BIOINFORMATICS:

A technology assessment of recent developments in bioinformatics and related areas of research and development including high-throughput screening and combinatorial chemistry.

Final Report for the *Science and Technological Options Assessment* (STOA) Unit, European Parliament.

Science and Technology Studies Unit (SATSU), Anglia Polytechnic University, Cambridge, United Kingdom.

Dr Nik Brown
Dr Annemiek Nelis
Dr Brian Rappert
Prof. Andrew Webster

Department of Human Genetics, Sylvius Laboratories, Leiden University Medical Center, Leiden, The Netherlands.

Prof. G. J. B. van Ommen

May 1999

Order of the Report:

Abstract
Options Brief
Executive Summary

Section One - Introducing Bioinformatics
- 1.1 Current Technological Developments
- 1.2 Key Actors

Section Two - Commercial and Practical Constraints
- 2.1 Training, Personnel and Disciplinarity
- 2.2 Scale, Size and Capacity
- 2.3 BI, Drug Innovation and Regulation
- 2.4 Software and Hardware Interoperability
- 2.5 Public / Private Alignment and Access to Data

Section Three - Changes in Health Care Delivery
- 3.1 Genetic Testing
- 3.2 Drug developments
- 3.3 Disease classification
- 3.4 Disease classification
- 3.5 Recruitment & Training in Medical Practice
- 3.6 Domestic Healthcare factors
- 3.7 Cost of Health Care Delivery

Section Four - Non-Clinical Products and Services
- 4.1 Bioinformatics and the Insurance Industry
- 4.2 Bioinformatics and Financial Services
- 4.3 Forensic Bioinformatics and Law Enforcement
- 4.5 Bioinformatics and Defence

Section Five: Public Acceptability

Section Six: Conclusion

Annex A: Technological Developments in Bioinformatics
Annex B: Key Actors
Annex C: Regulatory and Advisory Measures
Annex D: Study's Research Instruments

Abstract

In recent years, new gene science has become probably the most information and automation intensive activity in modern research and clinical innovation. In particular, gene sequence and functional analysis is now fundamentally dependent upon the global production, circulation and consumption of huge amounts of data. The exchanges between computational and biological sciences are both far reaching and reciprocal. On the one hand, masses of genetic information are being translated from their 'wet platform' onto the 'dry platforms' of silicon based databases. On the other hand, silicon is now becoming the basis for conducting 'wet' biological and chemical research using genechips and labchips.

However, the interfaces between life science research, clinical innovation and computational science are fraught with problems for policy makers. For example, with what consequences does genetic data become property; how is data-access controlled and distributed; who will benefit and who will be excluded from potential dividends; how will Europe's life sciences adapt to the rising access costs to modern biological innovation; how might it be possible to create seamless integration across Europe's bioinformatic resources; what are the difficulties in bringing biological and computational skills together in innovative combinations; how will the Parliament prepare for new therapeutic and diagnostic innovations; how will quality and safety be maintained?

All of these questions are addressed in this report beginning with a brief introduction to new developments in bioinformatics and the key actors involved. Section Two discusses some of the main technical, organisational and market barriers which inhibit actors from fully exploiting opportunities in the area. Section Three offers an assessment of the likely impact of bioinformatic-related technologies on healthcare. These impacts are then discussed in the context of non-clinical sectors like financial and forensic services. In each case, options for policy are offered.

Options Brief

The skills deficit: The future potential of Europe's buoyancy in the new biological industries depends upon the availability of a skills base in which computational and life science expertise is considerably more integrated. The Parliament may consider making greater provision for these realignments by encouraging new interdisciplinary University-based initiatives. Greater integration across the skills base may be improved by the Parliament putting in place grant-related incentives which reward increased formal relationships between different disciplinary research partners.

Adapting to changes in scale: In both clinical delivery and wider areas of R&D, new developments in bioinformatics entail fundamental changes in scale and capacity. If Europe's life science constituencies are to adapt to an emerging 'big science' environment, modification of the structure and organisation of research provision and healthcare delivery need to be considered. While maintaining the current stimulation, at local level, of transnationally interactive small-scale innovation, the Parliament should consider using the grant mechanism to encourage research partners to develop shared technological resources. Such initiatives have been common in other areas of resource intensive research like physics for example. In the context of healthcare delivery, since so many new bioinformatic resources have proved to be prohibitively expensive, the Parliament may seek to concentrate expertise in one or more centres of expertise throughout the Union. Quality accreditation may be one such method for the identification of such centres.

Improving integration: The report identifies a number of initiatives which have been useful in improving the level of integration between different operational platforms. The Parliament may consider making more formal arrangements to oblige research communities to adhere to standard protocols and specifications. In addressing problems whereby research actors use inconsistent biological terms of reference, the Human Genome Organisation's (HUGO) Nomenclature Committee has been particularly important creating some degree of consensus and standardisation. In respect to standardising operational platforms, the Object Management Group's Life Sciences Working Group has been instrumental in disseminating the use of the Common Object Request Broker Architecture (CORBA).

Intellectual Property and Copyright: The capacity of the EU to exploit proprietary opportunities in new genetics is identified as deficient within Europe, particularly in relation to the activity of SMEs. The Parliament should make further provision to create the conditions whereby public research actors and SME start-up firms are encouraged to exploit copyright and patent protection but within the context an international and mutual commitment to the free circulation of genetic data.

The Commercialisation of Public Health Genetic Registers: The report identifies a number of developments in which concern has been expressed in relation to public health-related genetic data being made available to private biotechnology and pharmacogenomics companies. It is a recommendation of this report that the Parliament continue to monitor such developments particularly in respect to assuring EU constituents that appropriate levels of confidentiality are being maintained. Such measures must also take into account potential and future uses which were unknown at the time consent was given.

Population Screening: The Parliament should consider a much more active role in encouraging and monitoring the integration of new genetic services into existing clinical professions. Genetic counselling services, for instance, need to be made available to the EU's constituents on an equitable basis between Member States. This may have to be done by ensuring that a minimum baseline of provision of diagnostic facilities are identified and made available on a trans-European basis. At the same time, new diagnostic capabilities need to be deployed in a way which takes into account local variations in healthcare delivery, cultural acceptance and available therapeutic expertise.

Non-clinical Products and Services: To varying degrees, bioinformatics has now been integrated into the future developments of a number of non-clinical sectors requiring policy responses from national and supranational government. In respect to forensic services, this report recommends that, in order to maintain confidence in the safety of convictions secured by the submission of DNA evidence, forensic laboratories are obliged to adhere to standard guidelines, as produced by the (US) National Research Council (Committee on DNA Forensic Evidence). The Parliament may also consider conducting further consultative work on the readiness of the EU data infrastructure for the increased use of criminal genetic registers and assess public and political desirability for such developments.

In relation to the use of bioinformatics in the insurance industry, the Parliament is encouraged to continue to closely monitor existing arrangements for self regulation. This approach takes into account the relative adequacy of social healthcare in most EU countries in comparison the US where regulation is somewhat justifiably more structured. Interim self regulation also reflects certain disincentives, such as expense, which has so far deterred the insurance sector from wholesale use of genetic diagnostic capabilities. Where insurers intend to use genetic data, careful attention needs to address their ability to properly interpret such data.

The Parliament may also consider petitioning to strengthen the Biological and Toxic Weapons convention in order to assure ethnic populations that policy makers are taking adequate precautionary measures against the potential of genomically targeted biological weapons.

Public Acceptance: The report argues that mistrust in bioinformatic developments

is not necessarily a consequence of a technically naïve public, but arises from questions raised in relation to the social and commercial motives of research institutions and their regulators. Whilst better technical education of the public may be welcomed by public constituencies and pressures groups, this will have little bearing on the way in which public constituencies form views about the desirability of bioinformatic developments. It is a recommendation of this report, that in all those areas identified above, the Parliament continues to use regulation as a mechanism to integrate science with the social contexts into which future developments will be deployed. This will depend upon both the effective institutional transparency of regulation and its responsiveness to social contexts of use.

The report concludes with contrasting scenarios characterising action on bioinformatics as more, or less, centrally co-ordinated. Each scenario produces distinct implications for action and depends on the deployment of distinct organisational and material resources. All of the issues discussed above would be affected by Parliamentary action that favoured one rather than the other of these scenarios.

Executive Summary

1. The European Parliament is aware that bioinformatics might have a major impact on clinical and non-clinical fields and provides new opportunities for the European Union's research agenda. This Report describes and assesses the current and future developments in this emergent field and identifies various areas of concern and outlines policy options for consideration by the Science and Technology Options Unit of the European Parliament. The issues and concerns thrown up by bioinformatics suggest that there will be a need to adopt a flexible, multi-dimensional approach towards the fostering, management and exploitation of the field, at both European and member state levels.

1.1 The report begins with a review of the current technological developments found in public and private sector actors, notably genomics and sequencing, high throughput screening, combinatorial chemistry and 'chip' based technologies, along with the infrastructural requirements (such as accessible databases and interoperable systems) on which bioinformatics R&D depend. It identifies a number of actual and potential difficulties here, especially in relation to securing interoperability between systems and databases. Key bioinformatic actors may be grouped according to whether they are research institutions and alliances providing access to data, whether they are involved in the provision, organisation and distribution of funding arrangements, or whether, like the pharmaceutical industry, they are purchasers of bioinformatic data services.

2. Developments in bioinformatics are emerging in relation to a number of key practical and commercial constraints which, in turn, will determine access to bioinformatic data and present certain limits to possible social benefits. Practical constraints relate to a scarcity of trained personnel likely to become more pronounced because of a rapid growth in demand for bioinformaticians, while the substantive content provided through current training in universities often lags behind actual industrial needs. Public research centres in the area still provide the majority of new staff for firms. In light of this, ***policy should consider greater provision of training which integrates biological and information sciences since both public and private sectors rely on disciplinary competencies being combined in entirely new ways. Incentives should be developed for universities to develop new bioinformatics curricula and break down disciplinary distinctions between departments.***

2.1 As the life sciences adopt the characteristics of 'big science' the entry cost of genomics research increases dramatically, especially for smaller actors, such as SMEs, and even for public sector organisations needing to invest in complex, expensive equipment. The results of this Report suggest that the primary economic impact of bioinformatics lies in its potential to reduce the costs of drug research upstream - i.e. discovery costs - rather than in reducing downstream trailing and related regulatory costs. In the context of health delivery, the ability of health providers to exploit new technical capacities will depend on new arrangements for

resource sharing and specialisation. If the relative position of public to commercial research and health provision is not to deteriorate, ***the European Union is advised to exploit its international scale and capacity in the formation of public health delivery services which traverse member state boundaries. These shifts in scale will be increasingly necessary as bioinformatic developments increasingly take on the attributes of 'big science'.***

Consistent with our attribution of differences in BI-competencies to size and scale it is relatively small public research and diagnostic facilities (such as public health laboratories) who are most excluded from new BI-related advantages. High costs are concentrated particularly in: access to skilled bioinformaticians as discussed above; institutional commitment to high cost investments in new gene sequence and array equipment; new data platforms for storage and analysis. Most respondents, considered that if left unattended, the relative position of public to commercial research would deteriorate.

2.2 Additional costs related to interoperability remain despite moves toward standardisation being fostered by relevant organisations and agencies. The temptation to avoid integration costs by going it alone will merely add to interoperability risks. Interoperability of bioinformatics data is crucial for ensuring that clinical benefits are realised. The problems are both inter- and intra-organisational and relate to standardisation, co-ordination and communication.

While the responsibility for solving this problem might best be left with the constituencies involved, *t**he EU could support efforts towards improved interoperability through programmes facilitating networks and fostering hardware and software development, especially with regard to ways in which a common platform could be put into place.*** Harmonisation across secure internet technologies is considered to be a key near term necessity by all actors as is a consensus on intranet and web-based media and the ongoing standards programme.

Interoperability relates to systems for exchanging information, an issue which the Report also considers in relation to copyright and related intellectual property rights over bioinformatics based data, such as sequences an genetic markers. While it is clear that open access to information is highly valued by the European (and other) research communities, it is suggested that ***the European Union should consider making further provision for copyright and patent protection.***

A recent development related to questions of access to and control over information is the construction and use by pharmaceutical firms of 'genetic registers' - such as has been recently agreed between the Iceland authorities and Hoffman La Roche. ***The Report suggests that concern over the use of public health genetic registers by commercial companies may well increase as a consequence of the Icelandic case. In order to better anticipate the wider incidence of these kinds of arrangements, the EU should closely monitor public and professional responses to these and similar developments.***

A final issue within the broad area of public/private relations and access to data relates to differences in funding provision between the two. Most public sector respondents believe there to be a funding bias in favour of the private sector that profits from the availability of publicly funded data, often crucial to the development of their commercial products. In light of this, ***the EU should consider whether the dependence of the private sector upon public institutions and the perceived funding bias this creates, should be addressed in future funding arrangements***.

3. Bioinformatics developments in areas such as high throughput screening and lead drug design are likely to have a significant impact on perceptions of health at the level of the patient, the determination and allocation of health resources, and requirements for counselling and - paralleling the need noted above for more research-related training - training of clinical and associated personnel. Changing perceptions of health risks, new disease classifications and the management of new health demands from patient advocacy groups along with the development of targeted drugs for correcting genetic disease, will pose a major challenge for clinical genetics.
Continuing efforts need to be made with regard to genetic counselling in European to ensure the ethical implications of human genome analysis for clinical practice are adequately handled. In addition, the EU needs to ensure that when large scale screening programmes are undertaken by members states, this happens in a manner which is sensitive to the indeterminacies of such activities.

However, the pattern of demand, the take-up of new bioinformatics-derived products and the depth of the clinical genetics skills base are likely to be very uneven across Europe. The market for clinical products will be shaped by factors at a national level which may work against the deployment of certain techniques. The agencies within European countries that are responsible for genetics (e.g., public health) are therefore unlikely to respond in an identical way to the available bioinformatics-derived techniques. ***The European Parliament may need to accommodate within its bioinformatics provisions considerable variation between public health regimes across member states. Attempts to foster European-wide health policy based solely on assumptions of clinical benefit may meet with limited success. Nevertheless, support for a broadening of the expertise base in both clinical genetics and counselling is required, just to cope with present levels of demand.***

4. Bioinformatics has given rise to new concerns about the relationship between insurance and the insured with regard to a mutually acceptable balance of risk between the two parties. There is some concern that the insurance sector is insufficiently prepared to make safe and reliable estimates of future risk based on an informed interpretation of genetic data.

Confidentiality considerations and the protection of data from third parties is a significant area of concern and will become more so in the future. The EU is

recommended to subject the use of genetic information to proper codes of practice and systems of financial redress to prevent non-anonymised genetic information on individuals from circulating between clinical and non-clinical sectors.

4.1 Bioinformatics has an impact beyond the clinical sector to include financial services, law enforcement (forensic bioinformatics) and military defence systems. Financial services and law enforcement sectors are interested in bioinformatics as the basis for automated recognition systems based on 'biological signatures' (such as retina pattern scans), and typically reliant on genechip technologies. However, the Report shows that there is considerable difficulty in developing a stable gene-based identity verification system, while their use in relation to areas such as DNA fingerprinting raise a number of concerns explored in the Report. ***The European Parliament should consider the commensurability between domestic and international legislation governing access to health records, since this will effect the development and use of gene-based bioinformatic systems. Acute technical problems in the use of gene-based biometric tools will, however, slow down the development of these systems, providing a policy-opportunity gap for further action at the European level, especially with regard to an assessment of existing legislation, such as the Directive on Data Protection, and the Directive on the protection of Personal Data (96/46/EC).***

Increased understanding of genetic variation raises concerns that biological weapons can be developed which either enhance existing biological weapons or target particular ethnic groups or individuals. While the sequence of human DNA does not vary significantly according to racial or ethnic classifications, what variations do exist cannot be excluded from ultimately providing a basis for weapons which are targeted on the biology of specific groups. ***The European Parliament might consider the option of supporting the call made by some groups (such as the British Medical Association in the UK) for a strengthening of the Biological and Toxin Weapons Convention, including the establishment of appropriate verification provisions, in order to minimise the possibility of a new class of biological weapons being developed.***

5. Finally, this Report examines the public acceptability of bioinformatics and notes that constituencies were sharply divided on whether or not public acceptability and ethical considerations would constrain to the clinical uptake of bioinformatic-derived products and services. However, attempts merely to promote greater public literacy of the science of bioinformatics is unlikely to produce more positive attitudes towards the field. ***An alternative to literacy-oriented education of facts would see the European Union encouraging its various science-based agencies to foster a more cautious approach towards the field, acknowledging the uncertainties that lie therein. This would include emphasising science as a collective enterprise, the limits of expertise, the contingency of scientific judgement, and the role of trust. It could be argued this would be more valuable in empowering citizens to form opinions on developments in bioinformatics.***

Section One - Introducing Bioinformatics

1.1 Current Technological Developments

Bioinformatics refers to a number of highly interrelated activities each of which is introduced below. Given the complexity of this discussion, for the convenience of the reader a more detailed description is included in technical Annex A.

Genomics and Sequencing: Genomic research involves locating specific genes which code for proteins within an organism's genome. Two distinguishable methods should be mentioned here since each have different implications for the research institutions using them. In the first place, 'clone by clone' sequencing methods involve copying short strings DNA repeatedly within a replicating organism until the sequences within the string have been mapped. This has been the dominant method used until recently by most public research institutions involved in the Human Genome Project. On the other hand, 'whole shotgun sequencing', fractures the DNA of an organism into small pieces and then proceeds by matching the bases at the end of each fragment with one another. This method has recently been pursued by the commercial company, Celera. To date, 20-30 genomes have been sequenced including micropathogens and small organisms with several groups competing to provide a sequence map of the human genome.

High Throughput Screening and Combinatorial Chemistry: *High Throughput Screening* (HTS) refers to approaches by which automated technologies are used to test compounds against gene targets which, in turn, have been identified through genomics. Combinatorial chemistry refers to the means by which vast libraries of chemical compounds are produced for the purposes of HTS. Both activities depend upon a high degree of robotic automation and on-line search capacity.

New Lead/Rational Discovery Drug Design: All of the developments in bioinformatics reviewed here are becoming more central to the emerging rationale by which the pharmaceutical sector produces drugs. Traditional drug design more usually began with a promising compound which would then be tested against physiological responses in animal models. Lead drug design reverses this rationale by starting with the clinical lead, DNA mined from databases, against which thousands of compounds are then tested for potentially therapeutic effects.

Biochips, Genechips and DNA Arrays: The miniaturisation of bioinformatic technologies offers substantial opportunities in terms of the acceleration and efficiency of the means by which the presence or absence of target genes in an organism can be determined. Instead of being coated in micro-processors, thousands of miniature wells are etched into genechips each containing DNA molecules which will snag corresponding genes in a test sample. Instead of being limited to a small number of genetic markers, a single genechip diagnostic is able to test for complex multifactoral conditions or for carriers of a whole range of relatively infrequent monogenetic diseases much more easily than before.

Time-of-Fight Mass Spectrometry: Another method for identifying DNA lies in using spectrometric methods whereby DNA ions are released into a 'flight tube' which measures their 'time-of-flight'. Larger DNA ions have a lower velocity than smaller ones and are slower at reaching the detector at the end of the flight. The process is extremely rapid, taking only nanoseconds to analyse each sample, thousands of which can be mounted on sample plate. High speed computational electronics and processors are needed to keep pace with the typing of each sample. The method is now becoming increasingly established in forensic pathology concerned with large scale DNA typology rather than determining the presence or absence of healthy (or unhealthy) genes.

