Data, Models and Theories for Complex Systems: new challenges and opportunities

Patrick Valduriez, Nicolas Arnaud, François Briand, Olivier Gascuel, Olivier Gimenez, Christophe Godin, Hervé Jourde, Pascal Kosuth, Pascal Neveu, Esther Pacitti, et al.

ModSysC2020

• Données, modèles et théories pour les systèmes complexes : défis et opportunités

• Premier rapport de prospective

Patrick Valduriez (INRIA-LIRMM)
Nicolas Arnaud (CNRS-OSU OREME), François Briant (IBM)
Olivier Gascuel (CNRS-LIRMM), Olivier Gimenez (CNRS-CEFE)
Christophe Godin (INRIA-DAP), Hervé Jourde (UM2-HSM)
Pascal Kosuth (IRSTEA-Maison de la Télédetection)
Pascal Neveu (INRA-MISTEA)
Esther Pacitti (UM2-LIRMM)
Andréa Parmeggiani (UM2-DIMNP)

UNIVERSITÉ MONTPELLIER 2
SCIENCES ET TECHNIQUES

Data, Models and Theories for Complex Systems: new challenges and opportunities

Final Report of the ModSysC2020 Working Group

1 July 2011

Group members: Patrick Valduriez (INRIA-LIRMM), Nicolas Arnaud (CNRS-OSU OREME), François Briant (IBM), Olivier Gascuel (CNRS-LIRMM), Olivier Gimenez (CNRS-CEFE), Christophe Godin (INRIA-DAP), Hervé Jourde (UM2-HSM), Pascal Kosuth (IRSTEA-Maison de la Télédétection), Pascal Neveu (INRA-MISTEA), Esther Pacitti (UM2-LIRMM), Andréa Parmeggiani (UM2-DIMNP).

Key-words: agronomy; environmental science; life science; climate; complex systems modeling and simulation; observation of natural and biological medium; scientific data management; data production, storage and archiving; high performance computing (HPC) and cloud computing.

Summary

At University Montpellier 2, the modeling and simulation of complex systems has been identified as a major scientific challenge and one of the priority axes in interdisciplinary research, with major potential impact on training, economy and society. Many research groups and laboratories in Montpellier are already working in that direction, but typically in isolation within their own scientific discipline. Several local actions have been initiated in order to structure the scientific community with interdisciplinary projects, but with little coordination among the actions.

The goal of the ModSysC2020 (modeling and simulation of complex systems in 2020) working group was to analyze the local situation (forces and weaknesses, current projects), identify the critical research directions and propose concrete actions in terms of research projects, equipment facilities, human resources and training to be encouraged. To guide this perspective, we decomposed the scientific challenge into four main themes, for which there is strong background in Montpellier: (1) modeling and simulation of complex systems; (2) algorithms and computing; (3) scientific data management; (4) production, storage and archiving of data from the observation of the natural and biological media.

In this report, for each theme, we introduce the context and motivations, analyze the situation in Montpellier, identify research directions and propose specific actions in terms of interdisciplinary research projects and training. We also provide an analysis of the socio-economical aspects of modeling and simulation through use cases in various domains such as life science and healthcare, environmental science and energy. Finally, we discuss the importance of revisiting students training in fundamental domains such as modeling, computer programming and database which are typically taught too late, in specialized masters.

Our recommendations in terms of interdisciplinary research actions are summarized as follows. In order to develop interdisciplinary synergies in modeling and simulation related to life and environment in Montpellier, we should capitalize on the experience already gained by existing and dynamically operating communities and collaborations. The research directions we describe are part of the LabEx NUMEV and EpiGenMed. A more focused subset of this program is also part of the IBC proposal or of the EquipEx GEOSUD. In these four projects, data management is viewed as important in conjunction with modeling, simulation and computation. The GRISBI group (Groupement de Recherche Interdisciplinaire sur les Systèmes Biologiques) gathers more than 70 researchers working on theoretical and experimental modeling approaches to biological systems and related environmental problems. GRISBI should be supported and developed because of its functionality towards the large variety of institutions distributed in Montpellier area. The IBC interdisciplinary project center aims at gathering about 50 researchers working on modeling, processing and analyzing data on a large scale in the fields of biology, health, agronomy and environment. IBC's flexible structure should serve as a test case for further similar interdisciplinary centers, for example focused on the interplay between ecology, mathematics and computer science, or on the impact of climate changes on human populations. To further foster interdisciplinary collaboration, we identify the need for an Environmental and Biological Data Center (EBDC), opened to access for a worldwide community, for common facilities for the storage of massive databases. This is a practical, operational project that should encourage the entire Montpellier community in modeling and computation to collaborate through the sharing of the most stable objects (data) they manipulate. To strengthen the interdisciplinary use of geoinformation for the analysis of environmental dynamics, the GEOSUD (GEOinformation for Sustainable Development) initiative that involves various teams in Montpellier should also be supported. These actions (NUMEV, EpiGenMed, GRISBI, IBC, GEOSUD and EBDC) should be strongly encouraged by UM2 and its partners, in particular, by dedicating significant amounts of human resources (PhD students, engineers and researchers).

We also acknowledge that training of scientists across complex disciplines is a major challenge. We discussed the importance of revisiting students training in fundamental domains such as modeling, computer programming and database, which are typically taught too late, in specialized masters. We proposed the following specific

actions for training interdisciplinary students: (1) moving towards teaching in English at UM2 in order to compete with top universities in attracting the very best foreign students; (2) interdisciplinary training, with the fundamentals of physico-mathematical and numerical modeling as well as Information and Communication sciences (STIC), in particular, programming and databases, and social sciences; (3) using platforms and databases in teaching; (4) training initiatives by levels of teaching (Licence, Master, PhD); (5) educating students to scientific independence; (6) making life on campus more … lively: UM2 students should be given the opportunities to become part of the university life and true actors of its functioning, not just consumers.

We expect these actions to have an important social and economic impact at the regional, national and international levels. At the regional level, it will foster collaboration between mathematicians and computer scientists involved in HPC, modeling and simulation, data management and data analysis with scientists of different disciplines such as agronomy and environmental science, life science, etc., which are strong in Montpellier. Although collaboration between these communities already exists in different forms, the interdisciplinary actions we propose will strengthen collaboration by giving it more impetus and visibility, which will in turn generate interest from other participants, including many local start-ups which have to deal with scientific data, and public authorities that have to define public policies in the domains of environment and health. At the national and international levels, the impact will be a natural continuation of the local actions and a consequence of the boarder-free, global nature of complex systems. Since all scientists participating in the project have ongoing collaboration with peer teams at the national and international level, these projects will simply strengthen those collaborations.

Table of content

1. Introduction

1.1. Motivations

The past decade has seen the emergence of new scientific communities whose needs in terms of modeling and simulation, data storage and data sharing are rapidly expanding. These communities typically try to understand complex systems (including living systems) and their interaction with physical and chemical environments. In Montpellier, these communities are those dealing with environmental science with topics like biodiversity, agronomic resources, water resources and water hazard, land dynamics and biological ones like genetics, cellular biology, developmental biology and neurophysiology in relation with computer scientists, theoretical physicists and mathematicians.

The study of Biodiversity must take into account the interlocking scales of basic elements of life, from genes to the individuals or populations, and, respectively, from molecules through the cells to tissues, and the relationship of each of these scales and their related phenomena with boundary conditions and constraints of physical and chemical nature. In biology, the choice of model organisms at world level and impressive progress of biotechnologies and genetics make it possible to decipher step by step how organisms develop and function under the control of genes and environmental conditions. At the level of the individual, and more so at that of the population, the interaction between living organisms is essential and shares similarities with the modeling of social networks. The study of water, as a resource (quality and quantity) or as a hazard, is also closely related to complex systems, where physicochemical couplings of nested scales are many and diverse. Furthermore natural systems are increasingly impacted by human activity: as a result biophysical processes are not the only drivers of their dynamics and socio-economical processes and management strategies must be taken into account in modeling approaches to provide integrated representations of systems dynamics.

These systems are too complex to be understood by simple observation or experimentation and the use of large-scale simulation in addition to modeling, is critical to understand their sensitivity and reaction to external drivers (for instance climate change impact or pollution impact on ecosystems) and to simulate potential scenarios. Furthermore, large-scale simulation tools are creating a huge data overload. In climate modeling for instance, simulated data sets are growing faster than the data set size for any other scientific discipline, with collections of hundreds of exabytes (1018 bytes) expected by 2020. Scientific data are also very complex because of the heterogeneous methods, with inherently multi-physics, multi-scale nature (spatial scale, temporal scale) of many sciences, resulting in data with hundreds of attributes or dimensions. Processing and analyzing such massive sets of complex scientific data is therefore very difficult as new, scalable data management techniques for HPC or cloud environments are needed.

At the same time, spectacular advances driven by technological progress in nanosciences and optics, have enlarged our knowledge of biological systems, ranging from single bio-molecules via cells to tissues and organisms, via massive production of quantitative data to assess the cell or organism life functions in time and space. The same occurs in the domain of environmental sciences that benefit from major breakthroughs and advances in satellite Earth Observation and distributed sensors technologies. Advances in modeling new phenomena in living systems are then rapidly becoming a prerogative for developing quantitative models to help data interpretation and reconstruction, and support biological, medical, and environmental sciences. Last but not least, modeling is necessary to reduce time and cost in fundamental and applied research, with strong implications for biotechnologies, agriculture, food science, etc. and the development of personalized and predictive medicine.

1.2. ModSysC2020: mission and organization

At University Montpellier 2 (UM2), the modeling and simulation of complex systems, including computation, algorithms and data management, has been identified as a major scientific challenge and a priority in interdisciplinary research, with high potential impact on research, training, economy and society. Many research groups and laboratories in Montpellier have been already working in that direction, but in most cases in isolation within their own scientific discipline (e.g. computer science, HPC, biology, soil science, hydroscience, geoscience, etc.). Furthermore, several local actions have been initiated in order to structure the scientific community with interdisciplinary projects in a bottom up manner, but with limited coordination among the actions.

Based on this vision and on the scientific community strengths, UM2 and its partners have set up the ambition to position Montpellier within the next decade as a world-wide leader in modeling and simulation of complex systems in domains where Montpellier has an excellence track record (i.e. computer science, agronomy, environmental science, and life science). In July 2010, Patrick Valduriez was appointed by UM2 to establish and conduct an experts' group, ModSysC2020 (Modeling and Simulation of Complex Systems in 2020), to work toward this ambition. The goal of ModSysC2020 was to analyze the local situation (forces and weaknesses, current projects), identify the critical research directions and propose concrete actions in terms of research projects, equipment facilities, and human resources. In order to establish a strong interdisciplinary scientific community, we also realized that researchers, engineers and students will need to be trained with new skills (e.g. computer science), across scientific disciplines. Therefore, we also propose specific actions in terms of training.

To work efficiently, we set ModSysC2020 as a small group, with 11 members chosen only for their scientific expertise, experience and commitment, not because of their institutions. The group gathered experts of the following disciplines (see Annex 1 for short bios): geoscience (N. Arnaud); information systems (F. Briant); computer science, bioinformatics (O. Gascuel); environnemental science (O. Gimenez, P. Kosuth); agronomy (C. Godin, P. Neveu), hydroscience (H. Jourde); computer science, data management (E. Pacitti, P. Valduriez); physical modeling of health science and biology (A. Parmeggiani). To get additional insights on two related topics (HPC and international training), at one of its meetings, the group organized presentations and discussions with representatives of the HPC@LR project (http://www.hpc-lr.univ-montp2.fr) and of the Webscience international master proposal from members of the Montpellier Webscience meetup group (http://www.meetup.com/webscience-montpellier).

1.3. ModSysC2020: results

During the fall of 2010, most members of the working group have been involved in federated projects such as the Equipements d'Excellence (EquipEx) and the Laboratoires d'Excellence (LabEx), which provided a good basis for inspiration. For instance, the modeling aspects are part of several LabEx awarded in 2011: Agro, Eau, NUMEV and EpiGenMed. In addition, the NUMEV LabEx involves modeling, simulation, HPC and scientific data management and the GEOSUD EquipEx is focused on the acquisition, processing and management of spatial information and Earth Observation for environmental systems. Since 2006, local modeling communities have been involved in the wide interdisciplinary group GRISBI (Groupement de Recherche Interdisciplinaire sur les Systèmes Biologiques - Interdisciplinary Research Group on Biological Systems) supported by the biology and physics departments of UM2 and sharing knowledge with all related disciplines in Montpellier (chemistry, computer science, mathematics, mechanics, medical and environmental science, opto-electronics, …), in particular, in the context of the "Virtual plants and ecosystems" program of the Agropolis Foundation ("Programme Plantes et écosystèmes numériques"). More recently a project of an Institute of Computational Biology (IBC) that focuses on the modeling and large-scale sharing of biological and genetic data has been proposed as a first attempt to gather the major forces of the region and promote international visibility. However, these projects are relatively focused and bottom-up. Thus, our working group proposes a more global perspective that links the different LabEx, EquipEx and projects to get overall consistency and foster synergy. To guide this perspective, we decomposed the scientific challenge into four main themes, for which there is strong background in Montpellier:

1. Modeling and simulation of complex systems, including mathematical, physical and numerical modeling;
2. Algorithms and computing, simulation on HPC platforms; production of very large, complex datasets.
3. Scientific data management, including organization, storage, archiving, sharing and analysis of very large, complex datasets produced by scientific applications.
4. Production, storage and archiving of data from the systematic observation of the natural and biological media, with the continuum model-system-data-expertise.

1.4. Organization of the Report

In the following sections (Sections 2-5), for each theme, we introduce the context and motivations, analyze the situation in Montpellier and its region, identify research directions and propose specific actions in terms of interdisciplinary research projects and training.

We also acknowledged the need to analyze the impact of modeling and simulation in the social and economic context, with expectations from potential users as well as potential business impact evaluation. In particular, if we propose new interdisciplinary training courses, we must also identify potential employers for the trained students. Section 6 provides an analysis of the socio-economical aspects of modeling and simulation through use cases in various domains such as life science and healthcare, environmental science and energy.

Finally, the group insisted on the importance of revisiting students training in fundamental domains such as modeling, computer programming and database which are typically taught too late, in specialized masters. Section 7 addresses this topic. Section 8 summarizes our recommendations

2. Modeling and Simulation of Complex Systems

Key-words: mathematical-physical-numerical modeling, dynamics, systems biology, developmental biology, quantitative biology, numerical simulation, model systems, coupled and multi-scale problems, non linear and non equilibrium processes, transport, stochastic processes, networks, energy-matter-information coupling, in-silico experiments, environment, life, simulators, virtual cell, virtual organ, virtual plant, simulation platforms.

