

EC-US WORKSHOP  
RESEARCH INFRASTRUCTURES



# New Research Tools for a Life Sciences Decade

THE REPORT

Ciudad de las Artes y las Ciencias  
Valencia June 21-23, 2001

Sponsored by

National Science  
Foundation



Generalitat  
Valenciana



European  
Commission





# New Research Tools for a Life Sciences Decade

THE REPORT

---

---

<http://www.cac.es/ecusworkshop>

# CONTENTS

	The Life Sciences Decade	4
	1990-2000: Ten years of progress in Life Sciences	5
	1990-2000: The dawn of new technologies	6
	1990-2000: The role of research infrastructures	6
	2000-2010: A new decade of knowledge	7
	The Workshop Mission	8
	The workshop dynamics – reports	8
	Expectations and achievements in 2010	10
	Moving across scales – The circle of Life Sciences	10
	What will it take in Research Infrastructures to get there?	11
	Specific examples of international efforts	11
	What kind of training will be necessary?	12
	How can a EU-US collaboration help to achieve this?	12
	Executive Summary	14
	What would you like to achieve in your community in the next 10 years?	15
	What will it take (in research infrastructure) to get there?	17
	Specific examples of international efforts	18
	What kind of training?	19
	How can a EU-US collaboration help to achieve this?	19
	Primary objectives for Biological Resource Centres (BRCs)	20
	What will it take (in research infrastructure) to get there?	21
	Financial Support	22
	International Initiatives	22
	Training	23
	Synthesis Statement	23
	The Global Ecological Observator Network	24
	What would you like to achieve in your community in 10 years?	25
	What will it take (in research infrastructure) to get there?	25
	Examples of international efforts	26
	Training of new scientists and exchange of ideas	27
EU-US	27	

ANNEX I  
INDIVIDUAL  
CONTRIBUTIONS  
p.29

David B. Allison	30
Philip E. Bourne	32
Graham Cameron	34
Berge Diderichsen	38
David Emerson	40
Johanne Fischer	42
Martin Hrabě de Angelis	44
Tony Mayer	46
César Nombela	48
R. John Parkes	50
Louis Réchaussat	52
Anna-Louise Reysenbach	54
John D. Strandberg and Franziska Grieder	56
Sarah A. Woodin	58

ANNEX II  
PRESENTATIONS  
p.60

P Zanella: Research in the Information Age	60
J Gosz: The Environmental Research Challenge	60
Peter Laggner: Nanobionics and Synchrotron Light	60
Thomas Kaufman: Centralised Resource Centres	60
Nancy Federspiel: Arabidopsis 2010	60
Les Grivell: E-Biosci	60

ANNEX III  
PARTICIPANTS  
p.61

ANNEX IV  
CONTRIBUTORS  
p.62

ANNEX V  
FINAL  
AGENDA  
p.63



# PREAMBLE



Carlos Martinez-Riera  
European Commission



Mary Jane Saunders  
National Science Foundation

## The Life Sciences Decade

We are certainly living in a Life Sciences decade. Completion of the human genome in June 2000 marked the beginning of the post-genomic era, a new era for Life Sciences, sometimes called the 'New Biology'. This 'New Biology' throws light on the way we perform research in the Life Sciences and provides a new angle from which individuals and the society as a whole perceive themselves and the surrounding world.

### The New Biology is characterised by the following attributes:

- **Multidimensionality** – Unlike most of the biology of the last half of the twentieth century, The New Biology increasingly involves collaborative projects with scientists working together – often in virtual centres or coordination networks that extend beyond our borders – to solve central questions about life on Earth. They must be able to integrate across several levels of analysis, ranging from macromolecules to ecosystems. The availability of new tools – such as the complete genetic and functional blueprints of organisms, massively parallel computers, sophisticated nanotools, and satellite imagery – is essential for understanding how components assemble into wholes, be they cells, organisms or ecosystems.
- **Multidisciplinarity** – As with much of modern science and engineering, biology in the 21<sup>st</sup> century has moved into a multidisciplinary phase, with the skills of physicists, mathematicians, computer scientists, geoscientists, chemists, social scientists and engineers required to answer the major ques-

tions. In fact, biology is fast becoming the lingua franca by which a large number of scientists originally trained in other disciplines are able to work together on exciting problems. In addition, biology is offering to other sciences many excellent models, such as DNA computing and molecular motors.