Laboratory Chips: In addition to the miniaturisation of gene identification tasks, tiny silicone platforms are also becoming available for combinatorial chemistry and high throughput screening. New 'lab on a chip' technologies have opened up the possibility of synthesising otherwise separate parts of the drug R&D process by simultaneously producing compounds which can then be tested against gene targets.

Databases: Bioinformatic databases can be characterised according to the way in which access is organised and according to the actual type of information stored. Nonproprietary libraries are entirely public and unrestricted by any access regulation. Genbank in the US, the European Bioinformatics Institute (EBI) in Europe and the DNA Database of Japan (DDBJ) are three of the most important nonproprietary databases. Some private 'niche' databases are also be held by companies for internal R&D purposes and unavailable to external subscribers. A large number of proprietary databases offer subscription access to paying customers alone. Proprietary privileges over the data itself, which can be in the form of stored DNA and or protein sequences, becomes the prerogative of the sequencing institution which may also hold the database. A map of access and data variation is provided in Annex A.

Storage, retrieval and dissemination: The software and technical infrastructures used to organise and store data have tended to develop incrementally and are therefore little different to techniques developed in the early days of bioinformatics. Flat files instead still commonly characterise the way in which data is managed. More flexible relational databases and fully object client interfaces are still fairly uncommon. Sequence search tools have evolved from simple keyword matches to highly complex alignment and pattern matching methods. BLAST is now the most established algorithmic search tool supported by the more rigorous but more time consuming FASTA.

Visualisation Technologies: The visualisation and modelling of molecular structures is important in refining potentially promising drug compounds and in better understanding the structure of proteins. Research activity in bioinformatics has produced a wide variety of tools to facilitate the better virtual manipulation of data.

Integration and Interoperability Bioinformatics is beset with the overwhelming problem of creating and enforcing standardisation. The failure to do so has led to failures of communication between research groups within and between firms, universities and genomics data libraries. Different teams working on the same gene or protein of a species often work in isolation because of various incompatibilities: competition and intellectual property considerations; variations in the structure of data bases; incommensurate biological nomenclatures used by research groups. Differences between data bases also produce a high number of mistaken search matches arising from different analytical views of the data. Such errors can be both misleading and costly.

1.2 Key Actors

Key bioinformatic actors may be grouped according to whether they are research institutions and alliances providing access to data, whether they are involved in the provision, organisation and distribution of funding arrangements, or whether, like the pharmaceutical industry, they are purchasers of bioinformatic data services. The characterisation below is intended to be illustrative rather than exhaustive. For convenience, a more detailed profile of each actor is provided in Annex B of this report.

Public and Private Sequence Research Institutions:

- **European Molecular Biology Laboratory (EMBL):** In the context of European public research, the EMBL is of utmost importance to the EU's standing in genomics and bioinformatics. With its headquarters located in Heidelberg, the EMBL operates three major outstations:

  The European Bioinformatics Institute (EBI) is Europe's largest repository of nonproprietorial gene sequence data. The EBI database is updated every 24 hours with new sequence data from the Sanger Centre (see below), Genbank (US) and the DNA Database of Japan. In addition to other databases, the EBI also maintains the SWISS-PROT and TREMBL protein sequence databases (see below). It also operates an Industry Support Programme involving more than twenty of the world's largest pharmaceutical firms and is significant in the dissemination of interoperability standards.

  EMBL Hamberg operates a research program using synchrotron radiation beam lines for molecular structural biology.

  EMBL Grenoble is concerned with applying nuclear physics to studying the dynamics of proteins and protein-nucleic acids.

- **The Sanger Centre** is supported principally by the Wellcome Trust and has generated as much as a third of the human genome sequence data available to the Human Genome Project. The Centre is the largest single contributor to the HGP. It exercises no proprietary rights to its data and by sequencing as much of the human genome as possible it seeks prevent others from exercising proprietary authority.

- **SWISS-PROT** is a curated protein sequence database and was established in 1986 at the University of Geneva and is now a joint project between the EMBL (with the EBI maintaining the database) and the Swiss Institute of Bioinformatics (SIB).

- **Celera Genomics** is a commercial company founded in 1998 to apply whole genome shotgun sequencing to the human genome. Celera's spokespersons have claimed that it will provide a more comprehensive sequencing of the human genome in advance of the publicly funded HGP.

- **Incyte Pharmaceuticals Inc.**[1] is one of the largest and most heavily subscribed proprietary database with as many as twenty firms currently paying for access.

- **Human Genome Sciences (HGS)**[2] is the commercial arm of the non-profit making Institute for Genomic Research (TIGR).

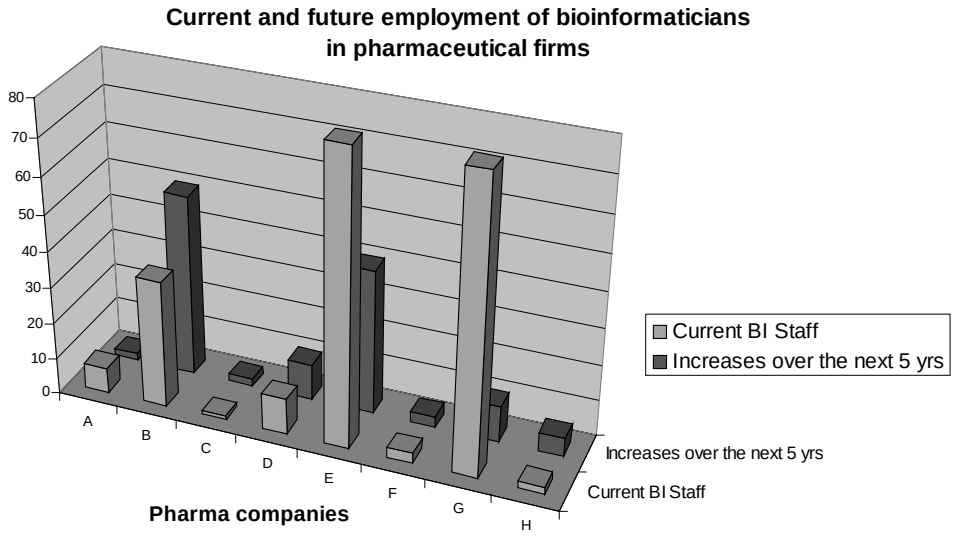
Section Two - Commercial and Practical Constraints

Developments in bioinformatics are emerging in relation to a number of key practical and commercial constraints which, in turn, will determine access to bioinformatic data and present certain limits to possible social benefits. Each of these constraints is addressed below and is considered in the context of findings generated from the study's questionnaire survey, interviews and sites visits.

2.1 Training, Personnel and Disciplinarity

Current and Future Needs: The emergence of bioinformatics has been accompanied by a well documented dearth in accessible expertise. This is cited as an acute problem both in respect to basic bioinformatic competencies and the rarer skills of those experienced in managing organisations as they enter and take advantage of new technological opportunities. The lag between demand for and availability of such personnel is a chronic problem, especially in the public sector and amongst small companies where salaries are less competitive. Indeed, all respondents consulted during the STOA study saw the present day scarcity and high value costs of BI personnel as one of the main barriers to SME access into BI-related opportunities.

Reflecting the scale of these new demands, the questionnaire study of the pharmaceutical sector revealed that most major companies anticipate increasing their bioinformatic staff by between 20% and 120% over the next five year period. More than half the respondents expected that their companies to at least double their bioinformatic staff over that period.



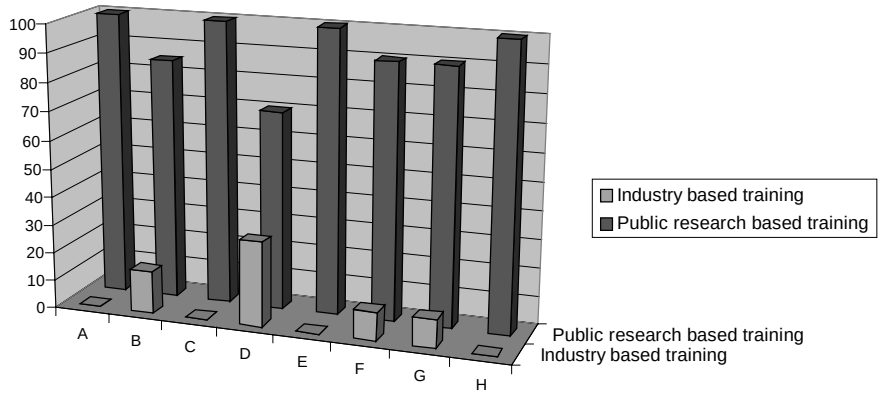
Towards a bioinformatic curriculum: As research increasingly shifts from biological entities at the laboratory workbench to the 'virtual' manipulation of such entities at the computer, individuals will require forms of graduate and post-graduate training which are current rarities in the University curriculum. Whilst information retrieval has traditionally been the task for professional science librarians, familiarity with on-line repositories now makes bioinformatic competence a professional necessity for all credible researchers. Bioinformatics presents formidable challenges to University education, not least because of its relative recency. The term 'bioinformatics' only appeared in the literature around 1991.[3] Yet Universities have been under considerable pressure to bring their curricula up to date.[4] A wide range of new university based taught courses have begun to develop in response to these needs. [5] Reflecting broad support for these developments, all respondents from the pharmaceutical sector agreed that over the next five years new University courses are likely to make a significant contribution to the availability of bioinformatic personnel. However, respondents expected that any real difference would be somewhat later, within the 5 to 10 year time frame. This indicates a severe time lag between industry needs and educational responsiveness. Indeed, both interviews and questionnaire responses revealed that this was not only a matter of delay but of educational content. Because the sector moves so rapidly, new University courses are unlikely to be able to do much than prepare graduates for the 'real' industry-based learning which pharmaceutical companies considered will continue to be their load.

A number of respondents and interviewees were particularly concerned that in comparison to the US, Europe is very poorly placed in terms of the expertise it requires to exploit the relevance of genetic data in the future. This is particularly the case in the context of biostatistics and fundamental population genetics as examples of disciplines which were once entirely theoretical but which now and in the future have the opportunity to be applied. Research and publications networks involved in

biostatistics are somewhat stronger in the US than they are in Europe and yet the historically high quality of health records in Europe makes the EU a potentially very strong mover in this area. These strengths may be consolidated by the population structure in Europe and the lack of controversy surrounding EU-based Human Genome Diversity (HGD) studies, in comparison to US led indigenous populations studies. Continued support of these fields and the designation of research resources would raise the profile of EU-based research which lends itself to many opportunities in population genetics and biostatistics.

The role of public research institutions: Notwithstanding the last point (above), the questionnaire study also indicated the importance of public research institutions in the provision of bioinformatic personnel and in bridging the skill gap between university courses and industry requirements. Whilst one company estimated that roughly 70% of its bioinformaticians had received their main BI training in public research, the majority of the estimates were in the region of 90-100%.

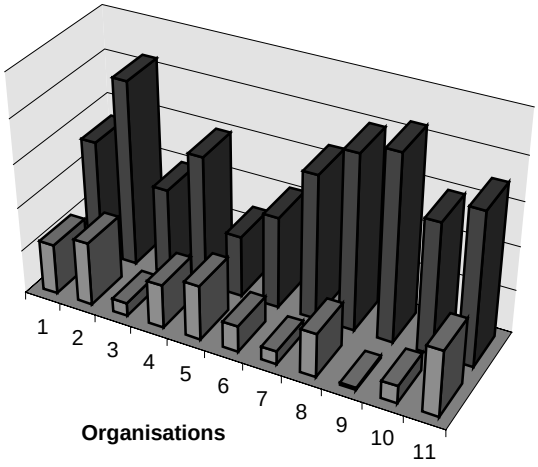
What proportion of your pharmaceutical bioinformaticians recieved their in a public research establishment?



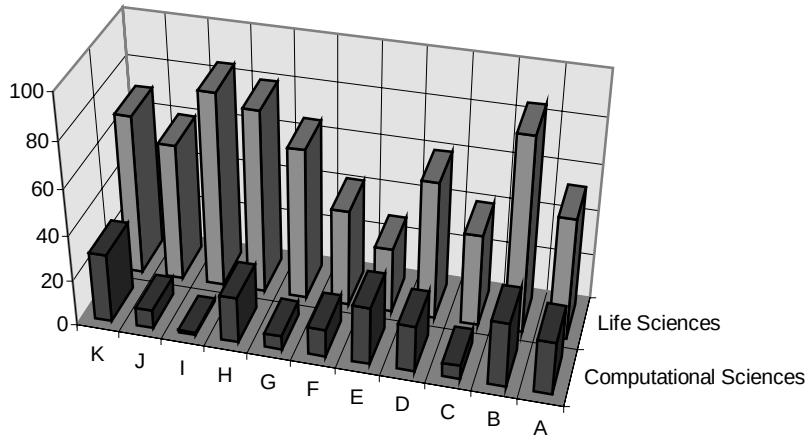
Bioinformatics and a new disciplinarity: How education and research institution-based training are to be reconfigured is far from straightforward. Bioinformatics is a highly unstable and heterogeneous mixture of different disciplinary heritages. Personnel frequently take a circuitous route into the sector from backgrounds in genetics, computational linguistics, software engineering, algorithmics, and so on.[6] The heterodox character of the sector has accordingly presented difficulties for curriculum developers – indeed, new organisational arrangements which cross-cut University-based disciplinarity are becoming essential in meeting the sector's training needs. Bioinformatics personnel often pursue graduate training twice across biological and computational disciplines which express poor integration between one another. New curriculum developments must be explored in detail to assess the degree of parity between education provision and the sector's needs.

The relative merits of different bioinformatics systems are seen to vary according to the disciplinary combinations which have produced them. For example, computational expertise often depends on data simplification, speed and efficiency. However, this can compromise the depth of complexity sought by biologists. This is an important consideration since credible systems rely on disciplinary competencies being combined in entirely unprecedented ways. Indeed, how different disciplinary assumptions impact on bioinformatics is a key consideration for sector's policy makers and innovators alike. The questionnaire of both the pharmaceutical sector and data-base providers indicates that a much higher number of bioinformaticians in the life sciences have a biological rather than computational disciplinary background.

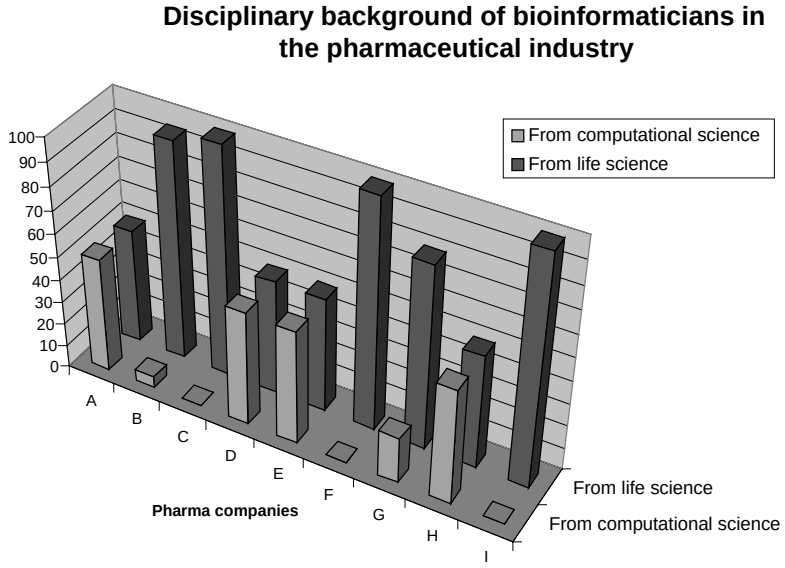
Disciplinary background of bioinformaticians in organisations that provide data-bases
(in percentage of the total number of employees)



Percentage of life sciences to computational sciences expertise in bioinformatic database facilities



New Institutional Learning: Bioinformatics, 'lead discovery' and 'rational drug design' have created entirely novel demands which represent radical departures from established ways of conducting research, communicating findings and producing therapies. A number of studies have noted the dependence of research actors upon being able to recognise new ways of managing their knowledge resources within highly accelerated computational environments.[7] A number of respondents considered this to be limited by limited vertical integration within research communities. Indeed, this has been noted as particularly acute in European research.[8]



Policy Options:

- To increase incentives for Universities to develop a bioinformatics curriculum which actively involves an Industry-based component. This would address disparities between University courses, where BI-resources will continue to be comparatively limited, and highly resourced industry research labs. Such incentives could be comprised of start-up funding arrangements linked to and, in some cases, eventually substituted by, industrial input.
- Many of the immediate recruitment problems in SME's and public research institutions relate to an insufficient volume of flexibly skilled recruits emerging from University training. Policy should concentrate on improving the volume and distribution of BI-skilled personnel as well as concentrating on specialist training through public research institutions.
- Measures must be taken to distribute BI-related approaches across all areas of the life science curriculum rather than in 'bioinformatics' departments or courses per se.
- Universities must be encouraged and supported in reducing disciplinary distinctions across computational and biological boundaries within their institutions.
- Revise funding criteria in research programs to accommodate disciplines which are best placed to exploit the future relevance of sequenced data - as in biostatistics and population genetics.

2.2 Scale, Size and Capacity

In addition to many new opportunities, developments in bioinformatics have generated equally novel competitive pressures which, in turn, may have serious implications for the long term preservation of public research activity in molecular biology. Given the costs associated with the bioinformatic research tools documented above, molecular biology is widely held to have assumed the 'big science' scale more usually associated with nuclear physics and space research. In turn, shifts in scale of these proportions necessitate the creation of new research alliances and specialisation if research institutions are to maintain there status in molecular biological fields. The implications of these pressures upon commercial, public and SME actors alike are considered in greater detail below.

Commercial Pharmacological Research: In the context of the pharmaceutical sector, market share continues to favour those companies able to dedicate smaller proportions of their overall R&D budget to meeting the very high costs of developing drugs, meaning much larger companies with a dedicated competence in maximising profits and reducing costs through bioinformatics. Questionnaire respondents from the database provider constituency were unanimous is identifying pharmaceutical organisations as the primary users of gene sequence data. Such organisations, having attained a certain size, are uniquely privileged in being able to afford both the high costs of bioinformatics and clinical trials.

The overall costs of bringing a drug to market can total in excess of £200m over a ten year period. Only ten per cent of the drugs that go into development actually receive registration, less than three per cent generate a remuneration which exceeds the cost of development. Greater access to gene sequence data through bioinformatics promises a reduction in the amount of investment put into drugs which eventually fail. Increasing this success rate depends upon pharmaceutical companies exploiting opportunities by which the 'attrition rate' of drugs in development can be reduced. As described above in relation to 'lead' and 'rational drug design', bioinformatics is essential in enabling R&D departments to identify promising clinical leads. Such leads can then help in determining proteins which may have a therapeutic bearing upon a condition. Equally, compounds with toxicological or deleterious secondary effects can be identified much earlier and in advance of failing expensively within the clinical trial phase of development.