2.1. Context and Motivations

The current explosion in information and observation technologies should allow us to better comprehend earth and our lives upon it. With fast growing amounts of information, we are pressed to develop theoretical and methodological frameworks for interpreting such information and using it to solve applied problems.

To master the modeling and simulation of complex systems, the steps involving (i) the observation and acquisition of environmental and biological data, (ii) their organization and distribution, and (iii) the development of advanced experimental techniques, are initial fundamental stages of research activities. The process for building knowledge of complex systems requires further development via mathematical, physical and numerical models, mutually unveiling and understanding the basis of new complex architectures, processes and functions of living and environmental systems. Within this approach, complex systems sciences provide testable and predictive understanding, shedding light on social and environmental impacts. These are key issues from the perspective of the use and management of natural resources to face climate change and promote sustainable development of human societies.

The technological breakthroughs in experimental biology and environmental sciences have revealed systems with novel properties in the biotic world and environment. Spectroscopy, ultrahigh resolution microscopy and non linear optics for biological imaging, techniques for manipulating nanoscopic and individual bio-molecular systems, both in-vitro and in-vivo, open unprecedented access to the quantitative description of biological constituents and related complex processes in space and time which characterize specific biological functions at different scales.

Equally important, progresses in molecular biology, genomics and functional genomics open the way to the understanding of complex regulatory mechanisms that control essential biological processes, which are prerequisites, for instance, for the development of novel therapies based on new generations of medicines specifically delivered into sick cells.

In parallel, the integrated systemic view related to “omics” approaches (genomics, transcriptomics, proteomics, metabolomics, interactomics, etc.) in the domain of systems biology, describes and rationalizes how diverse components, in all ranges of concentrations, can organize and regulate reciprocally in complex networks of interactions to assess cells or organisms life and functions. The systems biology frame applied to life and environmental sciences opens perspectives like the development of personalized and predictive medicine, reasonable agronomy and environment-friendly farming, or bio-technological food science.

In environmental sciences, rapid developments in satellite Earth Observation and distributed sensors technologies provide in depth information on spatial structures and temporal dynamics at multiple resolutions, allowing to study processes, trends and changes from local scale to global scale, and to couple systems such as atmosphere, ocean, solid Earth, hydrosphere, biosphere, anthroposphere, that used to be analyzed separately.

Due to the new challenges imposed by living and environmental systems complexity and biodiversity, quantitative approaches need to be developed that integrate mathematical, physical and computer sciences (Biological Physics, Theoretical Biophysics, Quantitative Biology, Computational Biology, Integrative Biology, Quantitative Ecology, Water cycle and Carbon cycle study) in order to describe and understand intracellular supramolecular constituents and assemblies, cells and unicellular organisms, tissues and organs, individuals and populations, ecosystems and anthroposystems (in a multi-scale approach), and correlate physico-chemical phenomena with information processing and regulation phenomena.

Along the years “hard sciences” have developed within their respective fields modeling tools allowing fine understanding of the physics involved as well as reliable industrial design. Large national or international programs have enabled these developments, such as the ASC program of the US Department of Energy (http://www.sandia.gov/nnsa/asc). In France, the National network of complex systems (http://rnsc.fr) encourages multidisciplinary approaches. Similar general programs aiming at knowledge transfer for the benefit of biological sciences already exist. In the US, major programs of the Common Fund for the Interdisciplinary Research Program of the National Institutes of Health (http://nihroadmap.nih.gov) have contributed to apply methods developed by physicists, mathematicians and computer scientists to biomedical sciences. In countries like Germany and Japan, it is not rare that foundations, with capitals from public and/or private organizations and companies, finance fundamental research in complex systems (see, for instance, the Wolkswagen’s Foundation role in biological sciences).

Modeling and simulation of complex systems is the result of longstanding interaction between mathematics, physics, mechanics and computer sciences. Theoretical physics approaches are based on advanced methods issued from statistical physics, physics of complex systems and stochastic processes. The dynamics of interactions between large numbers of particles, enables the modeling of complexity as emerging from simple processes. Stochastic and statistical approaches made possible to overcome the lack of information regarding a system under study such as: 1) the geometry(on macroscopic level, e.g. groundwater modeling in complex aquifer structures such as fractured or karstic rocks, or microscopically, e.g. nuclear chromatine structure modeling), 2) the initial state of systems, and 3) the multi-scale complexity of dynamic environments, such as an ecosystem or the cytoplasm. Statistical methods allow describing environmental collective processes (e.g. avalanches, erosion, pressure-impact relations), plant and organism growth, intracellular processes (e.g. traffic, assembly and cellular functions, genome regulation). In mathematics and computer science, inspired by the analysis of growing organisms and organs, a new form of dynamical systems is being investigated whose state has a structure that changes throughout time (dynamical systems with dynamical structure).

Theoretical and numerical tools may help sparing expensive and long field experiments. Furthermore, they are needed for the interpretation of experiments of increasing sophistication. Programs to integrate such upstream approaches have been launched, combining physical knowledge with mathematics and computing techniques, for environment and living objects. Nevertheless, few research groups possess the integrated vision of all disciplines required for modeling living systems or environmental processes. The existence of structures to facilitate such interactions would increase the potential for substantial progress.

2.2. Modeling and Simulation in Montpellier

Montpellier has an exceptional concentration of basic life and environmental science research and extended expertise in "hard sciences" like mathematics, physics, computer science and engineering. Many researchers from different institutions (CNRS, INRIA, INRA, CIRAD, IRD, IRSTEA, Montpellier SupAgro, AgroParisTech) and laboratories in computer science, robotics and bioinformatics (LIRMM), physics (L2C), mathematics (I3M), mechanics (LMGC), and chemistry (ICG) are already involved in inter-disciplinary collaborations supported by UM2 and national organizations like CNRS (e.g. Interdisciplinary program "Interface Physics/Chemistry/Biology: Support of risk taking"), ANR, INRIA, INRA, Agropolis foundation (e.g. virtual plants and ecosystems program).

In several laboratories, research teams combine computational and modeling approaches with biological experimentation and environmental observatories, to foster interdisciplinary collaboration (for instance, the MODEMIC INRA-INRIA team, the Virtual Plant CIRAD-INRIA team at INRIA-CIRAD, the AMAP CIRAD-INRA-IRD-UM2 unit, the "Biological Physics and Systems Biology" team in the DIMNP biology CNRS/UM2/UM1 laboratory, the TETIS and ESPACE-DEV laboratories in "Maison de la Télédétection" AgroParisTech, IRSTEA, Cirad, IRD working on spatial information for environment). The topics addressed are related to environmental and life sciences.

2.3. Research Directions

Biological and environmental sciences are rapidly becoming quantitative disciplines as a result of advances in computing power, data acquisition technologies, theoretical and numerical modeling. Central questions in natural resource management, environmental monitoring and restoration, ecological forecasting, "omics" sciences (genomics, proteomics, lipidomics, …), cellular and molecular biology, parasitology and virology, environmental and biological regulation of genetically modified organisms, animal and population behaviors, artificial life and neuroscience (to cite just few of them!), require more involvement from mathematical, physical and computer sciences in order to develop new methods and tools.
Interestingly, biological and environmental sciences share some common themes:

- Interactions of phenomena at different characteristic spatial, temporal and energy scales (multi-scaling modeling);
- Non linear and collective phenomena that couple scales (i.e. where multi-scale separation can be impossible to apply);
- Stochastic, or random, events that affect biological and environmental systems (forces, concentration fluctuations, individual behaviors and environmental fluctuations);
- Global constraints (mechanical, thermodynamical, food, resources, …) and internal constraints to the system (logistic, information treatment and sharing, …);
- Complex, interacting processes that produce data that are difficult to manipulate, analyze and visualize;
- Natural variability and observation imprecision (also incompleteness) that introduce unavoidable uncertainty in our ability to characterize biological and environmental processes, that uncertainty having to be qualified and quantified.

Addressing these common themes requires all branches of mathematics, physics, mechanics, and computer sciences. Ranging from intensive computer simulations, that attempt to mimic complex systems, to elegant theoretical analysis and "theorems" via model systems, the aim is to find and study examples that provide conceptual understanding, classification of processes and systems in classes, and generality to complex systems problems.

The Montpellier community in modeling and simulation has been actively working within such framework along the following directions. Some of these are already belonging to LabEx initiatives such as NUMEV and EpiGenMed, but others could be naturally integrated.

Principles of complexity: organization and dynamics in life and environmental sciences

The search for new physical and mathematical principles in biological and environmental systems is one of the main tasks of modern and future hard sciences. It addresses the principles of organization and dynamics of the structures involved in the gene expression process, like chromosomes, and the structures whose development is governed by this expression; the physics of self-assembly and interactions of intra- and inter-cellular structures, like microtubules and nanotubular lipid membranes, and their role in the neuro-degenerative diseases and cancer; the interaction of pathogens with living cells, like interaction of Malaria parasite with a red blood cell, viruses prevalent in warm regions and of pandemic risk like Dengue, West Nile, Japanese Encephalitis, Yellow Fever viruses.

Tissue organized collective cellular apoptosis and development induced by mechanical and topological constraints are developed between the L2C and ISEM laboratories, whereas theoretical modeling in complex and non equilibrium fluids are exploring biological microfluidics (L2C), with "Lab-on-chip" applications. Cell motility in confined geometries and mechanisms of lipid sorting induced by mechanical constraints are interests of DIMNP and L2C laboratories. In plant sciences, similar self-organizing processes are being deciphered: the Virtual Plants team for instance studies how mechanical forces within tissues interact with regulatory networks at the cell level to locally guide organ growth. A major part of the complexity of these developmental systems lies in the feedback loop between these basic processes and the interplay between local and global actions, thus calling for the development of new modeling approaches.

Topics like transport on networks of complex topology, primarily inspired by intracellular and inter-cellular traffic of cargoes and organelles in and between cells or protein activities in transcription and translation mechanisms, are relevant not only for biological sciences and medicine (understanding delivery pathways for new kinds of drug treatment to specific cells and tissues), but naturally apply to environmental sciences. For example, assessing the modalities of such transport processes within complex hydrosystems is a major challenge for the protection and management of groundwater resources or river ecosystems. On the same basis, modeling the management of logistic and territorial communities is necessary to better estimate the influence on world climate parameters (CO2 emission for instance). Risk evaluation and management in emergency events (e.g. escape exits in large common building or transport facilities), the study of the dynamics and optimization of internet data transmission coupled with minimization of CO2 production by Web activities, are other few examples of applications of traffic sciences applied to environment.

In all these cases, problems can bridge the study of the structure and the dynamics of a single molecule to macromolecular assemblies from the cell to the tissue to the organisms (see also multi-scale approaches in time, space and energies) or, equivalently, the collective behavior of populations in an environment and the coupling between environments. In this perspective, there is a common need to develop methods to integrate efficiently information at different scales, procedures of scalability and homogenization, and define emergent behaviors.

Not all classes of problems can be approached in this perspective. Some of them display intrinsic couplings, still difficult to decode and formalize. This is the case of coupled problems. For example, limited resources in energy or food imply intrinsic coupling of processes at different scales together with the coupling with environmental fluctuations. Non equilibrium living matter, but also traffic complex behavior of vehicles and granular matter are well-known examples that display long-range and collective phenomena induced by critical and low energy or behavioral fluctuations during the system dynamics.

Another significant example concerns biological systems. It is commonly believed that information processing and hereditary transmission are the fundamental processes that rule living systems and their evolution. However, recent works, awarded in major scientific journals of many disciplines, have shown how environmental and global constraints (like mechanical constraint or adhesive properties of the extracellular environment of an adhering cell) are decisive for the determinism of cells, organs and organisms development.
The emerging picture is that the variety and complexity of biological phenomena as well as life and organized systems on molecular scales are probably the result of a complex coupling between thermodynamic and information constraints. In order to explain these emergent concepts at cellular and multicellular levels, modern scientists must focus on the study of the complex exchanges between matter, energy and information, up to now a completely unexplored issue (with the exception, maybe, of scientists involved in artificial life experiments on single cell organisms or in behavioral studies of individuals and social communities). This interesting perspective, which does not appear in any LabEx, applies also to the complexity of environmental sciences problems and can open to unexpected technological applications, particularly in the perspective of systems resilience and sustainable development.

Systems biology

At the frontier between environmental and biological sciences, the activity on systems biology gathers communities of biologists, physicists, biophysicists and biochemists, mathematicians and computer scientists. High flow of data originated from genomic and proteomics analysis, from quantitative spectroscopy of a population of unicellular organisms (describing metabolic networks with a stochastic variability between individuals) or from neuron and cells activities, require advanced theoretical methods to interpret data and optimize experimental time and cost. Several teams contribute to the development in Montpellier of collaborations in systems biology – with models combining the principles of theoretical physics, fundamental computer science and mathematics, such as models of gene networks dynamics, models of tissue growth, or as inherently multiscale models.

The Virtual Plant team performs analysis and multi-scale modeling of plant architecture (e.g. analysis and modeling of branching structures, multi-scale plant geometry and topology), development of mathematical models of plant morphogenesis at different scales (e.g. 3D meristem reconstruction, models of organ morphogenesis, branching system development and root initiation) and software platform for plant modeling (e.g. OpenAlea). These models are often based on the analysis of new and complex data like 4D digital images and allow aggregation of heterogeneous numerical models at various organization levels.

In recent collaborations in systems biology, studies focus on the transcriptional network structure and dynamics of nitrate uptake in plants in order to quantify and optimize agricultural impact on the environment) (INRA/SupAgro, L2C, DIMNP, Catholic University of Chile, NYU), or on the metabolic networks of E. Coli bacteria populations via modeling and spectroscopy techniques (Fluorescent correlation spectroscopy, DIMNP, CBS). These studies are opening a way to implement the systemic approaches to entire organisms or populations of cells in view of applications in agronomy or bacteria endemic behavior. Other interdisciplinary collaborations (DIMNP) started on the modeling of phospholipid biosynthetic pathways in Plasmodium, the causal agent of malaria disease, while modeling and experimental studies on biofilms and bacterial growth will be further developed.

In cardiology, systems biology approaches are used to model the ability of one single heart cell to pulse via a complex and collective regulation of ion channels and pumps (IGF).

Merging systems biology with biophysics and biological physics

It is important to stress that systems biology approaches do not merge yet with the biological physics and biophysics descriptions of cellular systems. On the one hand, numerous mechanistic dynamical models of cellular molecular mechanisms have been proposed, characterized by rather high levels of abstraction and relatively small sizes, based on low-throughput and often unrelated data sets, which can overlook the co-existence and co-functioning of heterogeneous populations of cells in a tissue or a population in an environment. On the other hand, reconstruction of predictive mathematical models from functional genomic data sets remains a biological challenge, also in reason of the appearance of stochastic behavior in the same class of individuals or biomolecules. Moreover, many regulatory mechanisms interact and operate at molecular and supramolecular levels (DNA methylation, chromatin remodeling, transcription, protein modification) and depend on biophysical processes that are not usually taken into account in systemic approaches.