- **Information-Driven** – The fundamental commodity of The New Biology is information. Biology is quickly becoming a data-driven science, in which analysis of previously-gathered information, stored in interoperable databases, can lead to groundbreaking new discoveries and also help to focus the efforts of experimentalists and observational biologists. Vast quantities of information are being generated today, and vastly more information must be generated, stored, mined and analysed in order to turn that information into knowledge.

Many of today's major societal and scientific concerns are intimately related to the Life Sciences, including especially health, the environment and the preservation of biodiversity. Life Sciences are helping to provide solutions to the problems we are facing: preservation and sustainable management of the environment, conservation of biodiversity, prediction of global changes, new therapies and preventive health strategies, and safer and more abundant food.

The Life Sciences are interconnected by their very nature. Indeed, it is the integration of these scientific areas that will result in solutions. The new frontier for the Life Sciences is defined by the effective integration of these various scales of complexity.





The meeting place: City of Science and Arts, Valencia.

## 1990-2000: Ten years of progress in Life Sciences

From 1991 to today, there has been enormous progress within the fields of the Life Sciences.

Spectacular advances in genomics have had great impact on research and society. Some of these accomplishments include the following: the completion of sequencing the human genome and other important organisms; the progress in understanding the molecular mechanisms of cells; the capacity to design new targeted drugs; the understanding of the genetic origin of many diseases; and the enhanced knowledge of the role of genes and their complex expression interactions.

Since 1990, funding per year to genomic databases by the European Commission from its R&D programmes has increased from 10,000 Euro to 10,000,000.

The total number of sequences contained in

GenBank in 1988 was 20,579, almost exactly equal to the number of new sequences deposited per day in 2000. The total number of sequences contained in GenBank as of July 2001 was 12,412,541.

Biological collections, taxonomy and biodiversity, all considered as low-technology areas until recently, are experiencing spectacular progress as genomics opens new avenues for research. We are now able to design very precise mutant model organisms for research in medicine. For example, with the mouse model, thousands of mutations can be rapidly induced, characterised, and selected according to targeted criteria. Modern biodiversity and taxonomy incorporate the knowledge of genomes and molecular diversity and, at the same time, provide the proper conceptual framework to understand the genomics machinery.

In the early 1990's, the full characterisation of the gene responsible for one mutation in *Arabidopsis thaliana* took 4 years. In 2000, the full genome of *Arabidopsis* had been sequenced.

1990 was the beginning of gene targeting (knockout mice) which is widely used in 2000 as a standard technique. We now have the ability to create time and tissue specific knockouts as well as induce thousands of random mutations through the ENU chemical mutagenesis technique. Since 1997 there has been a boom in the development of standardised diagnostic methods (screens) to identify models for inherited diseases in man. Over 100 such tests for different diseases are performed on the chemically mutagenised animals. This creates a very valuable resource for future genomic and proteomic research. In 2000, 600 new mutant lines were created in Europe following this technique in the framework of national genome programmes.

Observations of the Earth and its ecosystems have expanded towards a global endeavour and have benefited from the developments in space teledetection, portable telecommunications stations, non-invasive sensors, and developments of sea observation devices.

## 1990-2000: The dawn of new technologies

Other fields of science and technology have also evolved in an exciting way since the early 1990's and this has further enabled progress in the Life Sciences.

The Information Technology and Telecommunications revolution has completely changed the parameters of any foreseeable evolution of knowledge only 10 years ago. The Internet has provided an unexpectedly rapid and easy way to access multimedia information; bandwidth and the quality of links unceasingly increase; data storage capacity and computing power have progressed at an exponential rate. Genome research, biodiversity studies, and global modelling would have been impossible without these breakthroughs.