The current high costs associated with attaining a competitive advantage through bioinformatics can only be met by much larger firms with greater capacities for flexibility in utilising gene sequencing, access to genetic data, pharmaco-genomics and combinatorial chemistry. In the questionnaire survey of the pharmaceutical sector and of public & private data-base providers, most respondents agreed that expenditure on bioinformatics is likely to continue to increase as companies search to maintain technological advantage. However, all of these respondents agreed that, whilst BI's importance is essential to increasing drug discovery capability, its contribution to reducing drug development costs (clinical trials and regulatory approval) is likely to be marginal (see *BI, Drug Innovation and Regulation* below).

Taking an early opportunity to invest in bioinformatics whilst maintaining the organisational scale necessary to see compounds through successful regulatory approval has created enormous cost pressures in the pharma industry. Changes in scale, size and capacity are certainly common features of the industry's adaptation to 'big (molecular biological) science'. The consequences of bioinformatics for such changes are evident in so called pharmaceutical 'supermergers' such as those between Astra and Zeneca and Hoechst and Rhone-Poulenc. Indeed, market analysts continue to speculate that an alliance between Glaxo-Wellcome and a suitable rival will mark a new round of such supermergers.

Public Research Institutions: The description above gives a clear indication of the potential for bioinformatics to increase asymmetries in resources between commercial and public research institutions. In responses to the questionnaire and site visits, public research actors were consistently anxious about their capability to maintain a research agenda which was relevant, up to date and able to take advantage of innovative developments in bioinformatics. In respect to public sequencing actors, whilst the price of producing sequence data is expected to fall (approximately $1/5^{th}$ of the current cost within the next five years), nearly all respondents agreed that the cost and complexity of analytical bioinformatics will continue to rise (see also section on public/private alignment below).

Consistent with our attribution of differences in BI-competencies to size and scale it is relatively small public research and diagnostic facilities (such as public health laboratories) who are most excluded from new BI-related advantages. High costs are concentrated particularly in: access to skilled bioinformaticians as discussed above; institutional commitment to high cost investments in new gene sequence and array equipment; new data platforms for storage and analysis. Most respondents, considered that if left unattended, the relative position of public to commercial research would deteriorate.

As a consequence, changes in the organisation of public research which take into account relatively limited budgets need to be considered (see policy recommendations below).

SME activity: High entry costs into BI-related industries is a current inhibitor on SME participation in the sector. This is considered to be particularly the case in Europe where venture capital investment is arguably less speculative than in the US. SMEs also tend to encounter similar difficulties to public research establishments in that bioinformatically competent staff are difficult to attract and equipment prohibitively expensive for start-up firms. Questionnaire responses from the pharmaceutical sector revealed that whilst this is likely to hold true over the next five years, access will ease considerably thereafter.

However, differences between SMEs vary according to the area of bioinformatics in which they are engaged. For instance, entry into the computational side of

bioinformatics is relatively more accessible than entry into high throughput, combinatorial and sequencing activity. This is largely because the risk distribution and product cycle duration vary considerably between computational and clinical-product companies (see chart accompanying the description of the pharmaceutical sector in Annex B). The former depend upon neat-term product markets whilst the latter depend upon much more long term investment. In the case of clinical-product SMEs, their bioinformatic competence usually involves investment by large pharmaceutical firms.[9] On the whole though, the price barrier for entry into the bioinformatic sector for SMEs is extremely high.

Policy Options:

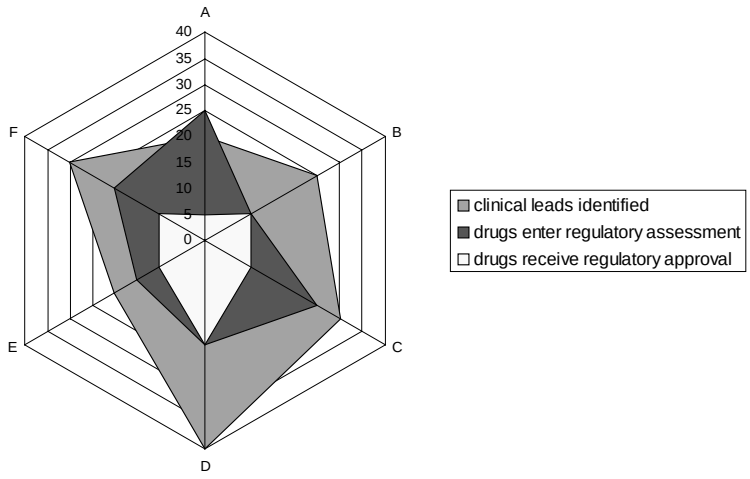
- The access of public research and diagnostic actors to bioinformatic tools also depends upon the availability of shared institutional resources such as that characterising other areas of high cost research, like physics for example. Measures for encouraging otherwise autonomous research teams to develop shared facilities need to be considered. Routes towards this can be established by requiring grant applicants to consider opportunities for consortium applications where an otherwise prohibitively expensive technology will be available to all members of the consortium.
- In relation to problems of scale for SMEs, policy will need to address the methods by which venture capital investment may be encouraged to support start-up firms. Also, some problems may be addressed by the development of incubator facilities where, again, facilities can be shared between a number of SME purchasers on the same site.

2.3 BI, Drug Innovation and Regulation

As mentioned above, investment in bioinformatics by the pharmaceutical sector is intended to be rewarded by improvements in the speed and efficiency of drug innovation. The relationship between BI-inputs and the speed of drugs through development, entering regulatory and receiving regulatory approval varied considerably by company and respondent though showed a progressive downturn in performance throughout the product cycle. This is represented in the graph below.

It was considered that total costs will continue to rise as a consequence of relatively fixed 'development' expenses rather than discovery costs where BI applies most. Over half the respondents believed performance increases to be attributable to developments in bioinformatics. However, all respondents from the pharmaceutical sector disagreed with the claim that BI might one day replace the 'wet-biological' verification of a compound's therapeutic value in animal models. The clinical geneticists surveyed shared the same expectations as those in the pharmaceutical industry regarding the potential of bioinformatics to speed drug innovation. However, in contrast to the pharmaceutical sector, nearly half believed that, within 10 years, bioinformatics would largely replace verification of drugs in animal models.

Pharmaceutical company estimates of the contribution of BI to % increases in the speed at which: leads are identified; enter regulatory assessment; receive approval



A-F refers to pharmaceutical company responses

In relation to regulatory capacity, most respondents thought that regulatory institutions across Europe were unprepared for an increase in the volume of therapeutic compounds being presented for approval.

2.4 Software and Hardware Interoperability

The Problem: Many things stand in the way of interoperability between different BI systems and research communities, not least the historical development of the sector. Bioinformatics systems were originally developed by relatively isolated research groups in response to local information handling problems. Established research groups have consequently exhibited a reluctance to part with their locally developed systems and their preferred vocabularies and terms of reference for compounds and genes. Considerable investments in established protocols exhibit a strong degree of lock-in which further stands in the way of interoperability. The degree of flexibility and openness to change by such actors is likely to be limited because of the cost (in financial terms and the loss of prestige) of reorganising nomenclatures and data handling systems. However, a large majority of the questionnaire respondents identified the main barriers to interoperability as organisational rather than simply matters of technical architecture.

The questionnaire study also revealed the extent to which anxieties around the security of data have played a part in research community isolation, and as a consequence inhibiting competitors from adopting standardised protocols (see also section of public/private alignment below).

The costs attributable to interoperability failure are considerable. Interviewees from the clinical sector agreed that the interoperability of bioinformatics data is crucial for ensuring that future clinical benefits are realised. For example, it is possible that different teams working on the same gene or protein may be unaware of identical work completed elsewhere. If cross matching between data bases is to be as automated as is hoped, these difficulties organisational and interoperability barriers have to be overcome.

Enhancing Interoperability: Many of the initiatives which seek to create consensus on interoperability are reviewed in Technical Annex A of this report. However, when asked which initiatives have been most valuable, questionnaire responses converged on a number of developments which have been significant:

- The Object Management Group (OMG) Life Sciences Research Task Force dissemination of the Common Object Request Broker Architecture (CORBA).
- The 'Bio-Standards Project' of the European Bioinformatics Institute (EBI) (partly funded under the DGIII Information Society Standardisation Programme).
- Community adoption of JAVA based interface.
- More widespread purchasing of multi-platform approaches.
- Respondents believed that the HUGO nomenclature committee should become a mandatory arbiter in generating agreed biological terms.

Future Expectations: Most respondents believed that their organisations will soon achieve interoperability across all important internal R&D systems and all saw this as a fundamental long term (5-10yrs) priority. However, in respect to interoperability between organisations, most respondent believed complete integration to be somewhat over-optimistic, even in the long term.

Interoperability and Monopoly Risks: A number of concerns have been expressed with regard to potential opportunity for monopolistic actors to take shape as the bioinformatic community struggles towards an acceptable degree of standardisation on hardware and software.[10]

Small software developers, once characteristic of the bioinformatics sector, now tend to sell their products to much larger companies. The fear is that the area may become increasingly dominated by a small number of commercial actors providing highly integrated visualisation, search and design packages. Indeed, the dominance of the administrative sector by Microsoft is seen as one paradigm for the way in which computerised biological research will be increasingly secured by monopolistic suppliers. More worrisome are the Microsoft changes to JAVA making it MS-specific, defying JAVA(MS) use under Netscape.

When asked whether BI might become characterised by a single 'Windows-type' common platform, dominated by one or more monopoly actors, interviewees were generally quite sceptical in the long term (5+yrs) and disagreed strongly on whether this would be the case in the short term (-5yrs). In broad terms, few respondents saw this as a realistic scenario, especially since: the BI field is thought to move too

quickly for a monopoly to establish; academic and public participation in the field is higher than in other computational markets; complete harmonisation is relatively unlikely; CORBA and Java are expected to provide a degree of diversity within standardised protocols.

Indeed, a number of respondents affirmed the view that great caution needs to be exercised in relation to fuelling or inciting anxieties about monopolies forming in response to interoperability problems. For example, a 'go your own way' response to such fears - even if intended to safeguard European independence from the US-dominated software-tools market - may actually contribute to more acute interoperability difficulties.

Policy Options: On the basis of the above review, it is clear that harmonisation across secure internet technologies is considered to be a key necessity. A number of institutional initiatives and policy initiatives will contribute to this objective:

- Initiatives which assist in the community adoption of Java and CORBA particularly amongst small public diagnostic and research facilities.
- Continued support for OMG Life Sciences Working Group.
- The EBI's standards programme is taken as a good illustration of the way in which a public bioinformatics facility can take a prominent position, even amongst large commercial actors, in the dissemination of standards. Such initiatives should be considered a long term necessity.
- Whilst the EU is in a good position to support such initiatives as those described above, many in the bioinformatics community believe legislative regulation to be only appropriate in exceptional circumstances and would therefore recommend a 'light touch' form of regulation, if any at all: 'Only if European Law forbids incompatibility changes unjustifiably being used to avoid free competition and capture users to proprietary formats or enforces use of clean open standards'.
- The role of concerted action programmes under existing and forthcoming Frameworks could oblige research institutions towards a standardised interface.
- In addition to technical improvements, organisational barriers between research communities need to be overcome. This can be augmented by obliging cross institutional grant holders to improve shared data facilities between one another thus increasing opportunities for the EU's to reduce restrictions in the movement of information.
- Grant holders must be obliged to accept names and symbols generated by the HUGO Nomenclature Committee for their findings.

2.5 Public / Private Alignment and Access to Data

Introduction: Both public and private actors play a crucial role in making available bioinformatics data as a publicly accessible good (see annex B). In this section we will look at the different issues that shape the relationship between both public and private constituencies.

While the division between public and private actors can be useful to understanding some of the differences and dynamics between both public and proprietary actors and the role of publicly supported research institutions, the reader should bear in mind such division is a somewhat crude distinction. Public-private alignments vary considerably thus highlighting the difficulties of identifying purely commercial or purely public research actors.

Bioinformatics can be said to be situated at the intersection of academic, government, and commercial interests. The mixture of different types of organisations in a setting that combines competitive and co-operation incentives creates a blurring of boundaries to do with the structures of ownership, funding, reward, and accountability. This blurring of boundaries complicates the impact of the public versus private characterisation on the movement and exchange of bioinformatic information. In the following, we will reflect on the 'blurred boundaries' between different constituencies as much as possible.

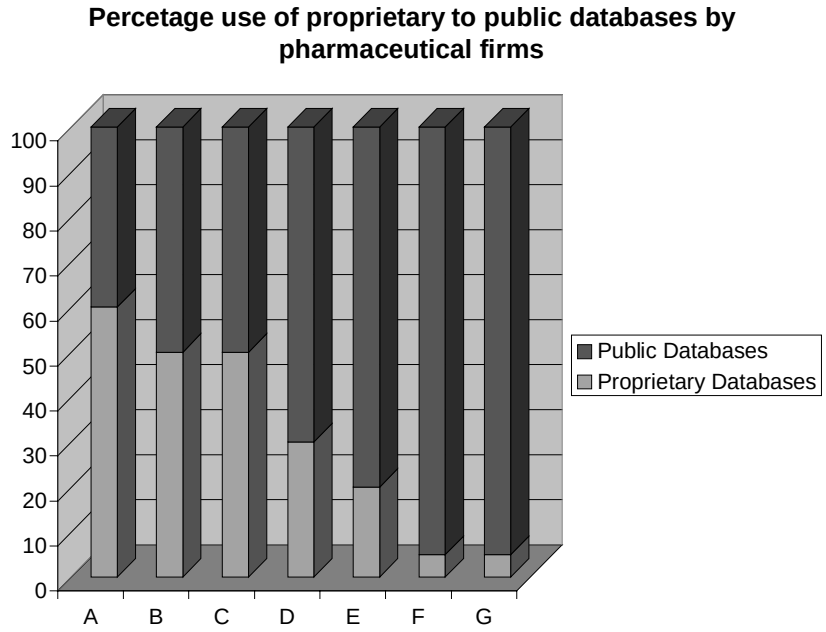
Accessibility of data: It is perhaps not surprising that institutional arrangements for determining access to bioinformatics vary considerably making it difficult to formulate a stable distinction between 'public' and 'private'. While the majority of basic bioinformatic knowledge might be probably better considered as 'pre-competitive' and to be make available as a public good, as this would collectively benefit the private sector more than if this knowledge were haphazardly spread across less (or non-) accessible resources, there is nothing inevitable about this viewpoint. On the contrary, in practice the public status of knowledge is becoming increasingly difficult to maintain in the context of a growing tendency to exploit data for commercial interests.

There is nothing inevitable about bioinformatic knowledge being made available as a public good: on the contrary the public status of knowledge is increasingly difficult in the context of a growing tendency to exploit data for commercial interests.

Academic or commercial alignment cannot be treated as the crucial factor in determining whether or not knowledge is exchanged openly. Industry often acts as an important source of scientific and technological knowledge as the case of TIGR reveals. Likewise, academics in competitive, high risk and high reward areas of research do not always share information freely. As universities assume the entrepreneurial role prescribed to them in much of European and national innovation policy, separating academic from commercial activities will be more and more difficult.[11] In bioinformatics, where the divisions between academic and commercially relevant research are blurred, ample opportunity exists for commercial ventures. Recently spin-off companies set up from the University College of London[12] and the European Bioinformatics Institute[13] are indicative of the commercial activities in bioinformatics. The formation and operation of such spin-offs involves a delicate process of negotiating the boundaries of 'public' and 'private' having implications for the accessibility of information.

Despite these considerations, for the sake of clarity (and because of the way in which we chose our interviewees - e.g. tried to define clear boundaries between interviewees) our discussion of the interview results draw on a very distinctive boundary between pharmaceutical companies, public database providers and proprietary database providers.

The use of Databases: The relative use and importance of public and private databases varied considerably from one pharmaceutical firm to another.



All pharmaceutical companies except one possessed mirror sites. When asked whether public sites were likely to be more significant to drug innovation than private data sites, respondents were fairly evenly divided though expressed marginal agreement to the question. Especially in the long term both were considered to be necessary.

Of the proprietary database providers we interviewed, we did not obtain detailed information on the number of subscribers. This was either confidential or subscribers were not yet available. The number of database subscribers to public databases varied considerably. Figures ranged from 400 to 8000 subscribers while the average number of 'hits' per day ranged from 100 to 8000.

Among pharmaceutical respondents there was very strong agreement that public database use is set to increase – but not necessarily proportionate to proprietorial databases. While a similar agreement was found among respondents from proprietary databases, this agreement was less strong among public database

managers. One of the reasons for the latter to be more sceptical about this subject is the feared withdrawal or lack of funding for public databases (see below). While the increase use of proprietary databases might be most visible in the short term, less consensus existed with respect to the longer term.

All but two respondents out of 16 thought large pharmaceutical companies (as opposed to SME's and academic researchers) would be the most important customers of proprietary databases.

Cost of proprietary databases: Bioinformatics research requires the exchange of DNA samples as well as written material. If the resources needed to access knowledge are high, support structures limited and terms of use restricted, this could produce considerable costs and time delays.

While all database providers thought the cost per sequence is likely to fall considerably[14], pharmaceutical companies were mostly divided on the future of the subscription cost to source genetic data. Most respondents from the pharmaceutical constituencies believed that considerable reductions in subscription costs were unlikely, especially in the long term (5-10 years).

Legislation: Openness & Ownership: Discussions about the openness of research typically revolve around discussion of intellectual property rights such as copyrights or patents. While, no doubt, the rules surrounding the formal ownership of knowledge have important consequences for the openness of research (see below), these need to be balanced against other considerations. For example the fact that public databases are updated every day and freely accessible, is highly valued by the research community. However, it is not only actors from public databases who perceive this provision of knowledge as crucial to the field, the same response was given by private constituencies. Both proprietary and public databases, in other words, will have to be valued in their own right.

Whilst public databases make a fundamental contribution to therapeutically relevant sequences, provision for subscriptions to proprietorial databases will continue to be important. Significant cost increases will be in the area of added-value data (biological and medical). As one respondent said about the future activities of commercial activities in the field:

> *"Our company data is currently proprietary but some estimate that most of it will be in the public domain inside 2 years. Innovation of new data products that add further value - e.g. genetic variants, genotype/phenotype correlates - will become the data products of tomorrow and are likely to be only achievable with private investment bolstered by industry subscription in the hunger for new data to build competitive advantage".*

The question whether EU legislation on gene patenting and data protection, compared to the US was in favour of actors own constituencies, was only answered by half of the respondents from public and proprietary databases. Most of them thought EU legislation was not favourable compared to US legislation, but a few thought it was. Within the pharmaceutical sector, respondents were evenly divided on which regulatory context favoured their sector most. Comments were made with respect to copyright and the fact that this had disastrous consequences (probably related to time-lag and the disadvantages with respect to US competitors). Others criticised the European rule that information must be confidential prior to patent submission.[15] Whether and when researchers 'publish' their results has important implications for encouraging innovation. Significant time gaps exist between research results and dissemination, to conceal knowledge from competitors, which then affects the incentives of others to pursue further work. Both pharmaceutical industry and public database providers criticised the EU for not allowing a grace period in patenting.