Regulatory network reconstructions merge data without almost any spatial and temporal characterization making extremely difficult experiments and actual bioinformatics and mathematical models to keep into account spatial and temporal behavior of networks. In this context, bridging detailed knowledge of molecular mechanisms with systems level description of cellular functions to unify systems biology, biophysical and biological physics approaches stands as a great challenge for future research.

Actions for the development of the above ideas have been proposed in the LabEx EpiGenMed that aims to turn new paradigms in biology into therapeutic strategies of tomorrow. These goals naturally converge with the LabEx NUMEV on modeling complex systems through ongoing collaborations like local research groups and collaborations such as GRISBI.

In the same logic, a national ANR project DYALOG will be submitted between different national laboratories (ENS, Institut Pasteur, and Institut Curie, Paris; DIMNP, IGH and IGMM, Montpellier; INRIA-Rocquencourt; IBISC, U. Evry/Genopole; TAGC, Marseille). The project development can provide an important expertise to contribute or build a modeling platform in the next years with the specificity of combining epigenetics to biophysics and biological physics (by including the biophysics of the chromatin and its dynamics, the intracellular transport processes and cytoskeleton dynamics and remodeling associated to signalization processes) with numerous outcomes and applications in biology and health. Theoretical teams as the “Biological Physics and Systems Biology” at DIMNP could host such platform due to its central position in interdisciplinary collaborations with IGH, IGMM,CBS, CRBM, L2C, HPC@LR and possible future LabEx projects.

Numerical methods applied to complex fluids, complex materials, plants mechanics and biomechanics

Numerical methods are applied to coupled problems in fluid mechanics and chemical physics developed to understand mechanisms of soil pollution, for modeling complex behaviors of granular matter systems, also within dedicated software platforms for simulations, and open to a wide range of applications for natural material (soils, rocks, plants), for civil engineering, transport engineering, food processing, and biomedical or pharmaceutical research. Mechanics and growth of plants and human biomechanics complete the pool of approaches in numerical modeling of living systems (LMGC).

Research in the domain of satellite Earth observation techniques (TETIS, AMAP, ESPACE-DEV,) lead to the spatial and temporal coupling of electromagnetic wave propagation models (SAR imagery, spatial interferometry, temporal interferometry, altimetry, Lidar, optical) with surface architecture, structure and texture models (water surface, trees, crops, soils), in order to simulate satellite data acquisition. Surface characteristics can then be derived from satellite data through inverse modeling. This approach is implemented to develop Earth observation techniques for the quantification of river water level, slope, surface velocity and discharge, for the quantification of biomass and forest 3D characteristics, for the quantification of soil properties, for the monitoring of agricultural crops and forecasting of crop yield.

Molecular modeling, macromolecular structures prediction and reconstruction, 4D organ reconstruction

Ab-initio computing and statistical physics coarse grained approaches applied to molecular systems under thermodynamical constraints (CBS), molecular dynamics combined with bioinformatics, theoretical methods to assess molecular and macromolecular structures from experimental data (CRBM, IGMM and CBS) and spatio temporal image reconstruction of organs in vivo (IGF) are topics that join the theoretical effort in describing macromolecular systems in living matter in strong relation with experiments.

Mathematical and statistical modeling in ecology and agronomy

Microbial ecosystems, such as the ones met in soil, are complex systems that present various spatial and temporal scales. There is a huge demand of mathematical models that take into account both individual and population levels. Microbial ecology raises new research questions on the role of biodiversity, in connection with experiments performed on reconstituted ecosystems and the huge data provided by the recent tools of molecular biology. Also, challenges related to plant adaptation to climate change, increasing demand for plants and plant by-products for food and non-food uses, require new mathematical and statistical models. There is an expectation of new concepts to be developed. For these purposes, teams like MODEMIC develop multi-scale approaches and gather the know-how of researchers specialists of dynamics systems, statistical and automatic control.

Functional ecology, organisms evolution and behavioral sciences (research field of several teams at CEFE and ISEM) also raise new research questions from a single individual to the population level, namely from the role of spatialization up to evolutionary schemes, and require new mathematical models. Interactions with experiments performed on reconstituted microbial ecosystems or animals monitored in their natural environment can test models and concepts in these disciplines, but also propose new concepts to quantify biodiversity, complexity of interactions, functional redundancy or evolution of lines.

Biodiversity characterization (structure and organization of a single plant and clusters of vegetations, relation with the Mediterranean territory, phylogeny, …) are goals of an important effort devoted to couple botany, ecology and agronomy disciplines with computer modeling (AMAP).

Characterizing and quantifying the impact of external pressures (both natural –ex. climate change- and man induced –ex. pollution) on biodiversity and ecosystems is a key objective of several research teams (ex. TETIS). They implement spatialized approaches, including characterization of habitats, characterization of land use, determination of spatial indicators of pressures taking into account morphology (fragmentation of ecosystems, connectivity) and geographical distribution (for instance pressures along river networks), and spatialized modeling of pressures/impact relations. These studies and model development aim both at improving our understanding of biodiversity dynamics in response to changes in the environment, and to elaborate mitigation and restoration strategies for implementation of public policies. Beyond the modeling issue, developing reliable methods to characterize the uncertainty of such models is a key stage for further operational application.

In the domain of agronomy a major challenge lays in developing global agricultural monitoring systems. This implies regional scale quantification of cultivated areas, identification of crops and forecasting of crop yield depending on meteorological scenarios and agricultural practices. Spatialized agronomical models must be fed with meteorological models outputs and assimilate in situ and satellite data, using stochastic approaches to simulate ranges of scenarios.

Mathematical and statistical modeling in environmental sciences and water sciences

Simulations of water resources with regards to climatic and anthropogenic forcing should rely on tools shared on an interdisciplinary basis: joint observation schemes on hydrological regions, experimentation media in controlled conditions, platform for modeling to serve the different fields of investigation and offer the opportunity to perform simulations under various simple or combined forcing phenomena. Some efforts should be done with regards to the downscaling of GCM output to assess climate change at local scale; these downscaled data will then help to better constrain regional groundwater or surface water models for a better management in the near future (short and mid-term).

Other modeling issues related to water science deal with the simulation of water resource dynamics and management at basin scale taking into account not only hydrophysical processes but also socioeconomic process related with users strategies through Multi-Agent models (GEAU, GREEN).

Simulating in silico spectroscopy and microscopy experiments and cell/organ simulators

Experimental advances also open the opportunity to model in-silico experimental signals in non-invasive, quantitative and combined detection with ultra-resolution in time and space in complex intracellular environments. This approach can evaluate possible counterintuitive or erroneous interpretations of experimental results, simulating the signal of combined experimental techniques and setting the bases for detailed models in the context of systems biology. First collaborations between modeling and experimental teams (DIMNP, CBS, CRBM, IGMM) started by simulating complex and collective behaviors from the knowledge acquired on single molecules and macromolecular assemblies. This approach is extremely effective to develop interdisciplinary collaborations, but relates also to the development of concepts from the modeling approaches of complex systems and other points addressed in this report like data management and data acquisition. Finally, this topic is a potential development for many cell simulators in the perspective of developing artificial life simulators and experiments or building a virtual cell or a virtual organ approach.

2.4. Proposed Actions

The research directions on modeling and simulation we described above are part of the NUMEV and EpiGenMed Labex projects for theoretical and quantitative approaches to complex systems in biological and environmental sciences. The unifying force and the national and international renown of Labex initiatives can already contribute to the development of research and training multidisciplinary complex systems and enhancement of our scientific community in Montpellier.

However, facing at the large diversity and dispersion of scientists and scientific communities involved in modeling and simulation over the all Montpellier area, a more efficient coordination is possible, not to say required. The sharing of concepts, methods, tools and "theorems", fosters working groups and research activities and helps in accelerating our investigations on complex systems in biological and environmental sciences. Individuals and working groups should be all involved in developing interdisciplinary synergies for modeling and simulations related to life and environments.
On the other hand, such large community requires also national and international visibility in research as well as in training for research, in order to better develop activities with other institutions (local, national and international), with territorial communities and organisms (hospitals, city-halls, regions, …) and industrial partners.

In doing so, we should capitalize on the experience already gained by existing and dynamically operating communities and collaborations. Such experience is gathered from renowned open-walls collaborations like the "Association de la Montagne Sainte-Geneviève" (organizing initiatives of many prestigious institutions in fundamental and applied sciences of Paris 5th District). It can be very inspiring for applying a similar structure to Montpellier research, facilities and human potential facing to world-wide competition.
In this context in Montpellier, interdisciplinary research on biological systems and related problems in environmental sciences has been promoted since 2004 by supporting activity of the biology and physics teams with openings to mathematics, mechanics, chemistry, and bioinformatics communities. The GRISBI group (Groupement de Recherche Interdisciplinaire des Systèmes Biologiques), created in the end of 2005, represents an open community (more than 70 researchers belonging to about 10 different laboratories) in theoretical and experimental approaches to biological systems and in related environmental problems. Such dynamic "open-walls" structure is very functional for the variety of institutions, sparsely distributed as in Montpellier. GRISBI proposes seminars on interdisciplinary topics, diffuses seminar announcements from other communities, organize workshops, develops among students the training for interdisciplinary research and, in particular, shares interdisciplinary competences on theoretical and experimental levels. Thanks to GRISBI multidisciplinary collaborations started, inspiring also more focused meetings and working groups (like for the "Biophysics Group", CBS, L2C and DIMNP laboratories, and the newly formed working group in "Physics of complex systems in life and environmental sciences", L2C and DIMNP), but also by giving a very important visibility to Montpellier interdisciplinary sciences on national (e.g. GDR) and international grounds.

Scientists belonging to GRISBI, indeed, count collaborations with the major national and international institutions involved in interdisciplinary approaches to biosystems (Ecole Normale Supérieure, Institut Pasteur, Institut Curie, Max Planck Institute (Germany), National Institute of Health (USA), to cite some of them) and their interdisciplinary activities are supported by UM2 and national organisms like CNRS Institutes and Interdisciplinary programs (e.g. "Interface Physics/Chemistry/Biology Interface: Support of risk taking") and ANR (if one limits to France).

In this perspective and at the light of the topics studied, it is our priority to foster GRISBI as a major scientific environment to 1) exchange ideas and tools by events (seminars, working groups, workshops, schools,…) and by dynamical databases for collecting and distribute information on instruments, techniques and human resources (i.e. competences, tools and approaches in experimental techniques and platforms, theoretical and numerical methods); 2) promote interdisciplinary research among scientists and students, 3) raise the awareness of scientists to interface their activities with other initiatives and institutions; 4) efficiently promote collaborations, among regional, national and international teams for answering to calls for projects by the ANR, the European Community and the Human Sciences Frontiers Program grants, to cite few examples; 5) improve Montpellier scientific area renown at national and international levels, and 6) develop collaborations with territorial communities, organisms and industrial partners.

The specific actions should support GRISBI "open-walls" collaboration participants in modeling and simulation activities with: human resources (PhD students and postdocs), engineers and technical assistant positions, researchers, teaching/researchers positions; grants for scientific collaborations (funds for collaboration start-ups between working groups, travels, short and long term visiting grants); support for collaborative and exchange activities (workshops, conferences, schools, …).

As an open-walls and dynamic collaborative group, GRISBI will efficiently boost attractiveness in research and education on interdisciplinary topics in life and environmental sciences, disciplines in rapid development and with a very high social impact.

3. Algorithms and Computation

Key-words: numerical methods, combinatorial algorithms, parallel processing, HPC, cluster computing, environment and life-science applications.

3.1. Context and motivations

In a number of scientific domains, data are growing dramatically, often at an exponential rate, and even faster than Moore's law characterizing the computational power of computers. Environmental data (e.g. satellite images) and life science data (e.g. the complete genomes of numerous species) are at the forefront of this deluge. Moreover, knowledge and models on these data are increasingly complex and accurate, and we have increasingly high expectations from their treatment and analysis, not only to resolve scientific questions (e.g. the origin of life, ecosystem dynamics ...), but also to address critical societal issues (e.g. health, medicine, impact of climate changes, demographic evolution, strategies for sustainable development...). This is a major challenge for computational methods and algorithms, since one must essentially be able to extract relevant information and perform more complex calculations on more data and in less time, with computers whose power does not increase as fast as data size increases. It is also expected for these computational methods and their results to be valid and certified, with proven guarantees of optimality and accuracy, including numerical precision.

Although the international community is largely involved in these issues, data are getting produced faster than our ability to analyze them with traditional techniques. The 1000 Human Genome project is exemplary in this respect (http://www.1000genomes.org). Launched nearly 3 years ago, the project has far exceeded its goals, now with about 2,500 sequenced genomes belonging to some 30 populations around the Earth. However, information extracted from these data is still low (only one article published very recently), whereas the data surely contain a wealth of information on human history and biology. Similar examples are found in other areas, such as that of environmental satellite images. More than 60 Earth Observation satellites currently monitor the planet using various types of sensors (optical multispectral and hyperspectral, thermal, passive and active microwave, interferometers, altimeters, lidar, accelerometers, .) to provide raw data on surface emissivity, reflectivity, roughness, vertical structure, from which information can be derived about land use, vegetation type and status, temperature, water level and colour, elevation models, soil moisture, precipitations, snow, Earth gravity field, ocean salinity, atmosphere chemical composition,… Extracting relevant information, coupling satellite data with in situ data, assimilating them in models and making full use of this extremely rich spatial and temporal information for environmental sciences is a major challenge still hard to reach.

The main current limitation in the scientific use of these data lies in the lack of innovative mathematical and computational methods to process the data, and the difficultly of scaling to match their dramatic increase.