While nanotechnologies are still young disciplines in 2001, they are being called upon to play an increasingly important role as a source of new and perhaps unpredictable engineering capacity and as a source of new tools for many other areas of research. The interface between biology and the emerging nanotechnologies is one of the most promising of intersections, resulting in biomolecular chips, cell nano-sensors, targeted probes, high-throughput determination of nano-structures, new nano-engineered organic materials, etc.

## 1990-2000: The role of research infrastructures

All of this progress would have been impossible without the rapid development of appropriate research infrastructures. The very concept of research infrastructures has also dramatically evolved over the past 10 years, unfolding from an idea of big installations to embrace virtual, networked, flexible, and distributed structures. Two forces have powered this evolution: the need to address scientific issues for which global approaches provide the right scale; and the





development of Information and Telecommunication technologies that have helped to overcome physical barriers.

The evolution of research infrastructures in the last decade is crucially linked to the high-throughput production of information and biomaterials and to the ability to collect them in a processable and accessible way. This is not only true for the enormous amount of data generated by genomic initiatives, but also for the increasing quantity of information resulting from ecological observation initiatives as well as for specimen collections and Biological Resource Centres, for all of which higher capacity is constantly being demanded.

In striving to maintain the growth path demanded by new research needs, the following have rapidly developed: large, interoperable, integrated databases containing information on sequences, structures, and molecular interactions; new networks of observatories of all kinds of ecosystems, including terrestrial and

marine; larger and more accessible biological resource centres for microbes, cell cultures, tissues, and model organisms; and new global initiatives to collect and access biodiversity information.

## 2000-2010: A new decade of knowledge

Until now, the leading concept has been to constantly increase capacity in order to cope with the data deluge. Biologists are now spending considerable time in front of their computers processing data; and too often, it seems that there is no time left for laboratory or "wet" science. There is a large consensus and a pressing need to move the emphasis from data to knowledge: how do cells work; how can we categorise the state of the planet; and what richness lies available within the Earth's genetic and organismal diversity. Only then will the accumulated efforts of past decades make sense.

## The Workshop Mission

With the above perspective, in October 2000, the EC-US Task Force on Biotechnology Research launched the idea of this workshop on research infrastructures for Life Sciences.

The workshop was given the task to determine the needs of research infrastructures within the Life Sciences for the next decade.

Such a broad and challenging demand requires framing the question within the limits of a conceptual challenge: to formulate the needs according to the expected achievements of Life Sciences ten years from now. Three focus areas that cover a large spectrum of research interests and that deal with issues relevant to society on a global scale were addressed:

- Biotechnology and Emerging Technologies
- Biological Resource Centres
- Ecological Observations

Workshop participants were then asked to imagine a global fully integrated functionality that would be operational by 2010 and to envisage and formulate some requirements, in terms of infrastructures, that would achieve this goal.

### Can we predict technological development?

#### March 1949: Popular Mechanics

"...experts predicted computers of the future would add as many as 5000 numbers per second, weigh only 3000 pounds, and consume only 10 kilowatts of power"

#### 1999: Interagency Working Group on Nanoscience, Engineering and Technology.

"...computers the size of a cube of sugar"

The idea is not to predict the future, but to set up reasonably achievable challenges. This approach exercises people's minds by requiring them to set a captivating goal. No one could foresee the details of the mission that would bring a man to the moon in the early 1960's, but the challenge made it possible; the completion of the human genome was technically unfeasible with the technology of the early 1990's, but it was eventually accomplished.

Furthermore, and probably more importantly, the spin-offs, in terms of scientific, technological, and economic developments that those challenges enabled, went beyond the direct benefits of the achieved goal. New technologies, products, and scientific advances were made possible by the endeavour.

## The workshop dynamics – reports

The group of 28 experts and programme managers brought expertise on functional genomics, animal and plant models, bioinformatics, high-performance computing, ecology, biodiversity, nanotechnology, and high-energy physics. The group included both early-career and more senior experts.

Two co-chair persons were nominated to moderate the overall discussions. James Gosz and Paolo Zaniella accepted this role.

All participants were requested to prepare inputs before the meeting. Their individual contributions are included in Annex I to this report.