From the questionnaire it is not completely clear how the issue of openness and ownership in light of current legislation should be interpreted. Does the legislative environment in Europe, where most information is freely accessible, disadvantage European research initiatives? Does the patenting climate in the US for example, outscore European research? From the site visits conducted for this report, the conclusion is that most actors are not worried about the fact that Europe's open access policy may detrimentally effect its ability to both exploit and regulate bioinformatics-based health technologies. As one of the interviewees claimed:

> *"I cannot see a US company will inhibit EU research because they patented data that the EU has been giving away freely".*

If actors reacted on the issue at all, they seem to see a much more fruitful way forward would be to boost European based patenting and create opportunities for intellectual property rights than changing the open access policy. As one of our interviewees explained, providing free access of information could also be an advantage to researchers. In this particular example researchers had made available a particular set of markers that became well used by (forensic) laboratories all over the world. The broad diffusion of his markers via the internet provided the researchers with a huge research network and extensive contact with other researchers. This network, according to the researchers, was in itself a very valuable and important resource for further research.

Access to Genetic Health Registers: Probably one of the most striking finds of the questionnaire is that there was almost unanimous agreement upon the question whether R&D will increasingly depend on making arrangements to access public health genetic registers – as with the recent 'Icelandic case'. This refers to an initiative in which an Icelandic company (DeCode Genetics) proposed the establishment of a multipartite database (including a core genealogical database, a disease specific database each of which to be supervised by an expert panel to

evaluate proposals). After much debate, the Icelandic Government eventually passed a law which assigned DeCode with the task of setting up the database with rights to commercially exploit it. The company can accordingly enter into commercial arrangements with pharmaceutical firms and recently agreed to pursue ten disease groups in a $200m deal with Hoffman-La Roche. This leaves the door open for other companies to approach DeCode on other disease areas.

Nearly all respondents agreed that R&D within their organisation would benefit from the availability of public health genetic registers.[16] Interviewees thereby foresaw relatively few problems in respect to the availability of genetic registers to commercial agreements. On the other hand, health care progesssionals are confronted with increasingly tight confidentiality regulations around the establishment and maintenance of registers.

Furthermore, one cannot rule out the possibility that without proper information and consultation, public opinion may increasingly turn against the use of genetic health registers, despite their potential therapeutic value. Hostile reactions to the Icelandic case are illustrative here. The debate around the use of genetic health registers resembles discussions around indigenous tribes and plant species for example, and the use of individual's DNA for databases. Within the public domain, the use of genetic health registers might trigger a higher level of concern than we found thus far among professionals consulted.

The relationship between public and private databases: One important aspect of the provision of genetic databases is the relationship between both private and public databases. While the existence of both public and proprietary databases might have created a stimulating innovation environment, at the same time it generates a number of important tensions.[17]

In May 1998, an alliance between Craig Venter and the biological devices manufacturer Perkin-Elmer announced the application of fully automated 'whole genome shotgun sequencing' to the human genome. In so doing, Celera Inc. promises to supersede the efforts of the publicly funded HGP bringing with it considerable criticism of the efficiency of the scientific method upon which the public HGP has been based. In addition, the alliance uses third generation Perkin-Elmer's sequencing machines which can run with less direct supervision than the majority of machines used in public research establishments thus reducing cost and increasing speed.[18]

Several significant areas of concern have been raised in relation to the application of the shotgun method to the human genome. First, shotgun sequencing is considered to be less accurate than more deliberate clone by clone strategies since it relies so heavily on repetitive sequencing. This is especially the case in large organism research where a significant proportion of the genome has no known function making the assembly of a genetic map particularly imprecise. This can only be offset by increasing the coverage and automation of alignment procedure.

Second, the Celera initiative increases the commercial opportunity for control of human sequence data in copyright law and the patenting of human genes. In response to the latter risk, the publicly funded parties involved in the Human Genome Project, have announced major increases in sequencing capacity and the strategic use of shotgun sequencing to arrive at a 'rough draft' (as opposed to whole shotgun sequencing) of the human genome. This will be eventually completed by using more targeted techniques.

Although all respondents felt there is a strong need for both public and proprietary database providers to co-operate, they differed on the question how feasible such co-operation would be. Commercial companies are more optimistic about the future of this co-operation than are the public constituencies. The latter for example said they suffered from inaccessibility of data from private companies (and sometimes felt the private initiatives as a real threat to their research) while private companies were more convinced of the mutual benefits (instead of mutual exclusion).

A majority of respondents from public database providers agreed with the statement that whole genome shot-gun sequencing, as is done by privately owned commercial firms such as the Venter initiative, forms a major threat to public databases. Whether this threat would be more of a current or a future problem, however, triggered very mixed responses. While some thought the problem was very pressing within the first five years to come and not so much in the distant future, others felt it was exactly the other way round. Respondents gave different reasons of *why* they felt private initiatives might be a threat to public databases.

One of the issues brought forward was the unequal share of investments put into staff development. Academics are often said to decide to move to private companies, taking with them both expertise and knowledge and sometimes even more than that:

> Most often academic developers just quit their organisation to join a commercial company taking their packages with them which benefits more the industry that gets them.

Free competition, in other words, according to the public sector, is generally lacking.

Questions on a second possible restrictive effect of proprietary data was included in the pharmaceutical questionnaires. Respondents from within the pharmaceutical industry were divided on the question whether there were any inhibiting effects of proprietorial claims to sequence data on drug R&D. Although one respondent urged for a shift in proprietorial emphasis away from genes and towards compounds instead, on the whole, the pharma constituency were broadly sceptical, or at least uncertain, about how inhibiting proprietary claims to sequences were.

A third remark made by a public database provider is the fact that the private sector secured its market by using incompatibility (interoperability) as a main asset. To the

question whether interoperability is a technical problem or not, this respondent answered:

> Linux, Emboss and other free and open projects clearly demonstrate that the problem is NOT technical, nor is it expenses. But just merely unwillingness of private companies to allow free competition.

Others claimed commercial soft-ware often changed its data format so that interoperability would be reduced. For public databases this means buying new software might take up a rather large part of their budgets. This relates to another constraint that public constituencies felt with respect to private initiatives, which might be called the *funding bias*.

The funding bias: The source of funding does not necessarily lead to an easy characterisation of whether that research operates with principles of free access. Publicly funded programmes can set rules that affect the degree of openness. In bioinformatics, many research programmes combine government, commercial, and non-profit funds and thus negotiate access principles across varied groups. However, most of the public database providers included in our questionnaire have an open access policy with respect to the provision of information.

Actors can be distinguished in relation to whether they generate sequence data, gather data (primary databases) or re-compile data (secondary databases); keep data up to date, monitor for errors, inconsistencies and repeat entries. In relation to public databases, avoiding an erratic position in relation to such problems as these depends upon secure continuity of funding. Respondents from public and proprietary databases commented frequently on the *funding* of public databases. Of major concern to both public and private constituencies is the question whether public databases will be able to secure funding in the future, especially for maintenance and up-dating software programs.

Funding issues are said to be *biased* in favour of the private sector. The latter do profit from the availability of publicly funded data which is often crucial to the development of their commercial products.

Most of the lessons that can be drawn from the relationship between public and proprietary databases, we feel, have to do with the issue of unequal competition.

Policy Options:

- In order to anticipate the full consequences of a broad use of genetic health registers, the EU should consider monitoring the full impact and potential reaction to this development.
- The EU should consider whether the dependence of the private sector upon public institutions and the perceived funding bias this creates, should be addressed in future funding arrangements.

- Funding arrangements must take into account that many research establishments have now become necessary infrastructural features (databases, skills and knowledge) of Europe's standing in bioinformatics. Funding arrangements must therefore be adapted towards a continuity which extends beyond strictly defined research projects.
- While the provision for open access to information is highly valued by the research community, the continued encouragement of copyright and patent protection, in the context of public-private alliance, continues to be a key priority for sustaining EU commercial standing in a new biological economy.

Section Three - Changes in Health Care Delivery

The Human Genome Project and derived genetic tests - in particular the development of tests for late-onset multi-factorial diseases- will have consequences not only for the way medicine is practised, but for the whole meaning of medicine as well. If realised, bioinformatics developments in areas such as high throughput screening and lead drug design will have a significant impact on perceptions of health at the level of the patient, the determination and allocation of health resources, and requirements for counselling and training of clinical and associated personnel. This section outlines some of the possible implications of bioinformatics for the importance of genetics in clinical practice and thus for health care in general. Overall, it can said that respondents from across all the constituencies interviewed for this report attested to the significant contribution of bioinformatics to clinical practice, although this would be in five to ten years time.

3.1 Genetic Testing

It has been argued that the clinical delivery of genetics has largely lagged behind developments in biotechnology.[19] Even in the US, despite a 15 year explosion in biotech companies, large scale genetic screening and therapy can hardly be said to be 'booming'. However, a number of clinical applications have emerged particularly in family screening. The BRCA1 test entered clinical practice within six months of its isolation and is now one of the most frequently used genetic diagnostics.[20] It is perhaps on the basis of such activities that nearly all the clinical geneticists contacted for this report thought the large scale screening of multi-factorial diseases (e.g., perhaps facilitated by genechips) will become increasingly commonplace over the next five to ten years.

Problems have arisen, however, in relation to the adequacy of existing genetic tests. Only 35% (BRCA2) to 65% (BRCA1) of mutations can be detected. This raises ethical and practical dilemmas by generating insecurity, false reassurance and misconceptions. This is the case for genetic diseases. Another consideration arises from the fact that mutations do not always express themselves in the expected disease or, instead, lead to different pathologies. Therefore, in considering large scale screening campaigns, such indeterminacies need to be carefully evaluated, especially given the professional requirement of 'non-directive' counselling (i.e., the provision of informed choice without coercion). Finally, a major problem remains over the most effective strategy for the prevention of genetic diseases, such

as breast cancer, and the role that screening can play in this. Screening can detect/confirm those identified as being at risk, but also those who have no awareness of being at risk.

A recent comparison on the organisation, legislation and finance of genetic services in Europe revealed that in countries such as Germany and the Netherlands there is great reluctance among professionals to be associated with large scale screening.[21] Eugenics and other problematic aspects of the history of genetics have been offered as reasons for this resistance. Presymptomatic and predictive testing for late-onset genetic disorders such as inheritable forms of cancer, Alzheimer and diabetes, in these countries are therefore expected not to diffuse easily.[22] Linked to this is the fear that diagnostic information may be misused or inappropriately distributed.[23] An essential prerequisite for any type of genetic test is the assurance of absolute confidentiality and the protection of data from third parties.[24] Yet, a large majority of the clinicians surveyed believed non-anonymised genetic information on individuals would circulate between clinical and non-clinical sectors.

3.2 Drug developments

Parallel to new diagnostic possibilities are the anticipated products and services seen to emerge from commercial pharmacology. The number of drugs which can be directly attributed to lead discovery from genetically sequenced data is set to increase considerably over the next ten years. One major pharmaceutical firm, for instance, estimates that this will increase from roughly 10% today to 50% over the next decade. Most other pharmaceutical companies have similar projections, including 1% to 12%, 0% to 60%, 5% to 100% for some companies.

It is hoped that developments in bioinformatics will mean knowledge of an individual's genetic constitution can help assess both a predisposition to a disease and the specific drug that should be used to treat it. Drugs that are effective and safe in one patient may be toxic or ineffective in another.[25] Accordingly, drugs are likely to become less generic and increasingly more heterogeneous. While most pharmaceutical companies interviewed did not foresee individualised treatment regimes (a drug being sold on the basis of results from a genetic diagnostic) becoming commonplace within the next five years, all agree that this will be very likely within the next 5-10 years. Indeed, many respondents agreed that, within this time period, most drugs will be sold in combination with a diagnostic as part of a kit. In some cases, the questionnaire revealed that respondents anticipate that this may reduce the role of GPs in diagnosis and drug prescription. Also, roughly two thirds of the pharmaceutical companies estimated that individualised treatment regimes will give rise to a much greater number and diversity of compounds entering the market over the next ten years. In addition to treating disease symptoms, compounds are also being developed to correct the underlying genetic causes of disease. Richard Sykes, Chairman of Glaxo Wellcome, anticipates marketing genetically-based drugs within five to ten years.[26]

3.3 Disease classification

The vast majority of respondents maintained bioinformatics is also likely to lead to new classifications of diseases. Diseases traditionally characterised through symptoms (phenotype) may be reclassified according to their genetic characteristics (genotype). Distinct disease categories may be found to share similar genotype properties whilst other classes may have to be internally differentiated.

3.4 Genetic Counselling

The introduction of DNA tests for frequently occurring cancers such as Breast and Ovarian cancer, has made current genetic services realise they are unable to cope with both the demand and the complexity of these disorders.[27] Increasing demand for BRCA1 and BRCA2 testing is only the beginning of an increasing demand for genetic counselling.[28] Questions around genetic counselling have led to a number of policy initiatives including the aforementioned study on European genetic services called CAGSE (Consorted Action on Genetic Services in Europe). Genetic counselling has also been addressed by EUROSCREEN, established in 1992 to analyse the ethical implications of human genome analysis for clinical practice in medical genetics.[29]

3.5 Recruitment & Training in Medical Practice

New genetic services and counselling demands have placed pressures on the recruitment and training of clinical staff. This not only relates to the formal training of professional geneticists but to the knowledge-base across all medical practitioners who will more and more be confronted with the genetic aspects of diseases. The uncertainties surrounding diagnosis, especially that of multifactoral genetic disease, or where gene-environment interaction is important in determining the onset of disease, can lead to considerable professional difficulty for doctors in providing appropriate counselling to patients and families.

Advances in bioinformatics and the possibilities therein entailed for clinical practice (e.g., with regard to classification of diseases), will give further impetus to general demands for genetics-related skills. For instance, the clinical geneticists surveyed were nearly unanimous in maintaining that skills in interpreting genetic data will be essential across the medical profession and that knowledge of bioinformatics is likely to lead to new disease classifications and the re-classification of existing nomenclature.

Despite this situation, education and training is perceived as a serious short-coming in most countries. All but one of the clinical geneticists doubted whether a sufficient skill-based existed within European health care services to handle bioinformatics-related developments. For instance, only a few countries so far have nationally organised training programmes in the area of ethics.[30] In terms of the practical and commercial constraints on the uptake of bioinformatics derived

products and services, clinical geneticists identified the clinician training and skills base behind health care service institutional inertia as the second most significant a constraint.

3.6 Domestic Healthcare factors

There is a need to recognise that the market for clinical products will be shaped by factors at a national level which may work against the deployment of certain techniques. The agencies within European countries that are responsible for genetics (e.g., public health) are therefore unlikely to respond in an identical way to the available bioinformatics-derived techniques. The European Parliament may need to accommodate within its bioinformatics provisions considerable variation between public health regimes across member states. Attempts to foster European-wide health policy based solely on assumptions of clinical benefit may meet with limited success. While in the short term the US is considered to be a better market environment for bioinformatics derived products, given the American appetite and incentives for new technologies, responses indicate that the gap between the US and European markets will close in the longer term (5-10 years).

3.7 Cost of Health Care Delivery

Respondents from clinical genetics and the pharmaceutical industry disagreed with the hypothesis that high expenditure on bioinformatics would be reflected in cost *increases* of pharmaceutical products. Consistent with points made above, this is largely because bioinformatics drug discovery is relatively cheap in comparison to drug development costs. The portrayal given from these constituencies is that the price for drugs will remain fairly stable. Any reductions bioinformatics is likely to make to the costs of clinical trials will be minimal. Even in the pharmaceutical industry, for instance, more than half of the respondents doubted bioinformatics would significantly reduce the attrition rate of the drug candidates that enter trials. The remaining respondents agreed that attrition would be 'lower but not much lower'. A long-term goal should be to determine the effect these changes will bring to the overall costs and nature of health delivery and health care across member states.

Policy Options:

- The European Union needs to address the way in which public research institutions, particularly those supplying genetic diagnostic services, can adjust their scale in order to access high cost bioinformatic resources. This may involve using quality accreditation mechanisms to identify which centres have developed an expertise in a particular area of genetic diagnosis. Such centres of excellence will in turn require EU-wide investment in return for services provided to member states. Respondents in the study agreed that whilst insufficient heterogeneity between public research institutions will compromise quality, too much creates a situation in which resources are inadequately targeted. It is hope that this recommendation will better facilitate necessary changes in scale, size and capacity.

- Continuing efforts need to be made with regard to genetic counselling in European to ensure the ethical implications of human genome analysis for clinical practice are adequately handled. In addition, the EU needs to ensure that when large scale screening programme are undertaken by members states in manner sensitive to the indeterminacies of such activities.
- The genetics-related skill base within European health care services should be enhanced to ensure the proper treatment of bioinformatics-related developments.
- Confidentiality considerations and the protection of data from third parties is a significant area of concern and will become more so in the future. The EU is recommended to subject genetic information to proper codes of practice and systems of financial redress to prevent non-anonymised genetic information on individuals from circulating between clinical and non-clinical sectors.
- A long-term goal should be to determine the effect of BI to the overall costs and nature of health delivery and health care across member states.

Section Four - Non-Clinical Products and Services

4.1 Bioinformatics and the Insurance Industry

Information relating to health is central to determining the mutually acceptable balance of risk distributed between insurer and insured. Bioinformatics has given rise to the anxiety that this balance may change considerably. On the one hand, insurers are concerned that an undisclosed test result will be used to insure high risk individuals for higher amounts. This subverts the terms of contractual arrangements in which both parties symmetrically enter into a contract without either having prior knowledge of predetermined events. On the other hand, applicants may be refused insurance or pay a disproportionately higher premium if they decline to undertake a genetic test or fail to reveal the results of a test previously taken. The availability of this kind of information to insurers may also detrimentally affect the uptake of potentially beneficial tests by individuals concerned about the misappropriation of their results. In addition, there are concerns that the insurance sector is insufficiently prepared to make reliable estimates of future risk based on an informed interpretation of genetic data.[31]

Anticipating changes in insurance practice arising from bioinformatics, it is important to take note of the existing use of medical data. The UK Human Genome Advisory Commission (HGAC) recently undertook a survey of the risk calculation methods used by the insurance industry finding that: [32]

- Genetic data is unlikely to figure in risk calculation.
- Underwriting is likely to be denied on the basis of genetic evidence only if the risk is more than five times the standard.
- Insurers have little experience of using information from genetic tests. Although they would expect to be told if a previous test had been taken. Not being told may constitute grounds for voiding a subsequent claim.

- Genetic tests are rarely, if ever, requested because administrative costs may well exceed savings made from excluding people less healthy than average. 95% of insurance applicants are grouped together and charged 'standard' rates thus avoiding prohibitively expensive admin. costs.
- Calculations of morbidity (health insurance) and mortality (life insurance) are made on the basis of age, height, weight, personal medical history (treatment and medication), family medical history, alcohol and tobacco consumption and dangerous sports. Hereditary factors are usually made on the basis of family history not genetic test data. Information is usually volunteered by the applicant and requested from the applicant's medical practitioner who is legally bound to reveal relevant information.
- Medical data is never shared between companies (except reinsurers).
- Companies tend to store relevant medical data on both successful and unsuccessful applicants for many years in case of a reapplication.
- An applicant's data is never used in assessing another family member.

