of genetics, molecular biology, genomics and systems biology. *History and Philosophy of the Life Sciences,* 29(1): 5-32.
Ramillon V. 2007. *Le deux génomiques. Mobiliser, organiser, produire: du séquençage à la mesure de l'expression des gènes*. PhD dissertation, École des Hautes Études en Sciences Sociales.
Rose, S. 1997. *Lifelines: Life Beyond the Gene*. London: Penguin.
Salter, B., & Salter, C. 2007. Bioethics and the global moral economy: The cultural politics of human embryonic stem cell science *Science, Technology, & Human Values, 32*(5): 554-581.
Sanger, F. 1975. The Croonian Lecture. *Proceedings of the Royal Society of London*, 191, B series.
----. 1988. Sequences, sequences and sequences. *Annual Review of Biochemistry*, 57.
Sanger F. and A. Coulson (1975) A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Journal of Molecular Biology,* 94: 441-448.
Sanger, F., S. Nicklen, and A. Coulson. 1977. DNA sequencing with chain terminating inhibitors. *Proceedings of the National Academy of Sciences*, 74.
Sanger, F., G. M. Air, B. Barrell, N. L. Brown, A. Coulson, J. C. Fiddes, C. A. Hutchison III, P. M. Slocombe, and M. Smith. 1977. Nucleotide sequence of bacteriophage ØX-174. *Nature,* 265: 687-695.

Sanger, F., A. Coulson, G. F. Hong, D. F. Hill, and G. B. Petersen. 1982. Nucleotide sequence of bacteriophage λ DNA. *Journal of Molecular Biology*, 162(4): 729-773.

Sanger, F., and M. Dowding (eds.). 1996. *Selected Papers of Frederick Sanger (with Commentaries).* London: World Scientific.

Sassen, S. 1998. *Globalization and its Discontents: Essays on the New Mobility of People and Money*. New York: The New Press.

Shreeve, J. 2004. *The Genome War: How Craig Venter Tried to Capture the Code of Life and Save the World*. New York: Alfred A. Knopf.

Smith, T. 1990. The history of the genetic sequence databases. *Genomics*, 6.

Smith, L.M. (2008) The development of automated DNA sequencing. Unpublished paper.

Smith, L.M., H. M. McConnell, A. Smith Baron, and J. W. Parce. 1981. Pattern photobleaching of fluorescent lipid vesicles using polarized laser light. *Biophysics Journal*, 33.

Smith, L.M., J. Sanders, R. J. Kaiser, P. Hughes, C. Dodd, C. R. Connell, C. Heiner, S. Kent, and L. Hood. 1986. Fluorescence detection in automated DNA sequence analysis. *Nature*, 321: 674-679.

Staden, R. 1982. Automation of the computer handling of gel reading data produced by the shotgun method of DNA sequencing. *Nucleic Acids Research*, 10(15).

----. 1984. A computer program to enter gel reading data into a computer. *Nucleic Acids Research*, 12(1).

Strasser, B. 2006. Collecting and experimenting: the moral economies of biological research, 1960s-1980s. In *History and Epistemology of Molecular Biology and Beyond: Problems and Perspectives.* Edited by H. J. Rheinberger and S. de Chadarevian. Berlin: Max Planck Institute for the History of Science, preprint number 310.

----. 2008. Genbank: natural history in the 21st century? *Science,* 322: 537-38.

----. 2009. Collecting, comparing, and computing sequences: the making of Margaret O. Dayhoff's 'Atlas of Protein Sequence and Structure', 1954–1965. *Journal of the History of Biology.* Available at http://dx.doi.org/10.1007/s10739-009-9221-0

Suárez-Díaz E. (2010) "Making room for new faces: evolution, genomics and the growth of bioinformatics" in *History and Philosophy of the Life Sciences,* 32: 65-90.

Suárez-Díaz E. and Anaya-Muñoz V. (2008) "History, objectivity and the construction of molecular phylogenies" in *Studies in History and Philosophy of Biological and Biomedical Sciences,* 39 : 451-468.

Sulston, J., and G. Ferry. 2002. *The Common Thread: A Story of Science, Politics, Ethics and the Humane Genome*. London: Bantam.

Sulston, J., and F. Mallet, R. Staden, R. Durbin, T. Horsnell, and A. Coulson. 1988. Software for genome mapping by fingerprinting techniques. *Computer Applications in the Biosciences,* 4(1): 125-132.

Thacker, E. 2004. *Biomedia*. Minneapolis, MN: University of Minnesota Press.

Thacker, E. 2006. *The Global Genome: Biotechnology, Politics, and Culture*. Cambridge, MA: The MIT Press.

Tutton, R. 2007. Constructing participation in genetic databases. *Science, Technology and Human Values,* 32(2): 172-195.