3.2. Situation in Montpellier

In this context, researchers of Montpellier and Languedoc-Roussillon should enable significant insights, both on the theoretical and practical sides.
The "Algorithms and Computation" theme involves teams from the fields of computer science (LIRMM, CINES, U. Perpignan), mathematics (I3M), microelectronics (SYSMIC), and physics (LMGC). This gives rise to a combination of a broad spectrum of complementary skills. This spectrum covers complexity and information theories, algorithmics, practical techniques to solve NP-hard problems, Monte Carlo probabilistic approaches, simulation of complex systems, design of dedicated architectures, high-performance computing, and processing of high-throughput data from biology (e.g. genomics) and environment (e.g. satellite and distributed field sensors data). The teams involved benefit from high international visibility in their fields. They have already published major results on various aspects of algorithmics, for instance to solve constraint problems with filtering [Bessiere 2005], in parameterized complexity [Heggerness 2007], or on approximation algorithms [Berry 2009]. The paper [Guindon 2003], which describes an efficient algorithm for likelihood-based phylogeny reconstruction, has been the most cited paper in ecology and environment since October 2007 (Science Watch – Thomson Reuters). To perform calculations, do research on various computer science aspects (e.g. massive parallelism) and distribute the results worldwide, they will use the HPC@LR facilities, the new cluster asked to the Avenir Infrastructure Call (1000 CPU Cores, 1PB storage), and the CINES computational power. Our goal is to gather all these skills for a twofold scaling challenge, both quantitative because of the huge size of the data, and qualitative because of the increasing complexity of the models and the diversity and specificities of the thematics they address. The theoretical results, the concepts and the tools that will be developed to address these challenges will have a generic impact that will reach beyond the primary objective of scientific data analysis. They will provide valuable contributions to computer and mathematical sciences and related fields.

3.3. Research directions

The emergence of new data in exponentially-growing quantities, combined with the essential questions posed by environmental and life (and other) sciences, requires fundamental research on the multidisciplinary formalization of these questions, their intrinsic algorithmic complexity, the design of effective methods to analyze them in the most effective manner, and the certification and testing of these methods on large simulated and real data sets. Specific works will be undertaken in each of these directions by the teams involved in this theme.

Combinatorial algorithms

Discrete combinatorial objects, such as graphs, trees and strings, lead to a large number of simple and meaningful models, especially (but not only) in the domain of computational biology (e.g. regulation networks, phylogenetic trees, genomic sequences...). Although these models may be well known and studied for some time, the drastic increase in data requires increasingly efficient algorithmic solutions. To that aim, new paradigms, techniques and thus algorithmic theories are being developed. In this trend of research, we will study the mathematical relationship between such discrete structures and geometrical representations or models. We will push our research further on the theory of parameterized complexity and approximation, with special interest in biologically motivated problems. We will consider indexing and string algorithms to tackle problems arising from massive DNA sequencing. We will work on the use of graph algorithms applied to image compression, protection and transmission and to the compression and reconstruction of plant structures.

Numerical methods

In collaboration with the researchers in modeling, the goal is to foster interactions between researchers around advanced numerical methods for problems arising from environment and life sciences: numerical optimization (e.g. for coastal erosion problems), discretization methods for partial differential equations satisfying restrictive physical and mathematical properties (e.g. for anisotropic diffusion in porous media), and Lagrangian-type discrete finite element methods, coupled with continuous media. Emphasis will be placed on probabilistic Monte-Carlo approaches, widely used for solving complex problems with no explicit numerical solution, for which we will develop adaptive techniques to adjust implementation parameter values during execution.

Filtering, problem reformulation, heuristics

Many of the problems faced when analyzing data from environmental and life sciences are inherently difficult (typically NP-hard). We can then use techniques that filter the problem at hand or reformulate it to produce a simpler and smaller instance. We can also use heuristics which do not guarantee optimality of the result but usually produce solutions of good quality. In this perspective, we will continue our work on kernelization methods in graph theory. In artificial intelligence, we will propose a constraint solver that uses on-line learning to select the best variable ordering and the best level of filtering during search. We will also extend our results on non-decomposability of global constraints to languages other than propositional logic. In bioinformatics, where the use of heuristics is standard, our goal will be to improve our theoretical knowledge of popular heuristics, demonstrate their mathematical properties, and reduce their time complexity without losing quality in the resulting solutions. Specific works will be dedicated to solving inverse model problems and the conditions of stabiliy to derive inputs from observed outputs with uncertainty characteristics.

HPC

To benefit from algorithm improvement, efficient and reliable implementations of large scale scientific computations are required, as it has been recently highlighted in Nature [Merali 2010]. Reducing the numerical precision (e.g. only using 32-bit floating-point arithmetic) is, for instance, a classic way to lower the running time. Hence algorithms and their implementations have to be re-designed to reduce any additional loss of accuracy introduced by the finite precision computations. Another approach is to benefit from high performance implementations of exact computations for algebraic or linear algebra issues. New architectures for massively parallel computing (e.g. many-core, GPU, Intel's Tera-scale project) justify important modifications of current algorithms and models. Such performance optimizations have already been carried out for current programs solving partial differential equations. However, since the optimal efficiency of the new architectures is difficult to achieve, the development of software tools that simplify the programming phase is mandatory. To stay on the cutting-edge of the next future parallel computation, The University of Montpellier has launched HPC@LR, a computing meso-centre that seeks to host and facilitate exchanges between architecture and software developers, and researchers from other laboratories more involved with applications.

Environment and life-science applications

Real and essential applications, where the scaling described above is already taking place, constitute the objective and test bed of concepts, procedures and techniques developed in this theme. Notably, we will process data from new high-throughput DNA sequencing techniques, whether to process this data at a very large scale using innovative algorithms, to characterize the genetic diversity of populations by combining coalescent theory and fast Monte Carlo methods, or to decipher pathogenic genomes using statistical learning approaches. In the treatment of environmental data, we will work on visualization and fusion, coping with the central issue of heterogeneity, such as spatial resolution heterogeneity that arises when combining satellite data with ground measurements. Another flagship application will be to simulate complex environmental systems, such as modeling and simulation of plant growth at different scales and under the control of genes, granular media (soils, sediments, rocks or snow), coastal systems, aquatic ecosystems, hydrological systems.

3.4. Proposed actions

A large group of researchers involved in these questions in Montpellier, with focus on molecular and cellular biology, will be grouped in the Institute of Computational Biology (Institut de Biologie Computationnelle – IBC), an interdisciplinary project center currently being elaborated as a GIS (Groupement d'Intérêt Scientifique), a Service and Research Laboratory (Unité de Service et de Recherche – USR) or a similar structure, supported by UM2, UM1, CNRS, INRIA, CIRAD, IRD, CEA…. The goal is two-fold: to bring together Montpellier researchers involved with modeling, processing and analyzing data on a large scale in the fields of biology, health, agronomy and environment, and to attract top national and international researchers working in these areas. Several branches of computer science and mathematics will be combined, in order to make progress towards resolving major issues in biology and related fields (e.g. cancer, pathogens, plant genomics, biodiversity). IBC will become a privileged interdisciplinary meeting place, not only bringing together researchers involved with this project, but also involving a large community of academic and industry researchers on regional, national and international levels. An important part of IBC activity will be dedicated to inviting world-class researchers to collaborate with, organizing scientific events, training young researchers, and promoting results and exchanging information with industrial partners. The service side, which is highly important in all biology sectors today, will be handled by platforms with which we are already involved, most notably RENABI-GrandSud for bioinformatics, which we are heading. IBC will serve as a interdisciplinary gateway between discipline-specific groups (PFRs, LABEXs) in Montpellier and the surrounding area. Strong ties will be established with regional companies and start-ups, leveraging the assistance of regional start-up incubators, in which IBC partners already play a key role.

IBC will leverage this gateway between industry and academia, and between scientific domains, by providing:

- technical staff to facilitate collaborations;
- PhD and postdoc grants to attract top international students;
- core funding to achieve long-term research;
- setup with specific accommodations; IBC will start with ~50 permanent researchers (part- or full-time in the institute, in agreement with their original laboratories) and nearly the same number of PhDs, postdocs and short-term contract researchers, but will most likely increase in the coming years.

IBC's flexible structure should serve as a test-case for further similar interdisciplinary centers, for example focused on the interplay between ecology, mathematics and computer science, or on the impact of climate changes on human populations.

4. Scientific data Management

Key-words: scientific data, data processing, data analysis, data mining, scientific workflow management, data integration, parallel query processing, cluster computing, cloud computing, data security, data privacy.

4.1. Context and motivations

Modern science such as agronomy, environmental science, life science, physics, must deal with overwhelming amounts of experimental data produced through empirical observation and simulation (http://www.computational-sustainability.org). Such data must be processed (cleaned, transformed, analyzed) in all kinds of ways in order to draw new conclusions, prove scientific theories and produce knowledge. However, constant progress in scientific observational instruments (e.g. satellites, sensors, large hadron collider) and simulation tools (that foster in silico experimentation, as opposed to traditional in situ experimentation) creates a huge data overload. For example, a recent report states that climate model data are growing faster than the data set size for any other scientific discipline, with collections of hundreds of exabytes expected by 2020 (http://www.er.doe.gov/ascr/Funding/Notices/LAB10-256.pdf, http://www.ncdc.noaa.gov/sds/index.html).

Scientific data is also very complex as the result of heterogeneous methods used for producing data and of the inherently multi-physics, multi-scale nature (spatial scale, temporal scale) of many sciences, resulting in data with hundreds of attributes or dimensions. Processing and analyzing such massive sets of complex scientific data is therefore a major challenge since solutions must combine scalable data management techniques in large-scale cluster, grid or cloud environments.

Furthermore, modern science research is a highly collaborative process, involving scientists from different disciplines (e.g. biologists, soil scientists, and geologists working on an environmental project), perhaps from different organizations distributed in different countries. Since each discipline or organization tends to produce and manage its own data, in specific formats, with its own processes, integrating distributed data and processes through scientific workflows gets difficult as the amounts of heterogeneous data grow.

Despite their variety, we can identify common features of scientific data [Ailamaki 2010]: very large size and massive scale; manipulated through complex, distributed workflows; typically complex, e.g. multidimensional or graph-based; with uncertainty in the data values, e.g., to reflect data capture or observation; important metadata about experiments and their provenance; heavy floating-point computation; and mostly append-only (with rare updates).

Unfortunately, generic data management solutions (e.g. relational DBMS) which have proved effective in many application domains (e.g. business transactions) are not efficient at dealing with scientific data, thereby forcing scientists to build ad-hoc solutions which are labor-intensive and cannot scale. The same observation generally applies to cloud data management as well. Cloud data can be very large (e.g., text-based), unstructured or semi-structured, and typically append-only (with rare updates). And cloud users and application developers may be in high numbers, but not DBMS experts. Therefore, current cloud data management solutions (e.g. Google Bigtable, Google File System, Amazon SimpleDB, MapReduce) have traded consistency for scalability, simplicity and flexibility [Özsu 2011]. As alternative to relational DBMS (which use the standard SQL language), these alternative solutions have been recently quoted as Not Only SQL (NOSQL) by the data management research community.

The overall goal of scientific data management is to make scientific data easier to access, reproduce, and share by scientists of different disciplines and institutions. In recent international interdisciplinary workshops, the following key requirements for scientific data management (that cannot be supported by current technology) have been identified : (1) rich representation of scientific data (multi-dimensional arrays, graph structures, sequences, etc.); (2) built-in support for managing and uncertain data (e.g. inaccurate data generated by faulty sensors or by imprecise observations); (3) distributed workflow execution involving large numbers of distributed processes and large amounts of heterogeneous data, with support of data provenance (lineage) to understand result data; (4) scalability to 100s of petabytes and 1,000s of nodes in high performance computing environments (e.g. very large clusters); (5) efficient data and metadata management, in particular, with semantics (ontologies), in order to help reasoning over data with different semantics (6) Open Source Software – a key requirement of scientists to ensure data is never locked in proprietary systems.

4.2. Scientific data management in Montpellier

The teams that can be involved in scientific data management in Montpellier have both strong expertise in data management and long experience of collaborations with the local scientific community in agronomy and environment.

In computer science, INRIA and LIRMM have extensive experience in distributed data management, in particular, data integration and workflow management. For instance, the INRIA-LIRMM Zenith research team has developed schema matching tools and is developing the WebSmatch environment. Within a collaborative project with UFRJ (Brazil), Zenith has been working on scientific workflow management. The SMILE team at LIRMM and the INRIA-LIRMM GRAPHIK teams have extensive experience with semantic data management based on ontologies. The TATOO team at LIRMM has longstanding experience in data mining. The SYSMIC team at LIRMM has strong expertise in data security.

In terms of High Performance Computing, Montpellier has much potential. First, there is CINES which is a national service for high performance computing, with the 7th most powerful supercomputer in the world. There is also a powerful computing grid at UM2 (http://webgrille.info-ufr.univ-montp2.fr/) that results from the collaboration between several laboratories (ISEM, LIRMM, IES and LPTA).

In agronomy, the main partners with data management expertise are CIRAD, INRA and IRD. The ID team at CIRAD includes a large team of bioinformaticians that has been involved in information system development and feature annotation on biological sequences. At INRA, plant biologists have used database management systems to store the data produced by plant sensors and perform statistical analysis. At IRD, in collaboration with LIRMM, scientists have developed data warehouses for analyzing massive amounts of agronomics data. In environmental science, the primary partner is IRSTEA, in particular, Maison de la Télédétection, with expertise in developing and operating environmental information systems with data coming from satellite imaging.

Overall, there is an excellent combination of teams with strong background on all aspects of data management and teams with expertise in using data management techniques for their applications.

4.3. Research directions

Addressing the requirements of scientific data management is on the agenda of a very large international research community of scientists of different disciplines and data management researchers both from academy and industry (http://scientificdatasharing.com). For instance, the SciDB organization (http://www.scidb.org) is building an open source database system for data-intensive scientific applications. SciDB will be certainly effective for similar applications for which the data is well understood (with well-defined models). However, SciDB addresses only one part of the problem (a centralized scientific DBMS). We need to better understand the fundamental aspects of scientific data management, in relationship with the main users, i.e. scientists, by capitalizing on the principles of distributed data management. Our vision is that no single system will be able to accommodate all the above requirements of scientific data management. Thus, we need to think in terms of interoperable solutions (implemented by components and services), to be used in different, even unanticipated, combinations, to address scientific applications with specific needs. Therefore, based on the expertise of the teams in Montpellier, we see the following complementary research directions where we can have strong impact. These research directions have been written by the Scientific Data Management group (headed by P. Valduriez) in the LabEx NUMEV.

Uncertain data management

Data uncertainty in scientific applications can be due to many different reasons: incomplete knowledge of the underlying system, inexact model parameters, inaccurate representation of initial boundary conditions, inaccuracy in equipments, etc. For instance, in the study of plant evolution at INRA Montpellier, each plant has several sensors monitoring different parameters, e.g. size and temperature. Every 15 minutes, new data from each sensor captures new data, sometimes inconsistent or contradictory. Instead of ignoring (or correcting) uncertainty, which may generate major errors, we need be able to manage it rigorously and provide support for querying. A promising approach is to model data uncertainty with probability theory, which provides rich computation capabilities. For instance, the uncertainty that a data item (e.g. a record) exists can be represented by a probability attribute while the uncertainty over an attribute's value can be represented with different values, each one with a probability. This makes it possible to compute a probability score for each data item that a query matches, and thus rank query results in a way that can be exploited by the user or application.