A number of domestic and international arrangements have taken shape in response to the possible use of genetic data by the insurance industry. However, legislative control and voluntary self-regulation is highly variable and often inconsistent depending upon local health provision arrangements and regulatory legislatures:

In the US, more than 30 states prohibit insurance companies using genetic tests as a precondition for cover.[33] Also, the federal Health Insurance Portability and Accountability Act (1996) prevents insurers from using pre-existing genetic conditions to deny cover. The US (NIH/DOE) Ethical Legal and Social Implications of Human Genome Research (ELSI) has also produced a number of reports.[34]

In Europe, several member states have initiated measures. In 1992, the Belgian parliament precluded insurance companies from requesting or using genetic information in the determination of life insurance contracts. The Dutch insurance sector was the first to self impose a moratorium which has recently been reaffirmed. The UK has seen the drafting of a self-regulatory Code of Conduct by the British Association Insurers. The Code discourages companies from requesting applicants to take a genetic test though allows them to request further information on a test already taken. It also permits companies to raise the premium of an applicant on the basis of genetic data. In accordance with UK Data Protection Legislation (1998), the Code requires prior consent stating the sole purposes for which disclosure of genetic data is sought before an application is processed. The test itself must have been approved by the National Health Service and is, therefore, restricted to single gene defect diagnostics.

At the level of the European Union, the Council of Europe Convention on Human Rights and Biomedicine, under Article 11, prohibits any form of discrimination

against a person on the grounds of genetic heritage. The Convention was adopted by Ministers on 19 November, 1996.

4.2 Bioinformatics and Financial Services

The financial sector illustrates many of the emerging technologies being developed to enable service providers to discriminate between legitimate and bogus clients. A whole range of new technologies highlight an increasing tendency towards the use of automated recognition systems based on biological 'signatures'. To date, a substantial number of companies have emerged to market security devices which scan iris or retinal patterns, facial features, vein patterns, ear shapes, fingerprints, vocal wave forms and even olfactory signatures.[35] In contrast to law enforcement, the financial service sector has not yet added gene-based bioinformatic technologies to this list of 'embodied' recognition systems, though genechip technologies (discussed above) may attract interest in the future.

In estimating the potential availability of bioinformatic technologies to financial services, it is important to recognise existing acute problems in biometric tools:

- A high number of biometric companies have failed because of a lack of confidence amongst potential system purchasers.[36]
- Boimetrics are costly, even when balanced against fraud prevention savings.
- Existing systems fail to demonstrate the kind of accuracy desired by the financial services sector. For example, approximately 2% of the population have prints which present difficulties to automated recognition systems. Similarly, contact lenses prevent iris recognition systems working properly. As a consequence, such technical difficulties are viewed as a uniform feature of biometric technologies.
- Intense civil liberties debate surrounds the use of biometric technologies acting as a further deterrent to commercial interest.

The use of gene-based bioinformatic identity verification systems present a number of additional problems and must be considered in relation to these and other inhibiting factors:

- Gene-based bioinformatic systems will have to conform to domestic and international legislation governing access to medical records.[37] Avoiding conformity to these aspects of legislation would require financial sector to rely upon in-house databases which would further increase cost reducing potential savings.
- Other legislation relating to medical devices will also apply to any procedure where various samples are taken from the body.[38]
- In contrast to other biometric technologies, gene-based bioinformatics relies upon a higher degree of direct contact with the body.

4.3 Forensic Bioinformatics and Law Enforcement

DNA technologies have now become an established part of evidential procedure in juridical and criminal investigation. However, many prosecutions secured on the basis of genetic evidence have been successfully challenged. Bioinformatic developments, particularly time-of-flight mass spectrometric analysis of large numbers of DNA samples, promise to extend DNA applications in a number of ways, for instance, by replacing current DNA profiling techniques with point-by-point comparisons of nucleotide bases. Whatever the accuracy of new DNA technology profiles, however, DNA forensic evidence will continue to be vulnerable to disputes over the procedures for collecting, analysing, and storing DNA material.[39] The quality of evidence is crucial in establishing the validity of DNA analysis. National Research Council (NRC - Committee on DNA Forensic Science) guidelines on the evaluation of forensic evidence are generally taken as the benchmark for European quality assurance.[40]

In addition, who has access to DNA material and on what basis (e.g., whether genetic profiles can be matched with personal information) are key central regulatory concerns. Responding to the complexities involved in DNA analysis, the US National Institute of Justice recently has established an inter-organisational working group called the 'National Commission on the Future of DNA Evidence'.[41]

Attention also needs to be directed towards organisational changes in policing brought by the increased accessibility of genetic data. As a consequence, adequate safeguards must be in place against wrongful disclosure and use of the genetic information. The use of such information for the purposes of identification, population surveillance, or population screening (e.g., political and economic refugees) may increase anxiety over the relationship between the State and the individual.

The STOA report entitled *An Appraisal of Technologies of Political Control* scrutinised existing regulations and practices of the use of personal information held by police forces in Europe.[42] The unauthorised and inappropriate access and use of (non-genetic) personal information, the sharing of data across government departments, and the use of highly inaccurate information were reported to be widespread. Although the report did not cover biometric forms of identification, the concerns it raised do pose worries for the likely future use of genetic information. Procedures for regulating police and juridical access to bioinformatic data bases need to be rehearsed and established covering areas such as:

- Restricting access of DNA information to third parties
- Controlling law enforcement handling and storage of DNA samples
- Implementation of formal, external laboratory inspections
- Regulating the exchange of data between European agencies

- The mechanisms for dealing with complaints and redress legislation, such as the Directive on Data Protection and the Directive on the protection of personal data (95/46/EC).

4.5 Bioinformatics and Defence

Increased understanding of genetic variation raises concerns that biological weapons can be developed which either enhance existing biological weapons or target particular ethnic groups or individuals. While the sequence of human DNA does not vary significantly according to racial or ethnic classifications, what variations do exist may provide a basis for targeting weapons on whole groups. The situation is complicated as the research needed to devise such weapons does not differ substantially from that necessary to develop therapeutic agents. Such research into DNA-based ethnic weapons was carried out, unsuccessfully, in South African during Apartheid.

However, a number of interviewees thought that the development weapons based on DNA profiles is far from realistic, especially since so few ethic groups have a sufficiently homogenous genetic identity. The opinion was expressed that, whilst it may ultimately be possible to develop such weapons, they will at best target only parts of groups and probably parts of the aggressors group.

> The BMA recently called for a strengthening of the Biological and Toxin Weapons Convention, including the establishment of appropriate verification provisions, in order to minimise the possibility of a new class of biological weapons being developed.[43]
>
> British Medical Association (BMA) Report on Genetics and Defence

Section Five: Public Acceptability

Sections Four outlined some of the implications of bioinformatics on clinical and non-clinical practices. The technologies underpinned by bioinformatics potentially pose fundamental challenges to established values and practices. It is not surprising that the public acceptability has been a contested area in discussions of genetics, one that has implications for bioinformatics. The clinicians contacted for this report, for instance, were sharply divided on whether or not public acceptability and ethical considerations would constrain the clinical uptake of bioinformatic-derived products and services.

A significant amount of recent work has been done on European attitudes to biotechnology and genetics and how the public comes to understand science and technology. This research raises important questions about the objectives of promoting a 'greater' public understanding. Discussions on public acceptability are sometimes framed in terms of a confused or ignorant public spurred on by a sensationalist media. Such accounts assume some deficiency in understanding by the public and prescribe some form of further education as a means of 'correcting' attitudes.[44]

The *Eurobarometer* surveys the public attitudes in member states to science and technology found that the level of knowledge is roughly correlated with levels of concern over the merits of scientific and technological developments. Therefore, it is not simply the case that attempts to promote greater understanding of science will produce positive attitudes. In-depth studies of the dynamics behind the formation of public attitudes have suggested a number of further important points:[45]

- Public understanding and attitudes are not simply affected by the presentation and communication of information, but by degrees of trust in the relevant institutions rather than amount of technical knowledge the public understands. Experts often present risk assessments as trustworthy or unbiased, while this is often the area of concern for the lay public. Consequently, claims to scientific authority on the part of government officials, corporate representatives or scientists may meet with scepticism or even hostility.
- Trust varies according to whether institutions are sufficiently policed, accountable, and responsive to the concerns of the public. So, with regard to bioinformatics, acceptability is likely to owe much to assessments made of the adequacy of safeguards to prevent the wrongful disclosure and use of the genetic information.
- Expert knowledge is ignored when it is not tailored to the needs or opportunities of particular public groups. More generally, decisions about acceptability made by members of the public are often taken in situations of conflicting claims and uncertainties. [46]

Policy Options: If public acceptance of new technologies cannot guaranteed by fostering a more 'scientifically literate' public, then alternative approaches are

required. At a general level, an alternative to literacy-oriented education of facts is providing better knowledge of the practices uncertainties that underpin science. This would include emphasising science as a collective enterprise, the limits of expertise, the contingency of scientific judgement, the role of trust. If not, otherwise normal displays of uncertainty in controversies are likely to breed cynicism. This sort of education would be more important than scientific literacy in empowering citizens to discuss and form opinions on the implications of developments in fields such as bioinformatics. Furthermore, diligence needs to be maintained to ensure organisations that handle genetic information are policed, accountable, and responsive to the concerns of the public.

Section Six: Conclusion

This report has mapped out current and likely future developments in bioinformatics and in each chapter we have identified a number of policy areas where Parliamentary intervention is suggested.

Intense information dependence is a unique feature of contemporary biology demanding a much broader approach which reconfigures the scale and capacity of local research communities or nation states. The core characteristic of bioinformatics is a trans-national form of organisation in which shared resources can be dedicated towards the storage and manipulation of huge amounts of data. Such trans-national networks must also include links with local clinical practices which would otherwise be excluded from the opportunities made available through bioinformatics.

However, exploiting bioinformatics requires more than processing data to integrate research and clinical practices, both internally and between the fields . Bioinformatics also transforms the nature of activities within trans-national networks. With the expansion, exploitation and development of bioinformatics, technologies undergo a radical shift in the way they are configured, in the way they embody specialised and expertise knowledge and the way in which they cut through conventional biological and physical barriers. Bioinformatics demands new forms of knowledge, new forms of engineering and new forms of intellectual and practical skills. The local organisation of molecular research and the configuration of organisations and traditional technologies within this field, in other words, have undergone major transformations which traverse national boundaries.

Only by virtue of managing *large* amounts of *integrated* data, bioinformatics can be brought to its full potential calling for an integrated approach which subsumes the capacity of national or local innovation regimes.

In the short term, the technological developments depend on the completion of the Human Genome Project's sequencing task and the associated, more complex work on genetic functionality and protein structures. Support for infrastructural and

training requirements here will need to be increased to address the problems public institutions will have in keeping up with the private sector. Public sector capacity will need to be maintained in the longer term not merely for research but for applications-based bioinformatics in clinical delivery. In addition, capacity will in part depend in the short to medium term on appropriate national and European-wide accreditation measures being adopted to ensure both quality assurance and harmonisation standards

Therefore, it is not difficult to see that inherent to the field of bioinformatics itself is a need to reorganise local (national) activities and integrate them across Europe. Accordingly, the composition of the EU and the Parliament as a central body, lends itself to the promotion and co-ordination of distributed research and clinical provision.

Given the necessity for a cross-national stimulation/promotion, the European Parliament still has a number of things to consider and a number of choices to make. What *type* of stimulation would it want to promote? The bioinformatics field can be characterised by a strong interdependency of both public and private actors. This interdependency is reflected in the different interests that they serve. There are, for example, some tensions between the two, - relating for example to interoperability, funding and intellectual property questions. Policy interventions, therefore, cannot presume a broad consensus on all fronts.

So far, the field has developed on a rather ad-hoc basis and has been driven mainly by incremental changes and local decision making processes. Nonetheless, the bioinformatics field in Europe has developed rather well and the Parliament may want to consider a much more central role in the management and co-ordination of it substantial bioinformatic resources. While this decentral, heuristic development has been fruitful in a rapidly moving, innovative field, its further integration and expansion may require targeted stimulation and co-ordination from central governing bodies.

The key questions for this report have centred on how to formulate a future for the European Union's engagement with the developments in bioinforamtics. How might the Parliament shape the EU's future capacity to exploit and take advantage of bioinformatic opportunities? In what way can these goals reflect an appreciation of local Member State contexts such that trans-European inclusiveness and equity is maximised. In what follows, these questions are addressed in the context of two concurrent tendencies or, what we might call, scenarios:

- ***shifts in the locus of power brokerage towards international organisations*** (European Union, United Nations, Human Genome Organisation, Operability Management Group, international research organisations, large commercial organisations, etc.).
- ***shifts in the locus of power brokerage towards local organisations*** (Non Government Organisations, devolution, patient advocacy groups,

pressure and lobby organisations, local research networks, self-organisation, micro-political trust issues, etc.).

Whilst both of these scenarios highlight tendencies by which power brokerage has been pulled both down (towards the local) and up (towards the international), neither is necessarily mutually exclusive. Indeed, the success of the EU Parliament's management of the future of bioinformatics rests firmly in reconciling the exploitation of its international scale and its capacity to reflect local conditions, demands and requirements.

Scenario One: The EU as a global actor (The EU exploits its scale and capacity in maximising the potential of bioinformatics).

In respect to its size, scale and capacity, the EU is uniquely placed to exploit opportunities which could not be met by member states acting alone. In this scenario, for the EU, pressures towards international power brokerage conveniently parallel pressures towards 'big science' in bioinformatics-related clinical delivery and healthcare research. The role of the Parliament lies in fulfilling two functions: First, creating the legislative, regulatory and incentive structures which oblige local actors (nation states, healthcare actors, research groups) to act in ways which maximise the EU-wide bioinformatics infrastructure. Second, utilising the aggregate competences of its member states in exerting influence on other prominent international organisations who are of relevance to the future of bioinformatics. This may include:

- aggregating networks of sequencing activity which both complement and yet maintain the public availability of data submitted by non-EU research communities. Maintaining the profile of EU involvement in such fora as HUGO with a view to consolidating the mutual commitment of non-EU partners (i.e. Japan, US) to the global dividends of bioinformatic activity.
- initiating trans-European healthcare arrangements whereby member states can draw upon bioinformatic resources and expertise located elsewhere in the Union.
- enforcing EU-wide statutory quality accreditation measures on EU domestic forensic and clinical DNA laboratories. In turn this will serve as a mechanism for the identification of niche expertise and enable the EU to concentrate resources and reduce unnecessary duplication. Such mechanisms will also include the means to maintain diversity thus providing some measure with which to judge performance.
- acting as the main power broker in obliging local research communities to adhere to the standards criteria of the Object Management Group Life Sciences Research Task Force.
- endeavouring to create multilateral consensus on appropriate degrees of confidentiality security which are reasonable between international partners.
- in the field of scientific research funding, stimulating and awarding integrated, transdisciplinary programmes between wet and in silico approaches, and the establishment of transnational expertise and resource

centers.

Limitations and Uncertainties: Limitations include the behaviour of other international actors in their management of sequence data, proprietary protection and research prioritisation. For example, will commercial sequencing activities (Celera, for example) concentrate the locus of the proprietary genetic economy outside the EU? Also, how will the EU, in utilising its scale and capacity, respond to those pressures emerging as power brokerage is pulled down by local actors within the Union (see below)?

Scenario Two: The EU as a local facilitator (The EU exploits its recognition of local power brokerage in maximising the potential of bioinformatics).

This scenario emphasises areas of power brokerage which the Union must take into account if it is to successfully foster bioinformatic opportunities locally as well as globally. For example, such areas of activity may be characterised by the influence of:

- new partnerships between specific interest organisations and professional groups. This can be seen in the mutual enrolment between clinical genetics services and organisations representing the interests of breast or colonic cancer patient groups in improving access to screening. Such alliances may be quite spontaneous, transitory and have immediate consequences for demand.
- the provincial historical formation of disciplines which may or may not lend themselves to external co-ordination. This will be a particular tension in the co-ordination of standardisation and quality accreditation across the Union.
- aspects of local commercial activity which may respond better to self organisation than EU-wide regulation. For example, the degree to which the insurance sector's use of genetic services is desirable or not will depend upon the healthcare context in which such practices are located.
- local constituencies who may oppose certain bioinformatic developments (whether on the grounds of confidentiality, biological risks, safety or quality), not because of inadequate technical knowledge, but because of mistrust in the motives of research communities and the regulatory frameworks to which they are supposed to respond.

As a result there are a number of implications that follow from this second scenario:

- increasingly influential pressure group alliances and the demands which they give rise to will require ready access to a EU-wide yet locally responsive bioinformatic infrastructure.
- locally developed and once relatively isolated disciplines or communities require new means of communication and voluntary consensus building. On the other hand, the EU also needs to consider exercising some degree of discretion in determining which standards are fundamentally necessary to integration and which measures may serve as over-demanding disincentives to compliance

- in instances such as the regulation of access to bioinformatic data by the insurance industry, regulatory intervention will have to discriminate between Member States where there is a high degree of social medicine (perhaps requiring little or no regulation) and those where there is little social medicine (perhaps requiring some form of external monitoring and perhaps control).
- in the formulation of measures intended to foster trust in future bioinformatic developments, the Parliament will need to observe the desire by NGOs and other public constituencies for the consultative transparency of regulatory and monitoring institutions (diagnostic laboratories, statutory bodies, the judiciary and regulatory institutions).

Limitations and Uncertainties: In this instance, the capacity of the EU to be effective in managing and maximising its bioinformatic potential will be constrained by an inadequate grasp of local and spontaneous conditions. The key question here rests upon the Union's knowledge of and responsiveness to the characteristics of such constituencies.

Annex A: Technological Developments in Bioinformatics

Genomics and Sequencing Technologies Genomics refers to the activity of locating specific genes and identifying the proteins with which they are associated. Whilst DNA refers to the basic constituent of genes which code for organic life, messenger RNA is instrumental in the production of proteins from DNA. DNA sequencing allows researchers to unravel the structure of an organism's genome and identify deleterious mutations. Protein sequencing acts as a basis for understanding the protein structures produced by DNA and ensuring that derived structures plausibly match the original DNA. Together, RNA and DNA sequencing facilitates a better understanding of disease by identifying the genetic variations present when cells and organisms develop abnormally.

A number of technological strategies are used in mapping the genomes of human and nonhuman organisms:

- *Clone by Clone* methods map the location of genes by dividing the genome into small blocks each of which is then copied repeatedly within a replicating organism (usually Escherichia coli) clone by clone.

- *Whole Genome Shotgun Sequencing* involves randomly fracturing the DNA of an organism into small fragments and then using powerful computer sequencing machines to identify the base pairs at the end of each fragment. If this is repeated time and time again, overlapping fragments can be matched allowing the system to map location of each gene. The technique has more usually been used to successfully map the genomes of small organisms but has recently been applied to the whole human genome by the US company, Celera.

High Throughput Screening and Combinatorial Chemistry

High Throughput Screening involves the use of fully automated robotic technologies to test compounds against a molecular gene target identified by genomic approaches (described above). As a consequence, large-scale screening programmes depend upon the ability of chemical science to produce huge variations in test compounds. Instead of drawing upon libraries of pre-existing known chemicals, combinatorial chemistry is used to rapidly generate vast on-line libraries of entirely novel chemical entities. Since every new compound is likely to register a different effect when screened against the protein expression of a specific gene, the pharmaceutical sector now depends upon the production of millions of chemical subdivisions and combinations in order to identify just one compound with potentially therapeutic properties.

The development of drug design can be characterised in terms of the difference between the traditional Drug discovery strategy (c1950's-80's) and Lead Discovery since the mid 1980s:

- *Traditional Drug Discovery* involved selecting a molecule for its pharmacological potential, based on animal modelling, and then adjusting that

potential by increasing the molecule's complexity. A number of limitations are evident here. First, an organic chemist would be unable to synthesise more than 50-100 compounds each year.[47] Second, the onus of discovery lay in finding a compound and then a relevant therapeutic target. Added problems arise in determining the point at which a the drug should be introduced in order to alter the course of a disease.