Six presentations on specific issues set the scene for the subsequent discussion (they are included in Annex II to the report).

Three breakout groups, one for each of the focus areas, were formed. Each was asked to structure its report around the following questions:

- Question 1.  
What should be achieved in the field in the next 10 years?
- Question 2.  
What will it take (in research infrastructure) to get there?
- Question 3.  
Are there specific examples of existing international efforts?
- Question 4.  
What kind of training is needed?
- Question 5.  
How can a EU-US collaboration help?

For each group, a moderator and a rapporteur were identified. Their names and those of each group's members are included in the reports.

The combination of the varied knowledge and experience of the participants, along with the input from the individual contributors, led to exciting and cross-disciplinary discussion, both at the plenary sessions and within the breakout groups. The results are described in the overall conclusions and in the sub-group reports. □

# EXECUTIVE REPORT



James Gosz  
University of New Mexico



Paolo Zanella  
University of Geneva

## Expectations and achievements in 2010

Today's major societal and scientific concerns are related to our health, the preservation and sustainable management of the environment, the conservation of biodiversity, and the consequences of planetary changes.

The challenge for science is to provide responses to the expectations of society, an overall concept that is supported by three pillars:

- a thorough understanding of life, whether it relates to the metabolism of individual cells, the development of diseases in the human body, or the ability to assimilate nutrients, water, and energy from the environment;
- information on biological resources and comprehensive knowledge of all the gene functions in important model organisms;
- capacity to assess the planet's status, to predict how key ecosystems are responding to changes, and to determine the biogeochemical and climatic consequences.

Therefore, a visionary global effort in the Life Sciences should aim at three major goals:

- build a model of life that is suitable for assimilation, evaluation, and use of the flood of biodata so as to understand fundamental biological processes;
- compile a comprehensive gene annotation of the genomes of major model organisms that would be underpinned by a coordinated global network of biological resource collections;
- apply new knowledge and technologies, including super-intelligent software and grids for bioinfor-

matics, aiming to allow the elaboration and testing of small and large scale predictive models of biodiversity, biocomplexity, community structure, and biogeochemical cycles.

Those grand objectives may be exemplified by the formulation of three major initiatives addressing those goals:

- *The Model of Life* (on page 14)
- *The World Genome Annotation Project* (on page 20)
- *The Global Ecological Observatory Network* (on page 24)

## Moving across scales – The circle of Life Sciences

There is an evident added value in moving across scales of complexity and integrating the knowledge already achieved in the different levels. Life Sciences must be capable of building new understanding on higher scales from smaller ones and vice-versa.

Modelling life, compiling gene functions, or assessing the planet's status are all challenges that require the integration of knowledge from the nano-scale to the ecosystem. Integrating scales of complexity and interfacing among the Life Sciences will be the research paradigm for the next decade.

---

**Individual Life Sciences are connected and interface with each other as well as with other disciplines including the social sciences. A circle can be generated that runs from biodiversity to biology to health to nanobiotech-**

nology to medico-technique to communication technologies to information technology and back to biodiversity. At each node of expertise there is a need to ensure compatibility and exchange of data as well as the need to facilitate connectivity with other disciplines. A clear understanding of the model of life, as well as society's needs, is critical for this circle to be established.

Connecting complementary competencies and needs within the circle of Life Sciences is the great challenge for a joint global effort in the next decade.

## What will it take in Research Infrastructures to get there?

Research infrastructures will continue building their capacity to cope with the collection of biological materials and increasingly complex multimedia data resulting from optimised high-throughput technologies.

Beyond the need to maintain their assimilation capacity, the integration and interconnection of autonomous nodes into global networks, provided with a common intelligence and/or organisation structures, will constitute the target of the efforts in supporting infrastructures.

Upon moving towards knowledge-oriented infrastructures, tasks linked to modelling and simulation will require greater efforts with respect to databasing and dataset intercomparability.

To this end, disciplines and technologies new to biology will be incorporated in the Life Sciences' corpus of knowledge and tools. Information and telecommunication technologies, space observation, electronic components and instrumentation, nanotechnology, and mathematics will increasingly develop applications to biological problems.