Data integration and visualization

In modern science, for understanding a phenomenon of interest, a major challenge is to extract, integrate, analyze and visualize data from a high number of heterogeneous sources. Source heterogeneity makes it hard, for instance, to perform data fusion (e.g. to merge data sources from remotely-sensed, in-situ observations with large-scale supercomputer-based models in physics and chemistry). Even a single data source can have a wide variety of datasets such as (1) datacubes representing annotated blocks of data in one, two or more dimensions (e.g. time-series or spectra for one dimension, images or frequency time spectra for two dimensions); (2) multi-parameter data coming directly from sensors or datacubes; (3) sequences of symbols (e.g. representing biological genes in genomics); texts coming from scientific journals or social networks, etc. Furthermore, data is typically multi-scale (micro/macro observation) and multi-granularity (different levels of precision or uncertainty), thus making it hard to integrate efficiently. Data visualization of large multi-dimensional datasets (e.g. to relate and visualize outputs from stellar astrophysics models that are essentially n-dimensional) is also important for scientists and requires efficient techniques that scale up.

Semantic-based data and metadata management

In science today, each research project creates and publishes its resulting datasets using its own representations. Thus, distributed Scientific Data Libraries, where data management tools (e.g. for storage, archiving, analysis, visualization) can work across collections, will allow capitalizing on data. But to produce usable information from those data, semantic metadata and representations of knowledge about the modeled scientific domains (ontologies) are crucial. This requires a new generation of data management tools, able to do large-scale reasoning on large data sets, while exploiting semantics encoded within ontologies. Dealing with heterogeneous conceptualizations makes this challenge even more difficult, in particular, to cope with inconsistencies, conflicting viewpoints, multi-scale and multi-granularity representations. Working across several domains makes these tasks even more complex, and requires the definition of cross-domain features of the data.

Online data processing

Scientific applications may directly produce streams of data with large numbers of sensors (e.g. in weather forecast, environmental studies, road traffic, power plants, etc.). As the volume of data increases, it becomes very hard, sometimes impossible, to store all available data before processing them. Recently, a great deal of research has been devoted to the development of methods and tools capable of processing data streams "on the fly", without storing the entire data. However, scientific applications pose new challenges for processing streams of heterogeneous, distributed and constantly evolving data. Furthermore, in many applications, there is a need to keep an historical view of the streams, e.g. to provide historical aggregate information from the streams or to detect anomalous behavior of monitored systems. This issue is critical when dealing with concept drift or changes due to environmental conditions. Since historical data may be not provided, how can we take into account periodicity of environmental data: diurnal, seasonal, weather-based changes? The physical concern of sensors, e.g. their limited energy, must also be considered, in particular to select the algorithms to be used in data mining. Finally, to deal with massive amounts of data to be processed, parallel algorithms that can exploit modern cluster or cloud environments are necessary.

Scientific workflow management

Scientific workflow management systems (SWfMS) allow scientists to describe and execute complex scientific procedures and activities, by automating data derivation processes, and supporting various functions such as provenance management, queries, reuse, etc. Some workflow activities may access or produce huge amounts of data and demand high performance computing (HPC) environments to improve throughput and performance. In addition, some of the data sources and computing resources involved in a workflow may be high distributed. Combining SWfMS with HPC environments and distributed resources remains difficult, involving complex middlewares that need to exchange data between heterogeneous environments.

Security and privacy of sensitive data

Some important scientific data may be very sensitive, e.g. patient medical exams, critical sensor data, and require secure transmission and storage. As the use of sensitive data increases rapidly, so does the likelihood of their interception and misuse. Therefore, data integrity and security become critical. Furthermore, it may be important to confirm the authenticity of a sender and protect users from being impersonated. These problems will only get worse as new data transmission technologies are developed. Thus, protecting critical data from fraudulent or inappropriate use is essential. Although many cryptographic techniques have been proposed (hashing, encryption/decryption, digital signature, authentication, etc.), the amount and variety of data that need to be protected should be taken into account in order to develop efficient algorithms that are suited to various scientific communities.

4.4. Proposed actions

The teams and laboratories involved in scientific data-intensive applications in Montpellier have both strong expertise in data management and long experience of collaborations in agronomy and environment. Thus, they are in a unique position to propose an ambitious, original research program based on the research directions we described. Based on our analysis of the local situation, we propose the following actions to foster an ecosystem (with training – research – transfer) of international visibility in scientific data management.

The research directions we described are part of the labex NUMEV. A more focused subset of this program is also part of the IBC proposal. In these two proposals, data management is viewed as important in conjunction with modeling and simulation. As specific actions, these two proposals should be strongly encouraged by UM2 and its partners, in particular, by dedicating human resources (PhD students, engineers and researchers).

We expect these actions to have an important social and economic impact at the regional, national and international levels. At the regional level, it will foster collaboration between mathematicians and computer scientists involved in data management and data analysis with scientists of different disciplines such as agronomy and environmental science, life science, etc. which are strong in Montpellier. Although collaboration between these communities already exists in different forms, focused projects such as NUMEV or IBC will strengthen collaboration by giving it more visibility, which will in turn generate interest from other participants, including many local start-ups which have to deal with scientific data.

At the national and international levels, the impact will be a natural continuation of the local impact and a consequence of the global nature of scientific data management that is boarder-free. Since all scientists participating in the project have ongoing collaboration with peer teams at the national and international level, such project will simply strengthen those collaborations.

The fact that the results will be released as Open Source components, like in SciDB, will help us be part of the more global community of developers in scientific data management and contribute to technology transfers that can be the basis for advanced services of all kinds.

5. Production, Storage and Archiving of Data from the Observation of the Natural Media

Key-words: dataflow, observation, satellite, experimentation, spatial and temporal scales, database, datacenter for environmental and biological observation and experimental data.

5.1. Context and Motivations

The inevitability of global change is proven, and its effects are now being felt by our citizens. Global warming particularly has been the subject of much debate until the recent demonstration of its undoubtedly anthropogenic origin in the last IPCC report. This demonstration has been possible thanks to the synergy between systematic observation of the natural environment and modeling, modeling which success owes as much to the increasing quality of models and computing power than to the intrinsic quality of data both as input and validation of the modeling results. This is also true of the study of all natural hazards and their impact on living systems. The systematic observation of the natural environment, through both in situ sensors and satellite sensors, is thus essential to the study of natural hazards and global change: it can drive the basic research that determines the relevant markers but also allows the establishment of operational systems for forecasting and warning, based on the data themselves or their analysis in terms of mechanism.

Similar arguments apply in monitoring living systems and their constituents at different scales in time and space from single macromolecules to cell, tissues and organisms. Technological advances are tracing a completely new path to understand the complexity of life and living systems. Modern techniques of microscopy and spectroscopy, for example, are non invasive approaches in order to follow, with ultra-resolution in space and time, the complex and basic processes of life biomolecule by biomolecule. On the same basis, experimental techniques developed to read living creatures' genomes, together with bioinformatics approaches, are strongly accelerating the lecture of the hereditary information transmitted by living systems. Moreover, the ability in mastering micro and nanosystems and physico-chemical and biological processes open the access to first examples of artificial life. Conceptually, a basic science like biochemistry has coupled its domain of interest with many other different disciplines (mechanics, physics, hydrodynamics, surface sciences, etc …), since the simple biochemistry processes are not enough to explain why a cell or embryo fates can be strongly dependent also on mechanical constraint or why a cell can sense the elasticity of its internal and external environments.

In this development, large production of data, information exchange and storage in public databases for the scientific community, or in private databases (e.g. to respect individuals' privacy on health information), are in very rapid evolution and constitute the basis for quantitative and predictive biology and for environment sciences. Together with these important developments, the large amount of data produced requires the elaboration of effective modeling. Statistical analysis of data can provide important hints and predictions, but improving the understanding of environmental and life processes also demands the elaboration of new principles and concepts, still unexplored, and their translation -with the help of experiments and simulations- in computational approaches for predicting possible behaviors in time and space. In this perspective, one can consider that the level of complexity of an ecosystem compared with a tissue of cells is probably very similar. Thanks to technological advances (data availability, data modeling, systematic data sharing and organization, computational power availability, nanostructures and nanoenvironments mastering) understanding the rules coding environmental and life processes are great challenge that mathematical and physical science communities are undertaking in view of future conceptual revolutions.

The success of systematic observation lies in the processing chain from raw data acquisition by sensors and human experts, processing the data using first raw treatment then more sophisticated treatment, sharing of this data and integrating it into models and simulations. Quality of the data depends on the overall processing chain, but the nature of treatment required by the division or modeling is also directly impacting on the nature of the sensor, taken in its broadest sense (type of measurement, spatial, frequency of measurement, scales of measurements). Therefore the processing chain can be considered as an information cycle.

The success of this project will therefore lie in its ability to establish interdisciplinary networks that share their experience in the domain of systems observation and information processing chain. Interdisciplinarity is needed between specialists of sensor technologies, computer scientists and designers of tomorrow's computers, and disciplinary scientists themselves since the complexity of living and environmental systems involves hard science for biophysical processes and human and social science for the interaction between individuals and within the human sphere.

The challenge of observation can therefore be seen more "operational" than strictly dedicated to exploring new frontiers in science, in the sense that dealing with large quantities of heterogeneous data is prerequisite to any attempt to understand complex systems, and that production of observational data is largely under way and requires rapid answers. However, experience in science shows that the development of innovative observation capacities has often been the key element for scientific breakthroughs.

5.2. Observation of Natural and Biological Medium in Montpellier

Montpellier is a major actor in the process of providing state of the art data from observation of ambient media or biological activity.

A national Observatory of Mediterranean environment rests on a dense network of multiparameter stations of continuous observation of physical, chemical and biological status of natural media, and most notably: the geophysical monitoring of ground and surface waters especially in karst systems, GPS monitoring of ground movements and tropospheric humidity, the monitoring of reference animal and vegetal species and overall biological state of selected areas, and the monitoring of coastal environment, in terms of nature and spatial variation of the coastal line and as well as focused studies of the continent- sea transition through lagunal systems as viewed from hydrosedimentary, physical and biological points of views. The observatory and several laboratories carry EquipEx and LabEx proposals that aim at reinforcing their observation capabilities, both through equipment and gathering of scientific excellence. Data produced in these platforms range from genomics to tree size, atmosphere humidity of size of fishes, thus enlightening the high heterogeneity of the data.

The GEOSUD Equipex (GEOinformation for Sustainable Development) is the first national satellite imagery infrastructure freely accessible by the scientific community and by public actors involved in environment management and land development. Initiated by the Montpellier scientific Community (TETIS, ESPACE-DEV, OSU-OREME, LIRMM,…) it gathers 14 national institutions. It will provide, among other observation data, annual high resolution (5m) coverages of the national territory for the 2010-2015 period, and will make available processing chain facilities, methods and training.

Another aspect of excellence in Montpellier is the richness of laboratories involved in health and biology research in strong cooperation with Montpellier University and hospitals and the support of organisms like CNRS and INSERM. Studies are performed within specific platforms conceived for developing and sharing new experimental techniques and interdisciplinary knowledge. Indeed, within the last ten years, the LabEx teams opened specific units dedicated to technology development and support to research teams for the use of these technologies in genomics, proteomics, pharmacology–screening, structural biology and biophysics, imaging, animal models facilities, experimental histology network, viral vector facility, protein production facility and genome-wide resources of ready-to-use ORFs and interfering RNAs. These facilities are all independent services open to the research teams in Montpellier, and will be coordinated within the newly created «Structure Fédérative de Recherche» (SFR) BioCampus. These facilities will play a major role in the technological offers to the LabEx teams, and through the coordinated action, and communication of the SFR, any team part of the Biology research campus of Montpellier will be informed of these new developments through meetings, web notes and communication. Such a structural organization will also favor exchanges between the teams interested in using the same technologies.

Among these technology facilities, five obtained a label from IBiSA, illustrating the strong performance of Montpellier technology development units. These facilities will play a major role in the technological offers to the LabEx teams and other laboratories, and through the coordinated action, and communication of the SFR, any team part of the Biology research campus of Montpellier will be informed of these new developments through meetings, web notes and communication. Such a structural organization will also favor exchanges between the teams interested in using the same technologies.

Concerning data for modeling biological systems issued from microscopy and spectroscopy techniques MRI facilities (distributed on nine different sites in Montpellier – see document included) and Structural Biology and Biophysics at CBS play a fundamental role.

Montpellier RIO imaging (supported by CNRS, INRA, CIRAD, UM2, CRLCC) has the following mission:

1. Promoting and ensuring, at optimized costs, the use of imagery (from single molecule to whole small organism) and cytometry, cell sorting (from single cell to embryo), XR imaging in all production processes of knowledges, results or data needed by public institutions or industrialists.
2. Promoting and ensuring the development of the use of imagery, and for the time being, optical microscopy, cytometry and sorting and XR imaging, specifically tomography in all academic and industrial domains where they are not or poorly at work.
3. Ensuring the education and training of a vast, highly qualified workforce from operational control of workstations to procedure, protocol and systems design.

The Structural Biology and Biophysics facility is more dedicated to the development of biophysical techniques like fluorescent correlated spectroscopy, laser and magnetic traps to manipulate single molecules, FRET/FLIM spectroscopy (a similar microscope is present in Montpellier RIO Imaging at CIRAD), NMR. Also a crio-electron microscope is present.

5.3. Research Directions

Although the major challenge in observation of natural and biological medium is more operational, there are also important research directions for which Montpellier could position itself as a leader.

Sensors and information: spatialization, discretization of time and data

The observation of the natural environment is becoming heavily instrumented and induces an increasingly large flow of data, more and more often in real time. This monitoring of key parameters physical, chemical, biological, human in the environment requires the implementation of innovative sensors, sometimes virtual and then assembling data intrinsically of first order.

Whatever the sensors, the key element for all observation network is the choice of spatial discretization and time period of measurement. The choice of mesh for space obviously depends on the dynamics of the phenomenon but the complexity of environmental and living systems often leads the observer to want to discretize space with a mesh infinitely small, which imposes a considerable volume of data and usually inaccessible cost of implementation and maintenance. A first answer is provided by an analysis of sensitivity uncertainty of modeling. Indeed, this type of analysis can identify the parameters having a real impact on the simulation results, whether the input parameters or internal parameter, by studying the response, in space and in time, of variables of the model to small changes in these parameters. The results of these sensitivity tests inform us about the variables less sensitive (boundary conditions, for example), for which there is no need to densify the network of measurement or accuracy of these measures, because of the low gain associated to the understanding and representation of physical processes modeled. In contrast, in the case of very sensitive variable (to internal parameters of the model for example), then these results can help identify areas where intensification of the measurement network will significantly improve the simulation results, and similarly, the optimal sensitivity of measuring devices may be defined in terms of the gain associated with better measurement accuracy.