- *Lead and Rational Drug Design* reverses the traditional order by beginning with a clinical target instead of a promising compound. As a consequence, most pharmaceutical companies have invested heavily in bioinformatics in order to effectively mine DNA and protein databases for potential leads. Having identified relevant targets, combinatorial tools are then used to produce a high number of compound variations which can then be screened for the desired active agent.

Highly automated procedures now enable the production of as many as 100,000 different molecules in a single synthesis cycle over the course of a year. Combinatorial chemistry also plays a crucial role in refining promising molecules by drawing upon structural protein analysis (using x-ray crystallography, nuclear magnetic resonance, spectroscopy and computer modelling) to generate three-dimensional images of new biological targets.[48]

Biochips and Laboratory Chips

Building upon the same silicon etching technologies that cover microprocessors with transistors, biochips or genechips are coated with an ordered grid of target DNA molecules which may function as molecular 'tweezers' which snag target genes. Laser technologies are then used to identify the colour florescence of the grid elements indicating specifically which genes are present in the sample and which are not. In so doing, such devices are able to perform thousands of simultaneous gene identification tasks at much greater speed and with greater complexity than has so far been possible. Where once, genetic research had to confine itself to the analysis of a very small number of genes offering insights into monogenetic disorders, biochips enable researchers to simultaneously monitor the activity of thousands of genes generating indicators for complex multifactoral conditions. Affymetrix, one of the lead companies in the field of genechips, now manufactures more than twenty different types of chip used in the rapid detection of a range of pathological genes, including those associated with HIV, liver disease and the p53 mutation thought to be responsible for as much as 60% of known human cancers.[49] A number of prominent pharmacogenomic research institutions now use genechips as a matter of routine, including the National Institutes of Health, Oncomed, Merck, SmithKline Beecham, Glaxo Wellcome, Hoffman-La Roche and many others. As research into genechips intensifies, the range of gene targets which can be detected using silicon based technologies is likely to increase dramatically. While the price is likely to reduce it remains an important factor in determining access into these and associated fields of research.

Having a similar material basis, the genechip industry has much in common with the microprocessor sector. Indeed, Affymetrix leases its production site from the microprocessor manufacturer, National Semiconductor. Another parallel lies in the exponential rate at which genechip and microprocessing power seems to double every eighteen months or so. For example, three years ago Affymetrix chips contained a mere 20,000 probes. Today, the company manufactures chips with 400,000 probes (using technology jointly developed with Hewlett-Packard).

As will be discussed below and in contrast to the semiconductor sector, a defining feature of the genechip industry is the way in which it relies upon a high degree of interdisciplinarity between biological, microelectriconic and Information technology competencies. Another feature which currently distinguishes genechips from microprocessors is the size of the market. Genechips are bought exclusively by pharmacological and genomic research sectors which, in turn, limits the actual and potential number of devices sold. Developers in the sector are increasingly looking towards broadening genechip applications to follow microprocessors into more distributed local and domestic markets where the number of items sold can proliferate. Illustrating the likely impact of genechips is the industry's intention to market over-the-counter kits containing the diagnostic chip to test for a specific bacterial agent and the antibiotic developed to treat it. However, as discussed further into this report, the health care market is unlikely to be able to support genechips which cost as much as the latest Pentiums. Instead, promoters look to disposable microprocessors as a model for the way in which the sector may develop within health systems characterised by chronic scarcity. The first commercially available genechip, produced by Affymetrix to detect HIV mutations inhibiting the efficacy of drug therapies, cost roughly $100 per chip but the scanning and analytical systems which interpret the chip currently cost in excess of $200.000. Respondents and sites visits revealed that most research groups estimate that they would expect to have to spend more than $1m to buy into genechip technology.

Parallel to the development of genechips, discussed above, research has continued to focus on miniaturising the synthesis of new molecules and their screening against gene targets. Accordingly, new 'Lab on a Chip' technology promises to open up the possibility of carrying out both combinatorial and screening tasks simultaneously. The US company, Orchid Biocomputer, recently marketed a chip containing thousands of small chambers each able to sustain a different chemical test. Lab chips vastly accelerate the process of drug discovery by reducing the space within which a test is performed and the quantities of chemical reagents used. The chip is composed of successive micro-thin layers of glass, each featuring a network of tiny vessels and channels, Chemicals are moved between different locations on the chip by using differently charged electrodes to acts as 'valves' and 'pumps'. The miniaturised scale of Lab and Gene chips, and their current tendency towards technical integration, lends itself to a variety of requirements. For example, since such devices operate as closed environments, promoters claims that chip devices are likely to be less susceptible to contamination thus improving accuracy. Miniaturisation also lends itself to research where the quantities of test substances

are very small, as in forensic science for instance. As with genechips, disposable Lab chip technologies are expected to become part of the standard diagnostic apparatus of clinical delivery.[50]

Time-of-Flight Mass Spectrometry

Mass spectrometry involves the detection of ions and measurements of their mass-to-charge ratio - the method has now been developed for identifying and typing DNA samples. Three developments have been significant in the emergence of what is technically known as MALDI-TOF-MS: 1. new matrices which ionise DNA without fragmenting; 2. a new laser ionisation method (matrix-assisted laser desorption-ionization [MALDI] which is able to free the DNA sample; 3. the liberated DNA ions then travel through a 'flight tube' which measures their time-of-flight to a detector. Smaller ions will reach the detector first since they have a smaller mass. The time value is then converted into a mass value in order to assess the mass of the sample. A single plate can be mounted with thousands of samples each of which will take a matter of nanoseconds to be analysed. Unlike eloctrophoretic methods, either in gel or in capillary, MS requires no allelic ladders in order to calibrate for variations in experimental conditions. For this reason, the process is considered to be highly accurate and extremely suitable for assessing large numbers of samples from many subjects. The main application of MALDI-TOF-MS is in rapidly typing large numbers of samples for feeding into forensic databases which have to be able to type hundreds of samples on a daily basis. Traditional methods, like slab gel and capillary electrophoresis, can only accommodate small sample numbers whereas MS and robotic sample preparation can handle several thousand samples every day: 'Time-of-flight mass spectrometry offers a rapid, cost-effective alternative for genotyping large numbers of samples. Each DNA sample can be accurately measured in a few seconds. Due to the increased accuracy with mass spectromery, STR allels may be reliably typed without comparison to allelic ladders. Mass spectrometry holds significant promise as a technology for high-throughput DNA processing that will be valuable for large-scale DNA database work'.[51]

Databases

Databases differ according to the type of information stored and the conditions of access imposed on institutions seeking entry. Repositories offer either unrestricted free access or are accessible to paying customers alone. The data itself has been sourced from a wide variety of routes including scientific literature, patent applications, and from sequences directly submitted by researchers. In respect to the type of information stored, biological data bases are dedicated to the storage of both nucleic acid sequences (DNA and RNA) and protein sequence information. Some of the larger nucleic acid databases are listed in the left hand column in fig.1 below. In addition, protein data bases are important in predicting the protein sequences produced by DNA and ensuring that derived structures match the original DNA (see right hand column).[52]

Nucleic Acid Sequence Databases	Protein Sequence & Structure Databases
Proprietary GENESEQ (Oxford Molecular) Derwent (Patent Database) Incyte Genetics The Institute for Genomics Research (TIGR) Human Genome Science (HGS) Celera Inc.	Proprietary OMIGA & MacVector (Oxford Molecular) Incyte Genetics Merck Human Genome Sciences (HGS)
Nonproprietary European Bioinformatics Institute (EBI) of the European Molecular Biology Laboratory (EMBL) GenBank at the US' National Centre for Bioinformatics (NCBI) DNA Databank of Japan (DDBJ) GDB (Genome Database in Toronto) Genecards (Israel) OMIM (DAN sequence / clinical description and gene variants)	Nonproprietary Protein Data Bank (PDB) of the Research Collaboratory for Structural Bioinformatics (RCSB) in the US EMBL (Heidelberg) Structural Classification of Proteins (SCOP) CATH SWISS-PROT (using EMBL & EBI DNA sequence data) GenPept (using US GenBank DNA sequence data)

Fig.1 Bioinformatic Databases

The table above distributes databases according to whether they are predominantly proprietary (paid access) or nonproprietary. However, a number of commercial data bases have, after a proprietary period had elapsed, been made freely available on relatively nonproprietrory terms (TIGR, for example).[53] On the other hand, public access libraries like the European Bioinformatics Institute and Genbank in the US, under the auspices of the Human Genome Project (HGP) have been consistently underpinned by international support for 'free access'.[54]

In addition to those databases described above, many research institutions (particularly pharmaceutical companies) own customised copies of proprietorial and nonproprietorial libraries called 'mirror sites'. Mirror sites offer a number of advantages over on-line systems: in-house searches are confidential preventing active areas of research becoming public knowledge; direct access can be faster than online web-based access; the sale of mirror sites to commercial companies by academic-public research actors is an arrangement by which public research expenditure can be reimbursed by profit making organisations.[55]

However, the question remains open as to whether the sector has seen something of a retreat from propriatorial access agreements with commercial data base providers which may suggest greater use of publicly available data.

Visualisation Techniques and Software

Visualisation and the computer manipulation of molecular structures is particularly important in refining lead compounds following the combinatorial and screening phase of drug development. Visualisation is also essential in the automated interpretation of laboratory and diagnostic results. However, visualisation in bioinformatics refers to a highly diverse range of activities and applications, some of which are described below:

- *Information Murals* - a technique developed to view information when the data exceeds the number of pixels available on the screen. The system is currently being used by the EBI to view particularly long gene sequences (up to 0.6M bases). Information murals were originally developed for statistical and mathematical research and illustrate the traffic in IT technologies across disciplinary boundaries.[56]
- *The Information Cube* - a method of 3D visualisation allowing the researcher to view hierarchically arranged information in a semi-transparent state.
- *Multi-Scale Surfaces* - Hierarchical information is presented on an infinitely scaleable 'wall'. The advantage of these systems is that they allow the viewer to keep a global perspective of the data while examining selected regions in detail.
- *Toolglasses, Magic Lenses and Portals* - appears as a virtual sheet of transparent glass providing a layered view of the data or application underneath.
- *Algorithm Animation* - used to visualise the workings of a computer algorithm.[57]

Integration and Interoperability

Many things stand in the way of interoperability, not least the historical development of the sector. Bioinformatics systems were originally developed by relatively isolated research groups in response to local information handling problems. Established research groups have consequently exhibited a reluctance to part with their locally developed systems and their preferred vocabularies and terms of reference for compounds and genes. The degree of flexibility and openness to change by such actors is likely to be limited because of the financial cost of reorganising nomenclatures and data handling systems. If cross matching between data bases is to be as automated as is hoped, these difficulties have to be overcome.

The institutional role of various organisations including the HUGO nomenclature committee and at a more practical level, EBI, NCBI and DDBJ in creating consensus on standards is of considerable significance to the development of standardisation and interoperability. Other significant developments in the area include:

- The Object Management Group Life Sciences Research Task Force was formed in 1997 to establish interoperable standards between users of the Common Object Request Broker Architecture (CORBA) in the Life sciences.[58] CORBA was recently adopted by the EBI and illustrates its role in generating consensus on standardisation.[59]
- EU's EUREKA programme has also dedicated funds to a consortium (£450,000 from the UK's Department of Trade and Industry) to improve interoperability. The 'BioLib' consortium consists of a number of commercial and public actors.[60]
- Programme languages like JAVA are able to provide bridges between different systems thus allowing incompatible networks to communicate. However, the question ramins open as to whether the widespread adoption of JAVA as the technological solution to interoperability might generate a competitive disadvantages for European.
- Remote Database Access (RDA) Project[61] is a communications protocol for remote database access that has been adopted and promoted as an International Standard by the International Organization for Standardization (ISO).[62]
- Semantic search protocols, or 'concept searches', have been introduced which can match terms which are syntactically different whilst sharing similar meanings.[63]
- Pharmaceutical firms have invested in improving software communication within their organisations. 'Wrappers' have been developed by Smithkline Beecham to integrate otherwise isolated research systems.
- The European Bioinformatics Institute has initiated a 'BioStandards Project' as part of its Industry Support program to disseminate solutions to interoperability amongst its industry partners. BioStandards, is funded jointly by the EBI, the EC (under their DGIII Information Society Standardisation programme), and a number of leading European pharmaceutical companies participating in the project.

Annex B: Key Actors

Research Institutions

Section One above introduced the different types of technologies and databases and some of the major institutions involved in bioinformatics activity. Here we extend that discussion by focusing on the institutional actors themselves and their different access and licensing conditions.

European Molecular Biology Laboratory (EMBL): Established in 1973, the EMBL is a mainly European, intergovernmental organisation dedicated to basic research, training and the provision of services in molecular biology.[64] Though the EMBL headquarters is located in Heidelberg, its three outstations are located in Hamburg (FRG), Grenoble (France) and Cambridge (UK).[65] The European Bioinformatics Institute (EBI) in Cambridge, discussed below, was the last of the outstations to be formed.

The European Bioinformatics Institute (EBI): The (EBI) is located at Hinxton Hall, near Cambridge (UK) and is described by the EMBL as 'a world centre for management of sequence databases and cutting-edge bioinformatics research'. Its establishment in 1993 was funded by EMBL with support from the Wellcome trust and the British Medical Research Council (MRC) to finance the present Data Library. The EMBL's Nucleotide Sequence Database, produced in collaboration with GenBank and the DNA Database of Japan (Mishima), is Europe's primary nucleotide sequence data resource. The EBI maintains the TREMBL and SWISS-PROT databases as well as various other databases. DNA sequence data is exchanged on a daily bases with GenBank and the DNA Databank of Japan, all providing the same sequence information. As a nonproprietary resource, the EBI database is free of charge. The EBI has also launched an Industry Support Programme, with the overall aim of helping industry to adapt quickly to, and maximise the benefits from, developments in bioinformatics. In co-operation with UNI-C (Denmark) and CASPUR (Italy), the EBI has served as a bioinformatics Technology Transfer Node (TTN), an intermediary between the EU, industry and academia.

Hamburg (FRG): The EMBL Hamburg Outstation is situated on the Deutsches Elektronen-Synchrotron (DESY) site and was founded in 1975. The role of the Hamburg Outstation is dual: to carry out an internal research program in molecular structural biology by utilising the special properties of the synchrotron radiation (SR) provided by DESY; and offer its SR beam lines to the international user community in molecular structural biology.

Grenoble (France): Soon after the 1975 Hamburg accord, EMBL signed a parallel agreement with the Institut Laue Langevin (ILL) in Grenoble to establish a second Outstation on the site of the world's leading research nuclear reactor. In 1976 EMBL began its fruitful collaboration with the ILL in developing neutron scattering techniques and instrumentation. Since that time, it has been EMBL's responsibility to provide biological support to visiting scientists doing measurements using these beams, which are especially suited to structural studies of the role of water in

biological systems, the dynamics of proteins, and protein-nucleic acid or protein-lipid complexes.

The Sanger Centre: Based at Hinxton Hall (the location of the EBI and the UK Human Genome Mapping Project Resource Centre (HGMP-RC), the Sanger Centre is a genome research centre founded in 1992 by the Wellcome Trust and the Medical Research Council (MRC).[66] Having contributed as much as a third of the total sequences to the HGP, the Centre is the largest contributor to the Human Genome Project.[67] All the centres data is made freely available and has assumed an institutional role in promoting free access. The Centre's provision of genomic information takes three forms: as sequence-ready maps, as assembled shotgun sequence data, and as finished and annotated consensus sequence of each bacterial clone.[68] Sequencing work which is partially complete is also released on a daily basis. The main concern is that the time between the generation of a sequence and its public availability serves as a window of opportunity for other, possibly proprietorial, institutions to make a claim on the sequence data: "It is not adequate to rely purely on release of the finished sequence of each clone as an indicator of progress; the risk [of duplication] is minimised by providing regularly updated maps of all clones as soon as they enter the process, if not earlier still".[69]

SWISS-PROT: One of a number of "centres of excellence" established by the Swiss government, SWISS-PROT opened in 1986 at the University of Geneva. More recently, SWISS-PROT has become a collaborative partner of the European Bioinformatics Institute[70] and the Swiss Institute of Bioinformatics (SIB)[71] providing detailed annotation and organisation to (mainly public) curated protein sequences, rather than holding primary information. While academic users have free access to products that are partially or completely funded by public grants, license fees are sought from commercial users. These ownership relations do not extend to nucleotide sequence data, which is assumed to be primary data which should be in the public domain. In practice though, this distinction must be complemented by a variety of other regulations regarding access to information and a recognised pragmatism in how distinctions are made between commercial and academics users. The question of whether a licensing fee is required is assessed on a case-by-case basis. Also, a license fee may also apply to subsequent derivative products.

Celera Genomics: Celera is a joint venture between J.Craig Venter, former director of the US National Institute of Health (NIH) and Perkin-Elmer Inc. Whilst at the NIH, Venter developed the methods which uses expressed sequences tags (EST's), used for whole genome shotgun sequencing (see section one), a method which at that time the NIH was unwilling to sponsor. Venter subsequently left to establish the non-profit making Institute for Genomic Research (TIGR).[72] Already mentioned in the discussion on shortgun sequencing in Section One above, Celera uses fully automated shotgun sequence technologies (the ABI Prism 3700) in attempting to complete its version of the human genome by the year 2001. Although the company intends to make the data publicly available, profits will be generated from

proprietary protection of the raw data and annotation parts of the database and patenting of some sequences. A number of concerns were raised in respect to the accuracy and rigour of the endeavour and these are more thoroughly discussed in Section One (Genomics and Sequencing Technologies). Initially it was hoped that the project might complement the International human genome efforts. However, the latter has proved to be difficult since co-operation between Celera Genomics and the federal Human Genome Project seems to be extremely difficult. In a hearing for the House Subcommittee on Energy and Environment both stated that although both projects can be seen as complementary, formal collaboration has so far had been difficult to achieve.[73]

Incyte Pharmaceuticals Inc.:[74] Incyte, a US based company, recently obtained a European office through the appropriation of Hexagen, the Cambridge based company Incyte took over in August 1998. Hexagen is now part of Incyte's new pharmacogenetics business unit, Incyte Genetics which is involved in gene-mapping, polymorphism discovery, genome sequencing and pharmocogenetics. Last year, about 20 companies, including GlaxoWellcome and Zeneca, took out paid subscriptions to the Incyte Database.

Human Genome Sciences (HGS): [75] HGS was originally the parent company of The Institute for Genomic Research (TIGR), discussed above and has provided profitable data access to a number of large pharmaceutical firms. Both HGS and Incyte, as well as Celera, use the expressed-sequence tagging method to complete their different genome libraries. However, their product is not so much the sequencing of the genome (something which is said they could do within a year) but instead concentrates on gene expression, something that is more profitable than sequencing. Until 1997, the company had secured a lucrative agreement to provide exclusive sequence data access to Smithkline Beacham.

EU as Bioinformatics Funding Actor

Throughout the 1990s, various European-level funding programmes have directly or indirectly supported bioinformatics research.[76] The DGXII Life Sciences and Technologies domain of the RTD Framework Programmes have been the most relevant.