Global approaches underpinned by robust local and regional expertise and resourcing will constitute the most adequate structure of those infrastructures. Unique central structures may be desirable, but are difficult to achieve in practical terms even within the scope of the next ten years.

Whenever common funding of unique structures is not possible, coordinated mechanisms of national or regional funding will be defined that assure the survival of the global initiative and that reduce the impact of a lack of local funding. To this end, some amount of redundancy may be necessary.

## Specific examples of international efforts

The Global Biodiversity Information Facility (GBIF)<sup>(1)</sup> is an international initiative to share the world's biodiversity scientific information in a distributed manner through an interoperable network of biodiversity databases and analysis tools.

GBIF constitutes a successful example of an emerging global initiative that is centrally coordinated by a

(1) <http://www.gbif.org>

secretariat with its own funding but that is built upon the contributions of independent nodes sharing and managing their own data and resources.

GBIF had its genesis from the work of the OECD Megascience Forum Working Group on Biological Informatics that was established in January 1996. The GBIF concept was endorsed by the OECD ministers as an international mechanism to make biodiversity data and information accessible worldwide.

More countries, including non-OECD states, were since then involved in the definition of the organisational structure and scientific mission of GBIF that culminated with the formal constitution of the Governing Board in March 2001 and the decision to establish the GBIF secretariat in Copenhagen in June 2001.

GBIF is a paradigmatic example of how a leading concept (compiling biodiversity information) can mobilise wills and resources and thus overcome organisational, technical, and political problems.

## What kind of training will be necessary ?

Training will be both the main engine and the principal spin-off of these initiatives. With such integrated and interdisciplinary infrastructures, training of biologists in new technologies and disciplines, as well as of technicians and scientists from other areas, will constitute a central asset of the initiatives.

Therefore, the new paradigm in training will focus on fostering transdisciplinary curricula particularly

between disciplines dealing with different scales of complexity that will help to close the 'circle of Life Sciences' as described on page 10.

International training schemes are easy to implement and expand, as several programmes already exist allowing transcontinental access. At the same time, they reinforce the networking nature of the new infrastructures. Geared by common work programmes, those exchanges will create a critical mass of combined expertise and links between centres, thus reinforcing the backbone of the virtual infrastructures.

The infrastructures themselves will provide the support for new forms of exchange and training through virtual conferences and courses.

## How can a EU-US collaboration help to achieve this?

Together, the European Union and the United States have the necessary expertise and resources to accomplish those global challenges.

The proposed objectives are global by nature and do not deal with areas where commercial interests could hinder a common approach, at least in the perspective of the next five years.

An initial consequence of the above is that there are, in principle, no major impediments to a coordination of efforts in which synergies will not only be mutually beneficial, but will also produce global benefits for science and society.



The second consequence is that those initiatives, developed by the momentum of the EU and US, should also incorporate the expertise and resources of other countries.

Existing funding structures in the US and the EU only partially support the launching of common large initiatives like the ones proposed in this report. There is some experience, though, in fostering and favouring the coordination of activities aimed at international objectives through national funding schemes.

Some experience is being gained in synchronising national funding of complementary projects, namely in transatlantic telecommunications networking, particularly between the European Commission and the National Science Foundation.

If common programmes are not envisaged in the short or mid term, the coordination of international initiatives will become a necessity as societal and planetary issues overflow the limits of the national funding programmes.

The EU and the EC, mainly through the EC-US Task Force on Biotechnology Research, may identify and foster the coordination activities that will produce a leveraging effect on mobilising the important resources of the national funding schemes. In this context, the ratio between each dollar invested in coordination initiatives and structures and the research efforts they enable may be as large as 1 to 1000, as demonstrated by the recent experience of GBIF.

For the three proposed initiatives on page 10, relevant national and international programmes already exist, or will exist in the next couple of years. Subgroup reports indicate that, at present, these initiatives are developing slowly and in an uncoordinated manner. A synchronisation and coordinated thrust will save millions of dollars, and, at the same time, will help to achieve the declared objectives better and sooner. □