Spatialization with finer mesh can be achieved through the development of mobile sensors that cover the space in a time step small enough to be considered negligible compared to the characteristic time of evolution of the medium, which is rarely possible. A more effective response lies in the processing of spatially distributed information in an attempt to extract information to a lower resolution than the spatial mesh offered as the model.

Another observation mode is remote sensing, mainly through airborne or satelliteborne sensors, which is not based on the location of sensors in the medium but in the processing of spatial information natively of a spatial nature. Satellite Earth observation is an extremely dynamic domain with high innovative potential for environmental data acquisition and environmental systems monitoring. As mentioned earlier, a large diversity of sensor types (optical multispectral and hyperspectral, thermal, passive and active microwave, interferometers, altimeters, lidar, accelerometers, .) provide raw data on atmosphere properties and on Earth surface response (emissivity, reflectivity, roughness) and vertical structure. From this data, information can be derived about land use, vegetation type and status, temperature, water level and colour, elevation models, soil moisture, precipitations, snow, Earth gravity field, ocean salinity, atmosphere chemical composition,… Research in this domain covers mainly three area : understanding and modeling of signal physics to design sensors and first order processing of raw data; designing satellite missions for research purposes; designing algorithms and processing chains to derive information about environmental systems in various contexts. The scientific community in Montpellier gathers a large number of teams involved in the development and use of satellite Earth observation for environment with a focal point at Maison de la Télédétection (AgroParisTech, IRSTEA, Cirad, IRD) that coordinates the GEOSUD EquipEx is coordinated by this group. Additionnaly, the Solarium group (Systèmes Orbitaux Liés aux Activités de Recherche Interdisciplinaires de L'université Montpellier 2) develops a technological platform to design, build and test small satellites.

The real-time transmission of data is also a lock. Transition to the operational level depends not only on this transmission and the ability of real time analysis of these data, but also on securing the data quality. Experience shows that the exclusive on-site storage of data is a risk of loss greater than the real time transmission. This transmission and its security involve putting in place a process streaming data to verify their integrity, and storage on arrival (indexing etc ...). In the case of satellite data, real time transmission and near-real time processing is of utmost importance to allow synchronized in situ and satellite data acquisition, thus enabling powerful spatialization of information. The GEOSUD satellite reception antenna will be a valuable tool for the Montpellier community.

Monitoring biological realms present different kind of problems and advantages. In first place, biological sensor for in vivo and in organism investigations requires non to be invasive in order to perturb as less as possible cell viability, process ruled by supramolecular assemblies, metabolites and reciprocal interactions. Spatialization in a cell or in an organ is a concept that one should consider in term of "dynamical spatialization" being the sensors applied directly or indirectly to biomolecules targeted by the investigator. In this sense, the intracellular medium of a single cell already presents a large and very dynamical kinds of compartments and scaffoldings rapidly changing to ensure cell functions. In addition, the stochastic nature of molecular events, coupled with the stochastic response of a cell or an organism, introduces variability in data which requires measurement repeatability and reproducibility. On the other side, small size naturally implies the use of all range of progresses in nanosciences, opening the eEra of the "Lab-on-Chip" approaches. In this sense, it is also possible to develop experimental devices coupling more and more detection systems thanks also to electronic miniaturization and computer assisted instruments, gaining in sensibility and consistence of experimental data.

Data storage and sharing

The first issue is that of securing data, which represent increasing volumes. These data must be secured so that their integrity is guaranteed. This imposes hardware processes as well as software and intellectual processes. The hardware approaches aim to develop the infrastructure systems and networks needed to safeguard such large volumes of data. There will be a choice between a backup closer to the user and the center of production of data, or remote in specialized centers. Software and intellectual approaches should build on current thinking about the storage of information including progress made in recent years in the backup splitting of data from "cloud computing". Indeed, the vision of cloud computing promises the provision of reliable, on demand, on the web facilitating access to virtually infinite resources storage, e-network or calculation. Thus, through a simple web interface and a low incremental cost, users can deport complex tasks such as backup and data management in large data centers operated by cloud providers.

But the issue behind the data storage is beyond the simple backup of course, because we are not talking about archiving but about data storage sometimes used in real time, and more generally to perform the nudging of models. Here again, the services of online data storage in the cloud can open new paths. However, one open problem is the interoperability of clouds, and the need to be able to easily change the storage provider without the data being closed.

This nudging of models is commonly used as part of a process of "postauditing" in environmental studies related to pollution of groundwater, for example, or in the oil industry after a phase of developing the exploitation associated with implantation of various drilling.

Above all this storage must not only allow but also induce data sharing. This data sharing also requires a dual approach. Thinking hardware (storage and network architecture) will be needed to make data exchange, physical upload and download quick and easy (especially if storage is remote from the primary user). This line of think must be coupled with a software and intellectual approach of the "intelligence" of data, the modeling of data and construction of metadata and ontology in general, and data mining from such depots. Unspecified information is generally useful only to its primary producer, even then at the condition that it stores the conditions for the measure. To do this, methods related to the definition of ontologies, data wa-

rehousing and data mining will be developed and used. They will contribute to the overall process being as smart as possible, and will also maintain a rational and optimized use of computing resources in a context related to green computing (or Green IT). This latest technological trend aims to improve the energy efficiency of hardware, including hardware solutions (ecological), but also software to optimize resource management.
The issue of intellectual property through the sharing is not addressed here although it is in itself an important issue.

Data archiving

Data must be secured beyond their "instant" use in a form organized and shared. It is common for old data to become necessary and even essential to new scientific questions, for instance analysis of trends and changes. It is then necessary that these data can be retrieved, extracted, with the guarantee that they are as complete as at their initial storage. This continuity requires that the modes of storage, organization and metadata organization be regularly revisited and updated. This work is beyond the capacity of most data producers because following the very rapid change of backup technologies (materials, location of data, etc ...) and ways of organizing data is a profession in full.

Lack of archiving can be a problem. Indeed, it is perceived by the scientific community in biology that many experiments could be just copies of elder works, thus constituting a useless cost for institutions, loss of time for researches and information redundancy. This latter aspect is particularly evident when a name should be attributed to a new protein or a biomolecule. In many cases since the same molecule can assure multiple functions, many different names have been used for the same object, thus making difficult the exchanges between interdisciplinary communities. In the domain of environmental sciences the problem arises for example with efforts to gather historical data on a given issue (for instance natural hazards), that can be found on various physical supports.

National centers, in particular, CINES, are responsible for the archiving of scientific data, including humanities but a national plan to safeguard the hard-science data is under development. Users and producers centers must be encouraged to draw on the expertise of these centers, including those present at CINES, and not to develop local solutions often weak and expensive.

5.4. Proposed Actions

In the objective of fostering strong interdisciplinary collaborations, in environmental and life sciences, a key factor is the local development of state of the art observation techniques covering production, storage and archiving of data. Design of innovative systems to observe and experiment are both necessary for premium science and a natural way for attracting companies in order to built new concepts of multi-users/multi-institutions and multi-partners platforms to generate data and perform data analysis (see the Nikon Platform at Curie Institute) in view of the elaboration of models and simulation. It is also very clear that promoting advanced observation techniques is another way to develop interdisciplinarity, a "tool" dialog being usually much more efficient than a "concept" dialog.

Abroad, there are already astounding examples of observational experimental platforms combined with data management. For instance the American USGS gathering all seismic data obtained in the USA whether it is produced via natural seismicity of experiments, and crossing these data with complex SIG, hydrological and geological data. Another example is located in the Max Planck Institute of Biochemistry in Martinsried (Germany) electron tomography can produce a 3D image of a cell with subnanometric precision and combine structural biology (via the Protein-Data-Base, for example) and bioinformatics knowledge, to automatically classify all sort of protein or organelle in the cell and therefore detecting also unknown typologies and species. Dissemination of satellite Earth Observation data by NASA and Universiy of Maryland is another example of such observational platforms for environment.

Overall environmental studies scientists and the biology community are starting to massively share knowledge as in large collaboration projects in physics (see CERN). Scientific reviews, for examples, are publishing journals to share experimental protocols (see Nature's Protocols) for an open access use, and more and more contributions are accepted at the condition that data bases be opened and shared. This is especially true of the biodiversity community that is starting the massively share gene databanks as well as more complex biodiversity information through the development of common ontologies for example (e.g. http://www.tdwg.org, or the Germplasm Resources Information Network (GRIN), or all developments behind the GBIF and SINGER groups).

Biologists share different opinions in the necessity or even the impossibility to define model systems to study by whole communities in order to better observe and quantify regularities in biological systems and define standard for experiments. Skepticism for this kind of approach in many cases is reducing the possibility to compare data of separate experimental teams. In this perspective data management can be very helpful in order to constitute efficient databases to recover information on performed experiments and data acquisition.

A lot can be done in this perspective, especially in organizing knowledge in large database with standards and interoperability with dedicated software. In biology this would open to combined systemic approaches ("omic" sciences: transcriptomic, lipidomic, proteomic, metabolomic, …), online analysis and data availability for in-silico modeling of systems and experiments. In environmental sciences this would strengthen and stimulate the necessary interdisciplinary approaches, teams from various disciplines sharing common data on environmental systems.
Therefore, there is a need for a strong effort to develop cultural and technical skills among the community to create, develop and share observation techniques and databases. This must be done at several levels:

- In the field, where the data "producers" are, by helping them very directly to the formatting and storing of data from the cropping, through the guidance of engineers skilled in both the scientific field of experiment/observation, sensor development and database construction
- The development of satellite Earth Observation techniques and the dissemination of resulting data among the scientific community
- The developments of local database storage systems insuring a first level sustainability.
- The presence on site of interdisciplinary groups dealing with second level ontology, data sharing developments and data mining and able to transmit this theoretical knowledge to the field level scientists.

- The creation of common facilities for the storage of massive databases, with shared engineering human capacities working for the whole community to constitute an Environmental and Biological Data Center (EBDC), opened to access for a worldwide community. Access to this center, that will by itself be a measure of the success of this initiative in Montpellier, could be measured as of now is the number of citations for example, and could even lead to an economic model based on the acceptance from the distant user of the datacenter to share some online storage capacity on his own remote machine, leading to secondary backup capacities which would not cost to EBDC and would create distant mirroring nodes of easier access for a worldwide access.

6. ModSysC Use Cases

Key-words: use cases, life science and healthcare, energy efficiency and environment protection.

6.1. Context and motivations

This section illustrates the impact of modeling and simulation (as promoted by our working group) through use case examples. It also illustrates the expected value for users, and a high level view of the economic and social context, as well as potential business impact evaluation. Today, modeling and simulation are extending their scope of usage. They used to be reserved to research and development domains, for certain industries. We see the number of interested industries increasing year after year, as well as the breadth of usage within specific industries, moving from a pure R&D usage, towards more and more operational applications. For example, the Aerospace industry which has been using modeling and simulation for many years, is now working on the convergence of their different modeling systems, for example combining mechanical, electronic, aerodynamics, thermodynamics and navigation systems, within a more integrated and operational global "modeling and simulation system".

This convergence will bring more value and will have a strong impact on the global development lifecycle of the systems produced, reducing cost and accelerating the time required to deliver systems (bring new aircrafts on the market for Aerospace example). It will drive new requirements towards the modeling and simulation systems, including increase of data volumes and data integration needs, as well as a broad variety of associated treatments; these treatments will result in the combination of applications to manage the data lifecycle (capturing, extracting, updating, storing, archiving, securing data, … often in real time) and applications exploiting these data (aggregating, correlating, modeling, analyzing, visualizing, simulating …) at the different stages of the lifecycle.

In the domains focused on by ModSysC2020 (life sciences and environment), we will see a similar type of scope extension, as well as a broader coverage of the Modeling and Simulation lifecycle.

Several examples of modeling and simulation usage have been described in the other chapters of this document. In the rest of this section, we provide examples which extend the actual usage, towards a more translational approach, enabling more synergy and coherence between the different phases of research, and taking into account the translation of research results, towards applications and large scale operational deployment.

6.2. Modeling and simulation use cases in Montpellier

6.2.1. Life science and healthcare

Several new organizations are being developed in Montpellier, and will gather a very large number of potential users for our "Modeling and Simulation" Institute.

- IHU (Institut Hospitalo Universitaire) is focused on chronic diseases. Its objective is to develop a translational approach, aimed at accelerating the transfer of R&D results towards clinical research and new therapy development. Modeling and simulation is well represented in the existing activities, but the IHU development will require to bring coherence and synergy between all the different users, the methods and tools they are using.
- CR2I (Centre Régional d'Innovation Industrielle) will host projects for IHU, which include Information Technology (IT), Diagnosis and Therapy. CR2I teams and customers will intensively use modeling and simulation in most of their projects. The Telemedicine project on modeling and workflows usage, as developed by IHU and CR2I teams, is in support of using telemedicine for chronic disease diagnosis and treatment monitoring.

IHU Project example: modeling of predictive factors and diagnostic

Our example is an IHU project on "Modeling of chronic disease states: exploitation of data collection oriented toward P4 (predictive, preventive, personalized, and participative) medicine". Today, medical research is highly influenced by new types of approach, methods, tools and associated data, globally called "Omics" domains and disciplines (genomic, proteomic, …). In addition to these data and discipline, and in particular for chronic disease, complementary factors and associated data, need to be integrated within the research scope (nutritional, environmental, social, psychological …).

Large cohorts (groups) of patients exist today, or are being formed, covering a very large range of pathologies. These cohorts provide very large volumes of data, having very different characteristics. Today, these data are mainly used to better understand, diagnose and cure a specific pathology. Though, chronic diseases are often combined and one disease often generates new ones (such as diabetes, obesity, cancer, acquired immunodeficiency syndrome (AIDS) …).

IHU scientists will work on demonstrating that different chronic diseases can be explained by similar mechanisms, from appearance to evolution and therapy (such as inflammation, for example). Both large cohorts and "omics" disciplines will play a fundamental role in the future of chronic diseases diagnostic and treatments. Relevant cohorts are part of IHU objectives (inflammation based chronic diseases, heart and neurological) as well as specific teams, focused on genomic, proteomic, medical imaging and psycho-social will combine their expertise and their data models.