Initial activities in the late 1980s to establish a suitable research infrastructure in Europe for genomic research led to the formation of the Human Genome Analysis Programme in 1990 under the Second Framework Programme. With a budget of 15.6 million ECU, the Programme set up study groups including:

- genetic (linkage) mapping and physical mapping
- data handling and databases
- training
- ethical, social and legal aspects

This included establishment of the collaborative European Human Genetic Linkage Mapping (EUROGEM) network providing training workshops in both informatics and laboratory techniques, to broaden the knowledge base in the Member States.

The Biomedicine and Health programmes, BIOMED 1 (1990-1994, ECU 150 million) and BIOMED 2 (1994-1998, ECU 374 million), have supported medical research activities in line with the priorities of the Council. These programmes aimed to improve medical knowledge and the health of the European population as well as strengthening the competitiveness of the European health industry. In particular BIOMED I focused on:

- Integration of physical and genetic linkage maps
- Mapping cDNAs in better understanding diseases development and treatment
- improvement of data handling and analysis

By BIOMED 2, the emphasis of the programmes shifted from infrastructural issues to enhancing mapping facilities and information management including an area (5) on Human Genome Research (see table 2).

- Gene mapping and analysis to provide a sequence-ready fragments of the complete human genome and to identify all genes with their regulating sequences as well as the non-coding elements with functional relevance, including research aimed at the exploitation of comparative approaches – Brain Research
- Construction of maps ready for transcript identification and sequencing
- High throughput sequencing methods
- DNA and chromatin elements of functional relevance other than genes
- Information management and analysis
- Development and application of database, and software technology for the management, integration and sharing of genome data
- Theoretical models for the analysis and understanding of genome data
- Development of software to facilitate experimental genome research
- Forums for communication and exchange of global data

Fig 2. BIOMED 2 Research Tasks of Area 5, Human Genome Research

Bioinformatics underpins developments in genetic understanding beyond those specific to humans. The BRIDGE programme (1992-94) on biotechnology and the 1994-1998 Biotechnology Programme (BIOTECH 2**,** ECU 595.5 million) aimed to increase the knowledge of biological systems for increase productivity in agriculture, industry, health, and the environment.

- Sequencing the yeast genome - Sequence data co-ordination. BIOTECH 1
- European scientists sequencing Arabidopsis. BIOTECH 1
- Continuation of the EMBL Data Library and Upgrade of the International Protein Sequence Databank. BIOTECH 1
- Provision of EMBL data library services. BIOTECH 1
- Permanent inventory of EU biotechnology research projects. BIOTECH 1
- Computer Network for EU Bioinformatics in Europe. BIOTECH 1
- Registry of sequenced plant genes. BIOTECH 2
- A comprehensive and integrated mitochondrial database. BIOTECH 2
- Methods/software for evolutionary genome analysis. BIOTECH 2
- Advanced database linkages in biotechnology. BIOTECH 2
- Provision of EMBL data library and related European IT services. BIOTECH 2
- Integrated protein sequence database. BIOTECH 2
- Database Linkage (CORBA). BIOTECH 2
- Common access to integrated immunogenetics database. BIOTECH 2
- A European database of biological volume images BIOTECH 2
- Sequence of divisions 1-3 of the Drosophila genome BIOTECH 2
- From the structure and function to the design of modular proteins BIOTECH 2
- Common access to biotechnological resources and information BIOTECH 2

Fig 3. Bioinformatics related projects under BIOMED and BIOTECH programmes

Although the Fifth Framework (1998-2002) has not been operationalised in detail yet, bioinformatics should play a key role in the *First Activity: Quality of Life and Management of Living Resources*. Bioinformatics underlines the 'Key Actions' (which includes areas such as the cell factory and the ageing population); 'Research and Technological Development Activities of a Generic Nature', as well as 'Support for Research Infrastructures'.

Pharmaceutical firms as BI actors

- ***Database Subscriptions***: Most major pharmaceutical firms have formal arrangements for accessing and sourcing bioinformatic repositories from both private and public sources. As mentioned above, this includes in-house copies of databases permitting speedy and confidential access to commercially valuable information. Pharmaceutical subscriptions to the Incyte databases in 1997 numbered 17, each paying as much as $15-25m per annum.[77]
- ***Managing Information Volume***: SmithKline Beecham recently withdrew its contract with the Human Genome Science database (worth $140m) because the company needed more time to assess the significance of the data that it had 'mined'. This illustrates the towering problem of keeping analytical pace with the relevance of new data and the newly emerging tendency for research establishments to shift their resources from sequencing to developing compounds from specific genomic targets.[78] However, whilst the use of public databases is increasing, it does not necessarily follow that research institutions view public data bases as necessarily better than private and commercial ones.

- ***In-House Software*:** A small number of companies have sought to develop in-house software responses to bioinformatic problems. This applies particularly to the integration of various data search and modelling tools. The capacity to develop tailor-made in-house information management systems is considered to be restricted to particularly large firms.[79]
- ***Wider Aspects of Pharmaceutical R&D*:** Whilst bioinformatics promises the acceleration of drug discovery, other aspects of R&D have continued to restrict the speed of drug innovation. For example, drug validation and animal testing via 'wet biological' laboratory technologies mean that the lengthy time-frame of the R&D life-span is unlikely to change significantly. Whilst such 'bottle-necks' may continue to characterise the drug development process, cost savings can be found in the use of bioinformatics to reduce wasteful investment in unproductive research paths.[80] However, improved simulation techniques (in silico) offer flexibility in adherence to 'wet biological' aspects of R&D.[81]
- ***Integration of IT and Pharmacological business properties*:** The ability of pharmaceutical and IT firms to align with one another is central to the effective use and application of bioinformatics. The defining features of both constituencies have been recognised as being distinct in respect to a number of key variables:

Variable	**Pharmaceutical Sector**	**IT Sector**
R&D Cycle	Long (10+ yrs)	short (6-18months)
Product Life Cycle	Long (10-15yrs)	Short (1-3 yrs)
Investment	Very High	Variable
Company Size	TransNational Dominated	SME Dominated
Failure Rate	Very High	Low
Regulation	Very High	Low
Product Market	Global	Niche
Product protection	Mainly Patents	Copyright (limited protection)
Profits	Early Losses	Early Remuneration
Drivers	'Discovery'	'Application'

Fig 4. Organisational characteristic of the Pharmaceutical and IT Sectors

Annex C: Regulatory and Advisory Measures

Statutory Legislative Measures

Directive on the Legal Protection of Data Bases: With the capacity for the storage of tremendous amounts of information presented by recent advancements in bioinformatics, regulatory attention has focused on the ownership and exploitation of data. The European Community recently increased the level of protection for databases. The most pertinent legislation for bioinformatics being the Directive on the Legal Protection of Databases. This Directive has sought to harmonise copyright laws on EU based electronic and paper-based databases. It grants rights

on both the structure and the content of databases to a length of 70 years after the death of the author. Member states will be able to grant certain exemptions for private use and teaching purposes. It also provides a new exclusionary right protecting those deemed to be the 'makers' of databases. The new *sui generis* protection relates to the extraction or use of parts of databases and will last 15 years from the finalisation of a database and will be available even if the contents do not qualify for pre-existing copyright protection.

Directive on the Legal Protection of Biotechnological Inventions: In 1998 the European Parliament approved the Directive for the Legal Protection of Biotechnological Inventions which seeks to harmonise the individual patent laws of Member States in this area. It extends patent protection to a range of biological inventions and specifies rules relating to the deposit of biological material at depository institutions. In addition, it specifies certain processes and products that cannot be patented on morality grounds including the human clones, genetic modifications of animals which are likely to cause unnecessary suffering, and commercial uses of human embryos. Since the commercial motives behind investment in bioinformatics are linked to the potential patentability of genes, the Directive serves is of considerable importance to EU bioinformatic stakeholders.

Medical Devices Directive 93/42/EEC: The directive sets general requirements relating to the design, construction and general safety of medical appliances *'for alleviating the effects of handicap or illness'*. Whilst excluding control via *'pharmacological, immunological or metabolic means'* the directive applies to 'any instrument, apparatus, appliance, material or other article, whether used alone or in combination, including the software necessary for its proper application'. Full compliance across the EU has been mandatory since June 13th, 1998. As such, the directive applies to many potential devices and products associated with bioinformatics and will be a condition of the sale of those products within the EU. The directive acts as the principal regulatory context for emerging 'non-in vitro' technologies and is therefore likely to exclude genechip-related technologies though it may apply to other areas of bioinformatics. The directive is one of a series of three directives (see below).

Active Implantable Medical Devices Directive 90/385/EEC: The directive applies to '*any medical device relying for its functioning on a source of electrical energy or any source of power other than that directly generated by the human body or gravity'* and '*which is intended to be totally or partially introduced, surgically or medically, into the human body… and which is intended to remain there after the procedure'*. Examples of embodied devices include pacemakers, neural stimulators and will apply to emerging bioinformatic devices intended to monitor real-time changes from within the human body. The directive has been in full force since Jan 1st 1995.

In Vitro Diagnostic Medical Devices Directive: The directive covers a wide range of devices an is intended to assure safety and quality to the degree stated by both

medical directives discussed above. Examples of the technologies covered include reagent products for the detection of markers in human specimens. It also applies to self testing technologies for the measurement of, for example, blood sugars etc. In view of the clinical and nonclinical products discussed in sections 3 & 4, the IVDMD is likely to be particularly important in regulating bioinformatic technologies where human samples are taken for diagnosis.

European Medicines Law: Pharmaceutical products across Europe are regulated by the European Medicines Evaluation Agency which has, to date, authorised the marketing of more than 60 drugs in the EU. Products associated with new bioinformatics will take one of three routes into approval within Europe. First, the centralised procedure which is mandatory for biotechnological products and optional for new medicinal products. Second, companies can pursue approval through domestic regulatory mechanisms and then apply for approval in other member states through the Mutual Recognition Agreement. Finally, products developed for marketing within a single country can be regulated by that country's authorisation procedure alone.

Directive on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data 95/46: Implemented in Oct 1998, the directive seeks to regulate the exchange of data between mainly commercial parties in respect to the preservation of personal privacy. The Directive also specifies that information flows from Europe should only be with nations that have adequate privacy protection. As a consequence, the measure generated some anxiety particularly in the US where intra-corporate trans-national data flow tends to be more self-regulated. Whilst the context of the regulation lies more in electronic commerce and direct marketing, it would also apply to any movement of clinical data where identification of the individual is possible (either coded identity or named identity). It does not apply to completely anonymised data, the form in which most bioinformatic data is currently exchanged. The Directive requires controllers to inform subjects of the purpose of the personal information and the identity of any other controllers with access to the data. If data is put to a second use, subjects must be offered the option of withdrawing their data. This would apply to the processing data for the purposes of preventive medicine, medical diagnosis, the provision of care or treatment, or the management of health care services. Subjects must also be permitted to have access to their records with the right to amend and revise stored information.[82] Exemptions from the directive are extremely limited but include legal or public interest grounds (see above on Law Enforcement).

NonStatutory and Advisory Measures

BIOMED-II (1994-1998): The BIOMED programmes include a number of issues that cross-cut through all research subjects including the Ethical Legal and Social Aspects (ELSA) of the Life Sciences and Technologies. The objectives of the ELSA programme are to analyse the ethical and social issues raised by specific applications of biotechnology as well as biomedicine and health research in view of

their being taken into account in public policy deliberations. Particularly relevant to bioinformatics, research focuses on questions of confidentiality and privacy in medical data, personal data protection, data bases and (different European) ethics committee, intellectual property rights, ethics of prevention and the ethics of insurance, for example. The programme has several mechanisms for dissemination and implementation of its results. The ELSA implementing unit within the European Commission for example is charged with supervising the ethical review of all scientific proposals by local ethics committees. Also, if appropriate, working groups are set up to advice the Commission and to report to the Council and to the European Parliament. [83]

EUROSCREEN 1 & 2*:* In 1992 the European Commission, under the Human Genome Analysis programme, funded a number of studies on ethical issues. One of these was carried out by the core group currently co-ordinating EUROSCREEN. Subject of this programme was the Ethical Implications of Human Genome Analysis for Clinical Practice in Medical Genetics, with special reference to genetic counselling.[84]

Euroscreen 1: *Genetic Screening and Predictive Medicine: Ethical and Philosophical Perspectives, with special reference to multifactorial diseases*.

- Monitor the development of genetic screening in Europe
- Recommendations concerning ethical criteria for the introduction, conduct and evaluation of genetic screening programmes
- Recommendations for addressing the public on genetic screening
- Recommendations on data legislation for genetic registers and biological banks, with reference to insurance, employment and legal procedures

Euroscreen 2: *Genetic Screening and Testing Toward Community Policy on Insurance, Commercialisation and Promoting Public Awareness*:

- The ethics of insurance
- Commercialisation at the point of service to the client
- Programmes for public education and raising awareness

Fig 8. Euroscreen Programmes

The Human Genome Organisation (HUGO): HUGO's Ethics Committee, recently released a statement on the control over and access to DNA Sampling addressing the key question of how data can be protected.[85]

The Committee issued the following recommendations on samples:

- Consent procedures should be clear on the potential use of the DNA samples indicating whether the sample and its information will: *identify the person, code the identity, or anonymize the identity*.

- Samples may be re-used if: there is notification of such a policy, the patient has not objected, and the sample has been coded or anonymized. Samples obtained and stored before such notification may be used if the sample has been anonymized.
- Security mechanisms must ensure the desired level of confidentiality.
- Desired levels of confidentiality must balanced against an awareness of the way in which they may impede retrospective validation or inhibit prospective therapeutic research.

The HUGO IPR Committee issued the following recommendation on patenting:
'HUGO… is worried that the patenting of partial and uncharacterised cDNA sequences will reward those who make routine discoveries but penalise those who determine biological function or application. Such an outcome would impede the development of diagnostics and therapeutics, which is clearly not in the public interest. HUGO is also dedicated to the early release of genome information, thus accelerating widespread investigation of functional aspects of genes… it would be ironic and unfortunate if the patent system were to reward the routine while discouraging the innovative'' [86]

Fig 9. HUGO Ethics Committee

Annex D: Study's Research Instruments

Site visits, seminar events and interviews:

A number of site visits and interviews were used both to inform the questionnaire rationale detailed below and verify findings made in the Interim Report. These included:

European Bioinformatics Institute: (2 separate visits) 3 interviews conducted during the second visit. The Sanger Centre and The UK HGMP Resource Centre were also included in the first visit.

Leiden University Medical Centre: 5 interviews were conducted with scientists from the public health and forensic laboratory.

Glaxo Wellcome: 4 interviews were conducted with senior members of R&D.

The SATSU team were invited to participate in the Genetic Interest Group's seminar on *Healthcare and the new Genetics in Britain and Germany* (27-29 Nov, 1998).

The SATSU team were invited to participate in a seminar event entitled *Combining Biotech and IT for the 21st Century* run by the Eastern Region Biotechnology Initiative.

Questionnaires

The Pharmaceutical Sector: 20 of the world's leading pharmaceutical firms were included in the questionnaire survey, with respondents being chosen for their responsibilities in bioinformatics.

Firms approached:	20
Firms responding:	9
Respondents identified:	25
Responses received:	10
Firms Responding:	AKZO Nobel Astra British Biotech Bender & Co Glaxo Wellcome Janssen Pharmaceutica Novo Nordisk SmithKline Beecham Synthelabo

Questions:

Commercial and Practical Constraints:

To the following questions, respondents were asked to indicate whether they: *Strongly Agreed; Agreed; Disagreed; Strongly Disagreed; Uncertain.* Respondents were asked to indicate their preference for both a <5 and <10 year time frame.

- The acquisition of bioinformatic personnel will improve as a result of new University courses.
- Your company's expenditure on bioinformatics will remain the same in real terms.
- The subscription cost of sourcing genetic data will reduce considerably.
- The use of public databases is likely to increase considerably.
- The cost burden of drug development will increasingly shift from bioinformatics to clinical trials.
- It will become easier for SME's to afford bioinformatic capacities and competencies.
- The speed of drug innovation is likely to increase significantly as a result of bioinformatics.
- The benefits of bioinformatics will continue to be overshadowed by the costs of regulatory considerations.
- Bioinformatics will significantly reduce the cost of drug approval .
- Bioinformatics is likely to replace wet biological verification of therapeutic value in animal models.

Respondents were asked to comment on the following:

- Is EU legislation on Gene Patenting more favourable to the your sector than in it is in the US?
- Is EU legislation on Data Protection more favourable to your sector than it is in the US?

- Is EU legislation on Drug Regulation more favourable to your sector than it is in the US?
- Will existing regulatory institutions cope with changes in the volume and complexity of new drugs?
- Over the next 5yrs, roughly by what percentage will bioinformatics increase the speed at which: Clinical leads are identified? Drugs enter regulatory assessment? Drugs receive regulatory approval? These changes are a direct consequence of developments in bioinformatics?

What problems relate to interoperability?

- 1. in your organisation? 2 Between you and other institutions?
- What initiatives have contributed most to improved interoperability?
- How many bioinformaticians does your organisation employ?
- By how many is this likely to increase over the next five years?
- Roughly what percentage have a health and life science back ground?
- What proportion of your bioinformaticians received their main training in a public research establishment?
- What proportion of your R&D budget is spent on bioinformics?
- What percentage of your organisation's software tasks are outsourced?

To the following questions, respondents were asked to indicate whether they: *Strongly Agreed; Agreed; Disagreed; Strongly Disagreed; Uncertain.* Respondents were asked to indicate their preference for both a <5 and <10 year time frame.

- Your organisation will soon achieve interoperability across all internal R&D systems.
- Biology and computational science will continue to speak different languages.
- In the future, interoperability between research institutions will be a non-issue.
- A Windows-type common platform, dominated by one or more monopoly actors, will characterise bioinformatics.
- The bioinformatic software sector will come to be dominated by a single monopoly actor.

Respondents were asked to comment on the following:

- What percentage of sequence data is sourced from proprietorial databases? … from public databases? What is the annual budget for subscriptions to private data repositories per annum? (Please specify currency) Does your organisation possess in-house mirror sites of sequence data repositories?

Issues relating to data access:

To the following questions, respondents were asked to indicate whether they: *Strongly Agreed; Agreed; Disagreed; Strongly Disagreed; Uncertain.* Respondents were asked to indicate their preference for both a <5 and <10 year time frame.

- R&D will increasingly depend on making arrangements to access public health genetic registers - as with the recent Icelandic case.
- Proprietorial restrictions on sequence data will significantly restrict drug R&D.
- Public data sites will be more significant to drug R&D than private data sites.
- Sequence data from public (nonproprietary) sites will become less reliable and useful than from private/commercial sites.
- Subscriptions to private data repositories are commercially more valuable than mining public sites.
- The problem of processing and absorbing data will increasingly move your organisation towards public data repositories as opposed to private/commercial data providers.

Expected changes in relation to the nature of health delivery

- What (percentage) proportion of your present drug portfolio can be directly attributed to lead discovery from genetically sequenced data?
- Individualised treatment regimes will become commonplace as a consequence of greater diagnostic capacity.
- Most pharmaceutical products will be sold in combination with a diagnostic.
- New diagnostic 'kits' will significantly reduce the need for a doctors' diagnosis.
- A much greater number of compounds will flood the market in response to individualised compounds are introduced.

Expected changes in the cost of health delivery

- The cost of pharmaceutical products will increase as a consequence of intensive investment in bioinformatics.
- Bioinformatics will assure a much lower attrition rate in drug R&D.