Waldrop, M. 2001. Data Mining. *Technology Review, January/February*.

Waters, M. 2001. *Globalization* (2nd ed.). New York: Routledge.

Zavestoski, S., Shulman, S., & Schlosberg, D. 2006. Democracy and the environment on the internet: Electronic citizen participation in regulatory rulemaking *Science, Technology, & Human Values, 31*(4): 383-408.

Zweiger, G. 2001. *Transducing the Genome: Information, Anarchy, and Revolution in the Biomedical Sciences.* New York: McGraw Hill.
Zuboff, S. 1988. *In the Age of the Smart Machine: The Future of Work and Power*. New York: Basic Books.

Appendix A

List of Interviews*

Graham Cameron	Member of the EMBL database team during the 1980s
Greg Hamm	Leader of the EMBL database team during the 1980s
Tim Hunkapiller	Member of Leroy Hood's group at the California Institute of Technology during the 1980s
Lloyd M. Smith	Member of Leroy Hood's group at the California Institute of Technology during the 1980s
André Marion	Co-founder of Applied Biosystems and manager of the company during the 1980s
Interview 1001	Population Geneticist
Interview 1002	Biostatistician
Interview 1003	Project Manager
Interview 1005	Bioethicist
Interview 1007	Director of NGO
Interview 1008	Lawyer
Interview 1009	Bioinformatician
Interview 1010	Population Geneticist
Interview 1011	Bioethicist
Interview 1013	Geneticist
Interview 1014	Human Geneticist
Interview 1016	Population Geneticist
Interview 1017	Bioethicist
Interview 2001	Biologist/Senior Scientist, leading global biotechnology company

* Miguel Garcia-Sancho conducted the named interviews between 2006 and 2007. Peter Chow-White conducted the numbered interviews with members of the International HapMap Project (except Interview 2001) in 2005 and 2006.

[1] The authorship is shared equally by Peter Chow-White and Miguel Garcia-Sancho and their names are listed in alphabetical order. Any correspondence regarding this paper can be addressed to both authors.
[2] http://hapmap.ncbi.nlm.nih.gov
[3] However, literature in the history and the philosophy of biology has shown that this association of the genetic material with the concepts of code and program has been more problematic and started in the late 1940s, much before the emergence of DNA sequencing and the personal computer (Kay 2000; Sarkar 1996; Fox Keller 1995; Moss 2004). Our paper will build on this problematization and extend it to the 1970s and 1980s.
[4] In his scholarship on information society, Manuel Castells considers genomics and the recombinant DNA techniques as symptomatic of this new social configuration, but does not draw further on the nature and implications of this socio-technical correspondence (Castells 2000 [1996], 54-59; Castells 2001).
[5] By analyzing different case studies, Soraya de Chadarevian, Joseph November and Sabina Leonelli have shown that different biological disciplines such as X-ray crystallography, biochemistry, molecular biology, physiology or plant genetics decisively shaped the design of biocomputing software and hardware, as well as databases (de Chadarevian 2002, ch. 4; November, 2004, 2006; Leonelli 2010). De Chadarevian, together with Lenoir, has placed the origins of the interactions between biology and computing in the late 1940s. This links with the claim of an "early information society" by historians of computing, who argue that the social concern with data processing much preceded the emergence of the personal computer and of the computer itself (Agar 2003; Black et al. 2007).
[6] Sequencing software and databases existed from the early 1960s, in the field of proteins and before the emergence of commercial text processors. This paper will exclusively deal with DNA sequencing applications, since previous devices are well investigated in the literature (see Strasser, 2009; Suárez-Díaz, 2010; Suárez-Díaz and Anaya-Muñoz, 2008).
[7] The generalization of string processing was a consequence of the gradual shift in the use of the computer from a mathematical calculator to an information processing device (Campbell-Kelly and Aspray 1996, 105 and ff.). This process was largely fostered by the emergence of the personal computer and word processing software between the late 1970s and 1980s (Ceruzzi 1998, chs. 7-9; Haigh 2006b).
[8] The reservations towards computer-minded staff which characterised the 1960s and 70s in biological institutions did not disappear automatically. Hamm and Cameron recall being regarded as "secretariat" during their early years at the EMBL and not being treated as equals until the success of the database (Interview with Hamm; Interview with Cameron).
[9] Diffuse team structures, as Jamie Lewis has shown, create a "virtual space of cooperation" that sometimes prevents interactions between researchers based in closer locations (Lewis 2010).
[10] Soraya de Chadarevian has suggested that genomic databases, in which you first access the genetic map of chromosomes, click a particular section and enter the physical map of DNA fragments and then the nucleotide sequence of a concrete fragment resembles the hypertext structure of the World Wide Web (de Chadarevian 2004, 95).
[11] http://hapmapncbi.nlm.nih.gov/cgi-perl/registration