The challenge is to find new ways of analyzing and correlating the vast amount of heterogeneous data, and develop new methods, tools and expertise to fully leverage these data, to better predict and diagnose chronic diseases. The real challenge is translating this base of knowledge into meaningful diagnostics and innovative new therapies and treatments. By modeling the clinical data, it could be possible to find new relationships that could have direct applications for conduction of medical and biomedical research, making prognoses and developing diagnostic tests. For example, modeling predictive factors can be applied in many chronic disease cases, such as diabetes or cardiology. Cardiac surgeons could use predictive analytics tools to better understand if a patient is an appropriate candidate for surgery, and how to best manage the case. By analyzing patient factors such as age, weight, current state of health, previous surgeries, his record of cardiac information's and number of procedures required, the surgeon is able to better understand the various risk factors on an individual patient basis.

The IHU modeling project proposed aims at defining and developing these new methods and tools, focusing on a modeling and simulation approach, applied to the full life cycle of chronic diseases, from detection of critical factors, to prevention when possible, or early diagnostic. Similar modeling methods and tools will be used, thus enabling a continuum of information exchange, from prognostic and diagnostics, to patient centric therapy and monitoring.

The different phases will be coherently managed. They will complement and enrich each other in term of knowledge and experience, thus contributing to the implementation of an "evidence based medicine" approach. The purpose of this new modeling and simulation approach will be to support research and clinician teams, by giving them access to a very broad range of data, modeling methods and tools, to easily access and exploit these data, correlate new types of data, and help them format and visualize the results of their investigations and simulations.

The results of these R&D activities can also be transferred to clinical research teams and to the CR2I. These teams will turn the Research results into new operational systems, which can then be moved from clinical trials to full deployment and usage by the medical eco-system (using telemedicine systems for example).

Expected value for life science and healthcare

The value and benefit of using modeling and simulation theory and tools, in the context of life science and medical research, will be as follows:

- Leverage available heterogeneous data and methods for new types of research and treatments, which require new correlations and combinations of data
- Develop innovative research directions across the various chronic diseases which are part of IHU scope of actions, and potentially support the discovery of common mechanisms between the various chronic diseases
- Accelerate the translation between research, clinical trials and deployment within healthcare eco-system.

This is crucial to fully support the objective of French government initiatives, "Investment for the future" , which not only aims at improving our research results, but also wants to more efficiently use research results to create value, and have an economic impact for our country.

6.2.2. Energy efficiency and environment protection

Project example : RIDER ("Réseau et Inter connectivité Des Energies classiques et Renouvelables")

As renewable energy becomes more and more present, at every level of our society (from houses, villages, cities, regions and countries), private and business environments will be in the middle of thousands of energy sources and energy consuming devices. RIDER's vision is that the global management and optimization of such a new energy environment will require Information Technology to be at the heart of it and form a «renewable energy IT platform». This IT platform will be able to structure the communication, the flow of real time data and the optimization between energy providers and energy consumers, combining centralized and distributed governance models, within a coherent architecture. The purpose of the RIDER project is to develop new modeling methods and tools, associated software & hardware, to treat energy related data in a very dynamic and innovative manner. Its objective is to optimize energy consumption at a scale and granularity which is not addressed today.

The scope of RIDER will include modeling and simulation, business intelligence and data analysis, optimization algorithms, visualization, technical architecture, communication standards development and tests. In term of modeling, RIDER's context requires to be able to link several domains, such as buildings physical and thermal characteristics, IT and Energy infrastructures, Energy flows and associated characteristics (temperature, air flow pressure, hygrometry, …), weather conditions, environmental characteristics, associated data models, etc.

A scientific model is based upon empirical data that is acquired from various sensors, measures and observations. These data are structured thanks to a data model to ease storage, access, and ensure integrity. Sensor measures are gathered by buildings supervisors and sent to the RIDER platform, where it is structured according to a data model. Statistical analysis can then be made on data in order to identify, for instance, correlations between different sensor measures or observations (variables). Thereby a statistical model can be built with mathematical equations to describe how the observed building behave. Data mining algorithms can also be used to find recurrent patterns of data, and association rules between variables of the model. Similarly, physical modeling is based upon mathematical equations describing a building behavior according to known thermal and physical properties. Both mathematical models resulting from statistics and physics can be seen as transfer functions representing the relation between input and output variables.

Operational research can leverage the aforementioned mathematical models in order to take optimal decisions accordingly to input variables and equations. Applying these models to different groups of buildings is a time consuming task because the mathematical models are bound to input and output data, specific to one context. We have the same concern with mathematical models for buildings, but we want them to be easily applicable to other buildings. Furthermore, the mathematical models use different variables that designate very different concepts (material conductivity, historical data, data patterns, occupancy schedules, sensors measures, actuators commands, optimal set points to reach, etc.). Hence, a simple data model is not enough to share a common understanding of the domains among people. Each modeling field stay linked to the data it uses. We need to separate the domain knowledge from the operational knowledge to give freedom of expressiveness to the mathematical models and ease their interconnection. This is why a formal description of the manipulated concepts must be achieved and gathered in a knowledge base, currently called pivot model.

A knowledge base is primarily used to enable making assumptions and inferring new information. We extend it by integrating processes and scenarios. Mathematical models can also be integrated in a knowledge base to help creating new information, which in turn enriches the knowledge base as a feedback loop, but also enable making high level scenarios. This knowledge base will help domain experts and users to easily understand the analysis, relationships, and decisions and to approve the results. 3D models will be able to represent not only data measures, but information about energy efficiency and help the user to understand how the buildings can be optimized.

Expected value for renewable energy efficiency and environment protection

The value and benefit of using modeling and simulation theory and tools, in the context of Renewable Energy usage optimization, is illustrated here after.

- Leverage available heterogeneous data and models for new types of research and treatments, which require new correlations and combinations of data and models. This will enable RIDER to deliver a holistic view of the managed domains and the optimization capabilities, within a totally new scale (from a couple of buildings, to hundreds of them, or even a full city).
- Develop innovative research directions across the various related domains (buildings, energy source and consumptions devices, weather forecast, energy usage profiling, …) which will enable to reach unmatched levels of renewable energy integration and energy efficiency optimization.
- Acceleration of translation between research phases, development and testing of new systems, and deployment of these systems for real usage. Modeling and simulation will the key solution for RIDER's "multi-scale" characteristics.

This is again crucial to fully support the objective of French government "Investment for the future" , which not only aims at improving our research and associated results, but also wants to more efficiently use research results to create value, and have an economic impact for our country. RIDER has already created 20 new jobs, and has the ambition to create a total of 80 new positions by 2013.

6.2.3. Environmental dynamics

Several organizations in Montpellier gather teams working in the domain of environmental sciences, that are potential users for our "Modeling and Simulation" Institute. These teams are strongly involved in LabEx and EquipEx.

- SFR Montpellier-Environnement-Biodiversité gathers 18 research teams from the Biodiversity domain. Its objective is to develop interdisciplinary and inter-team collaborations (within SFR and with external teams and research groups), particularly through research platforms, to promote the teams projects and develop their downstream application. The SFR Scientific policy is fully consistent with the national strategy defined by the Fondation de Recherche en Biodiversité (FRB, www.fondationbiodiversite.fr). It is organized in four main axes : model, explore, experiment to understand and manage anthropo-ecosystems.
- OSU OREME (Observatoire des Sciences de l'Univers : Observatoire de REcherche Méditerranéen de l'Environnement) gathers 6 laboratories in the domain of Earth Science, Ecology and Biodiversity, with agents from University, INSU, InEE and IRD. Its main mission deals with "Generic observation of mediterranean ecosystems and resources", with a specific focus on hazards and vulnerability. OSU OREME facilitates access to technical equipments mutualised between laboratories and ensures missions of communication and expertise towards public entities.
- GEOSUD (GEOinformation for Sustainable Development), based at Maison de la Télédétection and closely linked with teams from the organizations above, will evolve as an Institute dedicated to research on the multiple dimensions of spatial information : from data acquisition and data processing to spatial analysis and spatial modeling, information management and use of information by actors and stakeholders for environmental decision-making.

Project example: GEOSUD ("GEOinformation for SUstainable Development")

Understanding environmental dynamics requires multi-scale (both time and space) and multi-disciplinary approaches, from individuals to ecosystems, from local to regional and sometimes global, from daily behaviour and functioning to multiyear change and long term evolution. The spatial structure, organization and distribution of resources and ecosystems, their interface and interaction with anthropo-systems are key information to understand their dynamics.

GEOSUD will contribute to the Montpellier scientific community and the national scientific community research efforts by developing a national satellite imagery infrastructure freely accessible to all research teams and public actors, by developing and disseminating processing techniques and tools and spatial analysis and spatial modeling methods tuned to the various problems at stake : characterizing ecosystem structure, characterizing landscapes and land use changes, modeling natural and anthropogenic drivers of land dynamics, characterizing hydrosystems and parameterizing spatial models, analyzing impact of anthropogenic pressures on ecosystems and biodiversity.

This will allow scientific teams to develop innovative research in their respective fields of activity. using a large set of geoinformation and thematic data on their systems and using or developing relevant methods. While continuing to play their role in the development of space technologies, the partners will develop internationally innovative techniques for the management, processing and use of large volumes and fluxes of satellite data (data processing algorithms, multisource information coupling methods, data mining and assimilation methods, modelling approaches and tools, in situ data collection technologies).

Expected value for environment monitoring and management

Scientists working at the territorial level, from social sciences to environmental sciences and agriculture are at the moment underestimating, and the potential input and benefits of spatial information for their research subjects, and therefore under-using it. They lack the knowledge of its possibilities, access to data and skill to use spatial information properly. On the other hand they have no way to claim for new development fitted to their needs. GEOSUD will bridge the gap between ICT community and territorial and ecological science fostering new developments in both scientific fields.

Although targeted to science and technology GEOSUD will help policy makers to make a better use of spatial information. Many territorial and environmental scientists work in close relationship with land-use planners, agricultural development or environment protection agencies. Allowing them to make the best use of spatial information will pull all those professionals in a virtuous circle. The economical benefits from this are twofold: i) development of new service providers, especially SMEs able to support policy makers by the right spatial informations for their land-use management projects; ii) development of green technology markets. Those markets, claimed to be one major source of economical development in the future, need strong political incentives to allow new technologies to reach maturity. Such an incentive will be greatly enhanced by the ability, offered by a proper use of spatial information, to understand and plan the functioning of territories at different scales.

6.3. Concluding Remarks

Based on the examples developed in this section, we can see that modeling and simulation will be a fundamental domain of expertise for the coming years, and its usage will highly grow. It will not only be fundamental to accelerate research within specific domains (environment, biology, "omics" domains, …), improve their results, but it will also be fundamental for new "cross-domain" research, requiring the combination of data and modeling methods, in a coherent and organized manner, to achieve the new ambitious research projects' objectives (such as "evidence-based medicine" systems, distributed renewable energy management and optimization systems -RIDER-, and modeling of environmental dynamics for knowledge development and support to public policies).

Finally, modeling and simulation will be crucial for the added value of research projects, where the scientific ambition will be associated to business and economical impact objectives, including cost savings, revenue growth and employment generation, based on research results, in a much shorter cycle than it used to be.

7. Training and Life on Campus

Training of scientists across disciplines is a major challenge, in particular, because each discipline gets more and more complex and tends to specialize and generate sub-disciplines by partitioning complex disciplines. Below, we propose specific actions for training interdisciplinary students.

7.1 Moving towards English at UM2

A major challenge for universities is that science and higher education are getting more international, with English as the international language, and we must train many foreign students. A good example of international master proposed at UM2 is the Webscience (or e-science) master that is right on target. In particular, there are several courses on data management, one of them on distributed data management of interest to the scientific community. A concrete action would be to foster the proposal of Webscience at the highest level of UM2, as its international nature (and English teaching) seems to generate opposition. Other actions would consist in proposing specific updating courses in English (and French courses for those who are willing to learn the language) to students, and making official paperwork available in English. Overall, promoting the use of English at UM2 would help attracting foreign students, which is obviously a major way to disseminate French language and culture worldwide.

7.2 Interdisciplinary Training

Our group insists on the need to train students in the fundamentals of modeling in order to give them a solid background in Information and Communication sciences (or Sciences et Techniques de l'*Information de la Communication* - STIC) as well as social sciences. In particular, constant efforts should be made on the mastering of programming as well as databases. In that spirit, we note the existence of two Masters at UM2 (STIC-Health and STIC-Environment : the main goal of these Masters is to train students in two disciplines, one from STIC and the other from applied sciences, leading to bi-disciplinary profiles). This is achieved by pooling existing modules in both disciplines. However, beyond these ad-hoc solutions, our group insists on the necessity for UM2 to give itself the human and financial means to offer ambitious training in the modeling of complex systems via (1) the creation of new modules addressing specific needs in this domain and (2) hiring associate professors and professors at the interface of modeling and another discipline, hence via mixed committees (following the example of CNRS which opened 3 junior researcher positions between STIC and biology in 2010).

Considering the disciplines studying health and environmental systems, with social impact, quantitative and predictive outputs, and standards to define (as in engineering sciences), multidisciplinarity is a key requirement. Modern teaching concerning health and environmental sciences needs in general to be combined with the knowledge from mathematical, computer and physical sciences. Together with the examples presented above, several new Master Programs (see also below) aim at this kind of education, but much stronger efforts should be deployed in this direction.
For example, in order to produce multidisciplinary communities at the master level, it is necessary to easily activate at the beginning of the academic year special "proligomena" courses for allowing students to receive basic knowledge in all disciplines towards which multidisciplinarity is opened. This can be done also through the use of open teaching platforms, one to be cited as an "inspiring" example being OpenCourseWare (http://ocw.mit.edu) at MIT (Cambridge, USA) (see also next subsection).

7.3. Using Platforms and Databases in Teaching

The use of observational or experimental platforms is already partly embedded in the training of students, but should be reinforced. For instance, platforms like the computing grid of the Faculté des Sciences at UM2 or the spatial information platform at Maison de la Télédétection are also very useful to train students and should be encouraged. To insure progressive acquisition of higher technical knowledge, this should be started as soon as possible, especially through students' projects involving all levels of training from the first years up to the Master and PhD levels. This will create strong motivation, allow developing practical projects up to the operational stage and also promote project-oriented training that is difficult to transmit to students. In Montpellier, a student project is being established and tested via the platform of development of microsatellites dedicated to environmental studies (supported by the SOLARIUM group) gathering students of all levels and many scientific fields to a common project. This example should be extended as well as possible to most of high technology platforms.
Another significant step in training should be the creation, use and sharing of databases, up to its direct application in Master and PhD programs. Most students do not possess the background to develop databases that can be easily integrated and shared. This knowledge up to state-of-the-art qualification for database creation and use at the PhD level should be embedded in training. Note that such teaching units will be proposed within the framework of the Master in Water Sciences that will start in 2011.