Impact on domestic health care factors

To the following questions, respondents were asked to indicate whether they: *Strongly Agreed; Agreed; Disagreed; Strongly Disagreed; Uncertain.* Respondents were asked to indicate their preference for both a <5 and <10 year time frame.

- Bioinformatically-derived products and services will increase the need for locally (as opposed to globally) targeted products.
- Bioinformatically-derived products and services will be diffused more easily in the US than in the EU.
- The uptake of bioinformatically-derived pharmaceutical products across member States of the EU is unlikely change in the future.
- The EU is a legislatively/regulatory more benign environment for such products than the US

Public and proprietary database providers: This questionnaire was constructed in order to collect information on the expectations of database providing constituencies, one of the main constituencies in the field of bioinformatics. A distinction was made between proprietary and public database providers in order to allow us to comment on the their different outlooks. In respect to space constraints, the questionnaire is not included here but was based upon themes raised in the questionnaire for the pharmaceutical sector above.

Public database providers: 33 questionnaires were sent out to public data-base providers. Their names and addresses were collected through the internet. This list included, among other things, people from the European Molecular Biology network's EBM Nodes[87] but also names from individual researcher groups listed in publications and so on.

14 responses were received (and one R.T.S.)

Names of the Organisations:

LION Bioscience

Human Genome Mapping Programme - Resource Centre

MIPS
Institute Gulbenkian
Weizmann institute of Science
Sanger Centre
CNB CSIC
Swiss Institute of Bioinformatics
BioBase Denmark
Vienna Biocenter (University of Vienna)
Institute of Biochemistry and biophysics. Pas.
CNR-AREA (Universita' Degli Studi Di Bari - Italy)

Proprietary database providers: 7 questionnaires were sent out to proprietary database providers of which there were two in the US and three in the UK. Given the limited number of companies active in this field (especially in the UK) more than one questionnaire per company was sent out.

5 responses were received.

Names of the Organisations:

Incyte Genetics (UK) - two respondents
KIVA genetics. Inc.
University of Leeds
One anonymous

Clinicial Genetics Constituency: The clinical geneticists sample was determined by identifying key members of the European Society of Human Genetics (ESHG), a group dedicated to the development of research, education, and medical applications in human genetics. The sample included EU members of the ESHG Board as well as country contributors to a special edition of the *European Journal of Human Genetics* (the official journal of the Society) on genetic services in Europe.[88] In respect to space constraints, the questionnaire is not included here but was based upon themes raised in the questionnaire for the pharmaceutical sector above.

In total, 11 out of 26 questionnaires were returned with a response rate of 42%.

[1] Http://www.incyte.com/
[2] Human Genome Sciences (HGS) initially also funded TIGR (The Institute for Genomic Research), a not-for-profit research institute but parted from TIGR in 1997.
[3] Boguski, Mark (1998) 'Bioinformatics - a new era', *Bioinformatics* - trends supplement.
[4] See for example Liebman, Micheal (1995) 'Bioinformatics: an editorial perspective', *Netsci*, October.
[5] Seer for example : http://www.bioplanet.com/chat/jobs/index.html; http://www.TechFak.Uni-Bielefeld.DE/bcd/ForAll/welcome.html
[6] *Science*, 273. 12 July, 1996. p265; *Science*, 272. 21 June, 1996.
[7] *The Evolution of Bioinformatics*, Steve Gardner, Paolo Zanella, Tom Flores (Synomics Ltd [www.synomics.com/about/ev_bio.htm]).

[8] New Scientist, 18 October, 1997. pp36-40.
[9] See Martin, P and Thomas, S. (1996) The Development of Gene Therapy in Europe and the United States: A comparative Analysis (Science Policy Research Unit).
[10] Hopkins, M (1998) *An Examination of Technology Strategies for the Integration of Bioinformatics in the Pharmaceutical R&D Process*. MSc Dissertation for SPRU, University of Sussex.
[11] 75% of respondents from public databases felt the blurred division between academic and commercial research within many of today's entrepreneurial universities advantages them over commercial companies. It was commented it is hard to make complex, robust products in academic settings and that most often academic developers quit their organisation to join industry, in which case the later benefit from the academic knowledge and experience. However, none of the respondents from proprietary databases agreed with this point.
[12] 'UK academics, researchers, launch firm to meet bioinformatics outsourcing demands' *Bioinform News Service* ,1998. 2 (23).
[13] Bioinform News Service. 1998. 'Bioinformatics experts' start-up will integrate research systems' *Bioinform News Service* 2(17).
[14] Figures of 1/5 of the current price were mentioned.
[15] However, this was at the same time contra-argued by others. As one of the respondents from a proprietary database put it: "*This is often misunderstood by academics. The patent process does not need to delay dissemination of information into the public domain; most collaborations I have established between industry and academia have promoted rapid publication allowing 30 to 90 days to file for patent. The academics can then have the freedom to use the information for academic research purposes, whilst the company can seek commercial protection from business exploitation of inventions. The problem arises when data are to be maintained as trade secrets to ensure their commercial value. This typically has applied to raw genetic data where utility is undefined and so patentability is uncertain. After a proprietary period of 1-3 years much of these data will come into the public domain through other efforts."*
[16] Within the pharmaceutical companies, when asked whether the use of genetic health registers would be typical: within the next five years 6 agreed, 1 disagreed and 1 was uncertain; within the next ten years, 1 agreed strongly, 6 agreed and 1 was uncertain. Among public database providers, 2 disagreed R&D would benefit from genetic registers; 2 were uncertain and 8 agreed. Among proprietary database providers, 2 agreed while 2 didn't fill in this question.
[17] Science 280(5369), 1540-1542
[18] Promoters of shotgun sequencing discuss the application of the method to the human genome in *Science*, June 5 v280 n5369 p1540-3. See also, *Science News*, May 23, 1998, v153 n21 p334(2)
[19] Janson, Marie (1998) *Genetic Testing*, CEST report
[20] In some European countries predictive screening for a range of cancers is already in place and exposed to overwhelming demand from families. Harris, R. and M. Reid (1997) 'Preface: three Principles', *European Journal of Human Genetics*, 5(suppl 2): 1-2, p.1-21.
[21] Harris, R. and M. Reid (1997).
[22] This is obviously depending on the political, cultural and social contexts of different countries. See also Smith, R. (1997) 'The future of healthcare systems', *British Medical Journal*, 24 May. Vol.314, p.1495-1496.
[23] *Ibid*.
[24] The issue of confidentiality is not just related to bioinformatics itself but fits in with a broader concern about public databases. An other database that is currently the subject of debate therefore is the Electronic Patient Record.
[25] As is the case with Clozaril, a schizophrenia drug which induces a life-threatening blood disorder in about 2% of patients.
[26] NRC Handelsblad, Zaterdag 13 December, 1997.
[27] Nelis, A. (1998) *DNA-diagnostiek in Nederland*, Enschede, Twente University Press; See also, Royal College of Physicians (1996) *Clinical genetics services into the 21st century*. Report from the clinical genetics committee of the Royal College of Physicians, prepared on behalf of the committee by P.S. Harper., H.E.Hughes., and J.A.Raeburn.
[28] Marshall, E. (1994), 'Genetic Testing Set for Takeoff', *Science*, vol. 265, 22 July, p.464-467.
[29] See http://www.uclan.ac.uk/facs/ethics/esintro.htm
[30] bid.

[31] MacDonald, A. (1997) How Will Improved Forecasts of Individual Lifetimes Affect Underwriting? *Philosophical Transactions of the Royal Society*, vol 352, p1067-1075; Genetics Forum Opinion Poll, *Spice of Life*, April 1997, no 5, vol 5.
[32] Human Genome Advisory Commission (Dec 1997), *The Implications of Genetic Testing for Insurance.*
[33] *Time*, Jan 11, 1999, v153, p60-1
[34] (NIH-DOE ELSI Working Group and National Action Plan on Breast Cancer Workshop on Genetic Discrimination and Health Insurance. Genetic Information and Health Insurance Report of the Task Force on Genetic Information and Insurance May 10, 1993 NIH-DOE Working Group on Ethical, Legal, and Social Implications of Human Genome Research. Council for Responsible Genetic (US)
[35] *Computerworld*, Dec 14 1998, pp89-91; *Electronic Engineering Times*, Dec 7, 1998 p57-61; Bio-Identities, *PC Magazine*, Jan 9, 1999, p10; In the US, the current market is put at around $25m excluding the much larger market in law enforcement (estimated at $120m), *Forbes*, Aug 10, 1998 v162 n3 pp110-1
[36] Biometrics Doesn't Quicken Corporate Pulses, *PC Week*, Jan 4, 1999, p61.
[37] EU: Council of Europe. Convention for the Protection of Human Rights and Dignity of the Human Being with regard to the Application of Biology and Medicine. Strasbourg: Council of Europe 1996 (ETS 164). UK: Access to Medical Reports Act, 1988 (HMSO).
[38] European directive on medical devices. (93/42/EEC).
[39] Lynch, Michael. 1998. ÔThe discursive production of uncertaintyÕ *Social Studies of Science* 28 (5-6) pgs. 829-68).
[40] National Research Council, Committee on DNA Forensic Science: an update. *The Evaluation of Forensic DNA Evidence*. National Academy Press, Washington D.C. 1996
[41] McDonald, R. (1998) Juries and crime labs: correcting the weak links in the DNA chain. American Journal of Law & Medicine, Summer-Fall v24 n2-3 p345-363
[42] Wright, Steve. 1998. *An Appraisal of Technologies of Political Control.* Report to the Scientific and Technological Options Assessment of the European Parliament. PE 166 499
[43] British Medical Association. 1998. *Biotechnology, Weapons and Humanity*. London : BMJ Publications.
[44] Some scientific institutional reports are frequently cited as illustrations of this, including: Royal Society of London. 1985. *The Public Understanding of Science*. Royal Society: London.
[45] Irwin, Alan and Brian Wynne (eds.). 1996. *Misunderstanding Science?* Cambridge: Cambridge University Press: Cambridge. Shapin, Steven. 1992. 'Why the public ought to understand science-in-the-making' *Public Understanding of Science* 1.
[46] Slovic, P. (1992) 'Perception of risk: reflections on the psychometric paradigm', in Krimsky et al. (eds.) *Social Theories of Risk*, Praeger: Westport. Wynne, B. (1989a) Sheep Farming After Chernobyl: A Case Study in communicating scientific information. *Environment Magazine,* 31 (2), pp10-15, 33-39; Wynne, B. (1989b) Frameworks of rationality in risk management: Towards the testing of a naïve sociology. In Brown, J. (ed.) *Environmental threats* (pp 33-45). London: Belhaven.
[47] *The Economist*, Vol. 342, 1st February 1997, p1847
[48] For more detailed accounts of the techniques used in combinatorial chemistry see: D. Tapolczay et al (1989), Extracting Order from Chaos, *Chemistry and Industry*, n19 p772-4; See also: http://www.synthelabo.fr/strategy/uk/31_nof.htm
[49] For a related genechip articles, see: *Fortune*, March 31, 1997 v135 p56; anon, Gene Chip Technology Ready to Impact Diagnostic Markets, *Genetic Engineering News*, December 1997; anon, BioChips: Advances in DNA Array and Microfluidics Technologies, *Lehman Brothers*, November 21, 1997; For an illustration of GeneChip products, see: http://www.affymetrix.com;
[50] For further references to Labchip technologies, see: Running on Parallel Lines, *New Scientist*, Oct 25, 1997 (http://www.newscientist.com/cgi-bin/pageserver.cgi?/ns/971025/nnanolab.html); Orchid Raises $27M to Fund Varied DNA Chip Programs, *BioWorld Today*, April 15, 1998; Orchid: Microchemical Processors, Emerging Company Profile, *BioCentury*, March 16, 1998; A Hail of Silver Bullets, *Forbes*, January 26, 1998; anon, This Chip Could Save Your Life, *Newark Star-Ledger*, January 26, 1998; anon, Honey, I Shrunk the Lab, *Photonics*, January 1998; Outfitting Today's Drug Development Labs. R&D Future Trends: Miniaturization, *Hambrecht & Quist LLC Industry Report*, August 13, 1997; Discovery on a Credit Card?, *Drug Discovery Today*, July 7, 1997;
[51] Butler, J. M. (1999) STR Analysis by Time-of-Flight Mass Spectrometry. *GenePrint* 3-6; see also: Butler, J. M. et al. (1998) International Journal of legal Medicine, 112 (1), 45; Schumm, J. W. (1997) Profiles in

DNA 1 (2), 11; Monforte, J. A. et al (1997) U.S. Pat no 5,700,642; Monforte, J. A. and Becker, C. H. (1997) Nature Med, 3, 360.

[52] Jeremy Hayter (1997), Developments in Bioinformatics. *The U.K. Biotechnology Handbook 97/98*.

[53] Ibid.

[54] See HUGO Intellectual Property statements 1994, 1996, 1997 (http://www.hugo-international.org)

[55] Hopkins, M (1998), *An Examination of Technology Strategies for the Integration of Bioinformatics in Pharmaceutical R&D Processes*, MSc Dissertation, Science and Technology Policy Research Unit, University of Sussex.

[56] See: http://www.cc.gatech.edu/gvu/softviz/infoviz/information_mural.html

[57] A full account of visualisation technologies in bioinformatics can be found at the EBI web site: http://industry.ebi.ac.uk/~alan/VisSupp/VisAware/index.html

[58] The OMG task force recently developed five proposals for interface specifications after receiving responses to a consultation exercise from eight bioinformatics companies. See *Bioinform News Service*, Nov 23, Vol 2, No 23, 1998.

[59] For information on CORBA, see J. Hayter (1997/98), Developments in Bioinformatics. *The U.K. Biotechnology Handbook 97/98*. pp25-28; relevant sites: http://sunny.ebi.ac.uk/; http://sunny.ebi.ac.uk/BioObjects.html; http://adams.patriot.net/~tvalesky/freecorba.html

[60] Oxford Molecular, Glaxo Wellcome, Astra Bioinformatics Centre (Sweden), Silicon Graphics, Nottingham University, Trinity College Dublin and Turke Centre for Biotechnology. See: http://www.oxmol.co.uk/biolib/

[61] See: http://sdct-sunsrv1.ncsl.nist.gov/~jacki/RDA/index.html

[62] See: http://www.iso.ch/welcome.html

[63] *Science*, 275. 17 January. 1997. p327. *Science*, 269. 8 September, 1995.

[64] The EMBL was formed from an earlier organisation, the EMBO, the European Molecular Biology Organisation. EMBO and the laboratory project became separate legal identities in 1969. Goals and supporters of both organisations, however, have always largely overlapped and for over 20 years both organisations have collaborated closely.

[65] Http://www.ebi.ac.uk/ebi-docs/ebi-pr1.html

[66] The MRC total spend on this project has been some £16m pounds over 15 years. Similar sums have been committed to the project by the National Institutes of Health in the USA.

[67] In May 1998 the Wellcome Trust pledged full financial support to enable the Centre to complete one third of the 3000 million base human genome sequence, as part of an international collaboration. This decision made available an additional £110 million over seven years, bringing the total Trust investment in the Human Genome Project to £205 million.

[68] This practice was first used from the outset of the *Caenorhabditis elegans* genome project by the Washington University ST.Louis and the Sanger Centre and has now been implemented for the release of human genomic sequence data at both centres. This policy is endorsed by numerous of other centres as well (E. Marshall and E. Pennisi, 1996; National Science Council, 1988.

[69] Bentley, (1996), 'Genomic Sequence Information Should be Released Immediately and Freely in the Public Domain, (written on behalf of the Sanger Centre), *Science*, 274, 533-534; Marshall, E., and E. Pennisi (1996), Science 272, 188; National Science Council (1988), *Report of the Committee on Mapping and Sequencing the Human Genome*, National Academy Press, Washington, DC.

[70] Discussed above.

[71] The SIB, an academic, non-profit institution, was established in 1997 with the goals of promoting bioinformatic software and databases, offering various collaborative education and training courses, administering the ExPASy molecular biology WWW server, and offering support services to researchers.

[72] *Science News*, May 23, 1998, vol 152, n21, -334(2).

[73] Smaglik, Paul, 'Private Genome Sequencing Effort May Hasten Separate Public Venture', *The Scientist*, July 6, 1998.

[74] Http://www.incyte.com/

[75] Human Genome Sciences (HGS) initially also funded TIGR (The Institute for Genomic Research), a not-for-profit research institute but parted from TIGR in 1997.

[76] For a more in depth description of funding project see the following DG XII publications: Dujon, B. 1994. *Tools for genome mapping*. Publication N° EUR 15751 EN, ISBN: 92 826 8705 8; Hallen, M. (ed.) 1998. *Human Genome Analysis supported under BIOMED 1*. IOS Press, Amsterdam; Hallen, M. and Klepsch, A.

1995. *Biomedical & Health Research Series: Human Genome Analysis Programme* (Vol.8) IOS Press, Amsterdam; Hoet, J. and Gortebecke, C. (eds.) 1996. *The Human Genome in Europe - Scientific, Ethical and Social Aspects* Académie Royale de Médecine de Belgique; Suhai, Sándor (ed.) 1997. *Theoretical and computational methods in genome research*. Plenum Press, London.

[77] *Chemical & Engineering News* August 24, 1998 v76 n34 p11(2)

[78] *Nature*, 391. 12 February, 1998. p621. Delegates at the annual meeting of the American Society for Human Genetics (October 27-31, 1998) also expressed the problem that newly generated DNA sequence data has outpaced bioinformatic capability to make use of the data. See *Bioinform News Service*, Nov 23, Vol 2, No 23, 1998.

[79] Poste, G. (1998), Molecular Medicine and information-based targeted healthcare, *Nature Biotechnology Supplement*, vol. 16, May, pp19-21.

[80] Hopkins, M (1998) *An Examination of Technology Strategies for the Integration of Bioinformatics in the Pharmaceutical R&D Process*. MSc Dissertation for SPRU, University of Sussex.

[81] Testing, Testing, *The Economist*, Vol 342, 1 Feb, p1847.

[82] In practice consent is not normally obtained for a tissue sample or suchlike used for diagnostic purposes to be stored and possibly incorporated in a subsequent research project with the approval of the scientific ethical committee system. Consent for storage is regarded as being implicit in the actual donation of the sample

[83] Under FP4, bioethics was extended to three programmes in Life Sciences and Technologies: Biotechnology (BIOTECH); Biomedicine and Health (BIOMED); Agriculture and Fisheries (FAIR).

[84] See http://www.uclan.ac.uk/facs/ethics/esintro.htm

[85] Http://www.gene.ucl.ac.uk/hugo.sampling.html. This statement was released February 1998. See also Knoppers B.M., Hirtle M, Lormeau S., Laberge C.M., Laflamme M. (1998) "Control of DNA Samples and Information", *Genomics.*

[86] HUGO Statement on the Patenting of DNA Sequences Authored by: Dr. C. Thomas Caskey, Professor Rebecca S. Eisenberg, Dr. Eric S. Lander, Professor Dr. Joseph Straus, Max Planck. Edited by: Dr. Belinda J.F. Rossiter. See for full text: http://www.gene.ucl.ac.uk/hugo/patent.htm

[87] http://biomaster.uio.no/brochure/html/national.html

[88] ESHG. 1997. *European Journal of Human Genetics* 5(2).