7.4 Training Initiatives by Levels of Teaching

Training initiatives must take into account specific actions at three different levels of teaching: Licence, Master and PhD levels. In the worldwide competition for excellence among academic and research institutions, Montpellier can definitively play a major role thanks to the excellence of the disciplines in research. A first step for this action is the development of training at all levels (Licence, Master and Doctorate). In this context, interdisciplinary studies should be promoted among students with strong motivations in fundamental and applied research.

Licence Level. Concerning modeling and simulation approaches, efforts should be directed to foster the education of students in environmental and life sciences to mathematics, physics and computer sciences, thus towards environmental and life sciences quantitative approaches. Since disciplines like physics and mechanics have experience in modeling systems and processes, it is important to start from very specific examples in order to present the operational approach of modeling, explore strategies to advance in the understanding of phenomena, unveil mathematical and analytical computing power, proposing numerical approaches for numerical computing and simulations. This can be done with courses organized in lectures on the analytic approaches and complemented with practical lectures to study the same problem via numerical approaches. Teaching units of this kind already exists in Montpellier (see for example the 2nd year teaching unit "Modeling and Algorithmic in Physics") and could constitute an example to introduce quantitative approaches to student communities that have weaknesses in mathematics and physics education. Interdisciplinary modeling should then be associated also with reinforced education, with additional teaching units for motivated, top students to assess solid bases in mathematical, physical and numerical approaches; this reinforced education should be coupled with a training on environmental and life sciences concepts and questioning.

Master level. One of Montpellier' strengths is the dynamics expected to be derived from the strong articulation of research forces with higher education. Several master programs are training in interdisciplinary fields, for instance: (1) Molecular and Cellular Biophysics, BioMed Master in biology and health sciences (supported by the universities of Montpellier 1 and Montpellier 2 since 2006); (2) Physics and Engineering of Living Matter (created by UM2 in 2009); (3) Bioinformatics; (3) Statistics in life sciences and health and Biostatistics (supported by UM 1 and 2 and SupAgro); (4) Master in water sciences (supported by UM 1, 2 and 3, Agro Paristech, INRA, IRSTEA, IRD) that started in September 2011. Complementing these more specialized areas, interdisciplinary «STIC-Health» (development of a UM1, UM2, Institut Telecom, Ecole des Mines d'Alès partnership) and «STIC-Environment» Masters options have recently been setup, involving most of the disciplines contained in NUMEV.

In many cases, teaching modules between different masters (see for example the Biophysics masters) have been pooled, thus bringing students with different backgrounds to merge in an interdisciplinary context necessary for the scientists and engineers working in interdisciplinary fields of future generation. Nevertheless, different LabEx initiatives (NUMEV, EpiGenMed, …) have noted that important efforts need to be done. To develop the whole interdisciplinary teaching, it is necessary to set up specific introductory courses and improve the quality and amount of teaching. In particular, it is necessary to introduce more teaching modules that introduce different modeling approaches to complex systems.

Another aspect to stress is the absence of training or teaching modules dedicated to systems biology approaches, which require different expertise in mathematics, physics, computer sciences and biology. In view of a development of this discipline together with biological physics and biophysics approaches (see also research proposed actions), it is necessary to propose a master offer on systems biology of the University of Montpellier.

In the same perspective, discussions concerning a specific master in Modeling and Simulation Approaches to Complex Systems had been initiated, but we are still not in a condition to propose a formulation that could be exhaustive for covering efficiently all different kinds of modeling developed in Montpellier. The discussion should be continued.

PhD level. In order to create a visible profile on the site, it is necessary to setup a complementary graduate program focusing on the topics of its scientific project, in collaboration with the I2S, SIBAGHE and CBS2 doctoral schools. This program will leverage the completeness of graduate modules, and above all, the creation of recurring annual schools that take place over the course of a week and include appearances by renowned researchers. The targets of these schools will include doctoral students, postdocs, and young researchers from France and abroad who wish to benefit from an advanced platform on one of the fields of excellence. An international school of this type has already been setup with a bi-annual calendar in Montpellier, on the theme of surgical robotics, funded in part by Europe and by local administrations. The school will serve as a model for other schools: for example, a proposal for a «Physics of Living Matter» school has emerged with a focus on two topics, «from cells to tissue» and «intra-cellular assemblies». It has the potential to reach a large population of doctoral students whose research lies in the middle-ground between hard sciences and biomedical sciences: physicists, mathematicians, mechanical experts, electricians, chemists, biologists... All these initiatives (graduate programs and schools) should be characterized by specific modules and topics oriented to modeling and interdisciplinary approaches in the perspective of better merging disciplines that, because of very fast progress in the last decades, have followed in many cases a separated development.

7.5 Educating Students to Scientific Independence

All points presented above should be encouraged also in the perspective of stimulating as much as possible the scientific independence of university students. This process should start at the first year of academy, whereas nowadays' undergraduate students are still used to integrate their studies in a too much "scholar" (in a "high-school" format) form of education. As a matter of example, it is frequent to host foreign students in courses and realize that they demonstrate a higher independence in developing their personal project. Moreover, French PhD students tend to require a lot of supervision, while students in high-level universities abroad work to look for their own PhD topic during their Master years. In the same perspective, students abroad usually integrate the laboratory-life much earlier than in our universities, learning how to distribute their effort in research and studying, sharing their experience with young colleagues and learning various aspects of modern science and the related life. All these facts enormously motivate the scientific careers. We believe there is a long way to change an "ingrained" vision of the university, seen just as a natural continuum of high-school education, into the optimal place to receive the best scientific education.

Scientific independence is the way to improve human resources necessary for innovation and progress and young scientists are central factors in such a social process. The University, due to its multidisciplinary culture and foundation, is the best place for developing this kind of education.

7.6 Making Life on Campus More … Lively

Our group insists on the importance of providing a scientific culture to students. To do so, we acknowledge the existence of several initiatives to popularize scientific areas, but we think efforts of centralization are needed to make them easily available. To evaluate whether students are aware of the existing actions, we suggest that surveys should be conducted at UM2 using questionnaires. We also encourage extending the opening hours of the central library later in the evening and over vacations (including summer), possibly through the employment of students. In the same vein, thinking should be conducted on how to keep the university open over summer, possibly by organizing scientific events like summer schools. Overall, UM2 students should be given the opportunities to become part of the university life and true actors of its functioning, not just consumers.

8. Summary of Proposed Actions

For each of the four themes in ModSysC2020, we proposed specific interdisciplinary projects for which strong support from UM2 is needed to reach the objective of international leadership. We summarize these projects below.

Interdisciplinary research projects

In order to develop interdisciplinary synergies in modeling and simulations related to life and environments among the researchers scattered in Montpellier, we should capitalize on the experience already gained by existing and dynamically operating communities and collaborations. The research directions we described are part of the LabEx NUMEV and EpiGenMed. A more focused subset of this program is also part of the IBC proposal or of the EquipEx GEOSUD. In these four projects, data management is viewed as important in conjunction with modeling, simulation and computation.

The GRISBI community (Groupement de Recherche Interdisciplinaire sur les Systèmes Biologiques) gathers more than 70 researchers belonging to about 10 different laboratories in theoretical and experimental modeling approaches to biological systems and related environmental problems.
GRISBI should be supported and developed because of its functionality towards the large variety of institutions distributed in Montpellier area.

The IBC interdisciplinary project center aims at gathering about 50 researchers working on modeling, processing and analyzing data on a large scale in the fields of biology, health, agronomy and environment. IBC's flexible structure should serve as a test-case for further similar interdisciplinary centers, for example focused on the interplay between ecology, mathematics and computer science, or on the impact of climate changes on human populations.

To further foster interdisciplinary collaboration, we identified the need for an Environmental and Biological Data Center (EBDC), opened to access for a worldwide community, for common facilities for the storage of massive databases. This is a practical, operational project that should encourage the entire Montpellier community in modeling and computation to collaborate through the sharing of the most stable objects (data) they manipulate.

To strengthen the interdisciplinary use of geoinformation for the analysis of environmental dynamics the GEOSUD (GEOinformation for Sustainable Development) initiative, that involves various teams on Montpellier Campus, should also be supported. It is dedicated to research on the multiple dimensions of spatial information for environmental sciences: from data acquisition and data processing to spatial analysis and spatial modeling, information management and use of information by actors and stakeholders for environmental decision-making. These actions (NUMEV, EpiGenMed, GRISBI, IBC, GEOSUD and EBDC) should be strongly encouraged by UM2 and its partners, in particular, by dedicating significant amounts of human resources (PhD students, engineers and researchers).

These actions should lever activities in order to:

- **foster the sharing of competences and merging approaches** in modeling and simulation of complex systems related to biological and environmental sciences,
- **strengthening** the synergy between experimental and theoretical modeling approaches and simulation platforms,
- **giving** the research community in Montpellier international visibility;
- **promoting** "open-walls" collaborations (e.g. GRISBI as structuring group of reference), for research and education, in modeling complex systems in life and environments, at the national and international levels;
- **supporting local teams and working groups** in starting local interdisciplinary projects and establishing future external interdisciplinary collaborations based on approaches involving modeling and simulations and possibly integrated to other supporting initiatives (see next points);
- **interfacing** complementary initiatives presented hereafter like GRISBI, the Institute of Computational Biology (IBC), the Environmental and Biological Data Center (EBDC), the project GEOSUD;
- **raising the awareness of scientists** for analogue initiatives and progresses in data management, on algorithmic and computing in biology and environmental sciences, scientific data management, production, storage and archiving of data, that can improve their scientific activities and collaborations;
- **promoting collaborations efficiently**, among regional, national and international teams for answering to calls for projects by the ANR, the European Community and the Human Sciences Frontiers Program grants, …
- **creating a dynamical database** by collecting information on instruments, techniques and human resources (competences, tools and approaches in experimental techniques and theoretical and numerical methods) about each individual, team and laboratory of the collaboration for:
  - providing information for a dynamic search of competences and instruments between dedicated platforms and experimental teams for instrumental and protocols development;
  - helping theoretical, experimental or mixed working groups to form and focus on new problems, approaches and ideas.
- **sharing knowledge, experience and scientific benefits** that

participants can bring from their own external collaborations (see for example, the DYALOG project in Systems Biology);

- **developing**, via simulation, modeling and experiments, oriented technologies for studying biosystems and related environmental problems together with hospitals, industrial partners and **territorial organizations**.
- **promoting** interdisciplinary research to the core of young researches that will develop the interdisciplinarity of the future: master students, PhD students and postdoctoral fellows
- **enhancing** young scientists interdisciplinary work (e.g. prizes, fellowships for international conferences and schools)
- **promoting** national and international workshops and schools,
- **contributing** to CNRS "Groupement de Recherches" national initiatives,
- **inviting** renowned scientific personalities for lectures and seminars.

Furthermore, we acknowledged the growing importance of HPC and cloud computing for these themes. In HPC, Montpellier enjoys a strong position of competency center and enabler, with CINES, HPC@LR and the computing grid of the Faculté des Sciences at UM2. Furthermore, IBM Montpellier has been promoting and selling HPC or cloud solutions, which is useful to help implementing large-scale applications. Although the relationship between applications, users and HPC providers is well on track with HPC@LR, what is lacking is a strong local eco-system with a continuum between research, development and applications. First, research in HPC is relatively scattered, limited and specialized. We clearly need to attract one or more senior HPC researchers with international stature, to organize and lead the community. Second, we also need different industry actors including startups that provide HPC services (like the ASA company involved in HPC@LR) as well as other actors (e.g. Bull, Orange) that could be interested in joint R&D projects.

Training Actions

We acknowledged that training of scientists across complex disciplines is a major challenge. We discussed the importance of revisiting students training in fundamental domains such as modeling, computer programming and database which are typically taught too late, in specialized masters. We proposed the following specific actions for training interdisciplinary students: (1) moving towards teaching in English at UM2 in order to compete with top universities in attracting the very best foreign students; (2) interdisciplinary training, with the fundamentals of modeling as well as Information and Communication sciences (STIC), in particular, programming and databases, and social sciences; (3) using platforms and databases in teaching; (4) training initiatives by levels of teaching (Licence, Master, PhD); (5) educating students to scientific independence; (6) making life on campus more … lively: UM2 students should be given the opportunities to become part of the university life and true actors of its functioning, not just consumers.

Social and economic impact

We expect these actions to have an important social and economic impact at the regional, national and international levels. At the regional level, it will foster collaboration between mathematicians and computer scientists involved in HPC, data management and data analysis with scientists of different disciplines such as agronomy and environmental science, life science, etc., which are strong in Montpellier. Although collaboration between these communities already exists in different forms, the interdisciplinary actions we propose will strengthen collaboration by giving it more impetus and visibility, which will in turn generate interest from other participants, including many local start-ups which have to deal with scientific data, and public authorities that have to define public policies in the domains of environment and health. At the national and international levels, the impact will be a natural continuation of the local actions and a consequence of the global nature of complex systems which is boarder-free. Since all scientists participating in the project have on-going collaboration with peer teams at the national and international level, these projects will simply strengthen those collaborations.

We also described use case examples which extend the actual usage, towards a more translational approach, enabling more synergy and coherence between the different phases of research, and taking into account the translation of research results, towards applications and large scale operational deployment. Based on these use case, we can see that modeling and simulation will be a fundamental domain of expertise for the coming years, and its usage will keep growing. It will not only be fundamental to accelerate research within specific domains (environment, biology, "omics" domains,…), improve their results, but it will also be fundamental for new "cross-domain" research, requiring the combination of data and modeling methods, in a coherent and organized manner, to achieve the new ambitious research projects' objectives (such as for "evidence-based medicine" systems, distributed renewable energy management and optimization systems -RIDER-, and modeling of environmental dynamics for knowledge development and support to public policies).

Finally, modeling and simulation will be crucial for the added value of research projects, where the scientific ambition will be associated to business and economical impact objectives, including cost savings, revenue growth and employment generation, based on research results, in a much shorter cycle than it used to be.

References

[Ailamaki 2010] Ailamaki A., Kantere V., Dash D. 2010. Managing scientific data. Communications of the ACM, 53(6):68-78.

[Berry 2009] Berry V., Guillemot S., Nicolas F., Paul C. 2009. Linear time 3-approximation for MAST problem. ACM Transactions on Algorithms 5(2):1-18.

[Bessiere 2005] Bessiere C., Régin J.C., Yap R.H.C., Zhang Y. 2005. An optimal coarse-grained arc consistency algorithm, Artificial Intelligence, 165(2):165-185.

[Guindon 2003] Guindon S., Gascuel O. 2003. A simple, fast and accurate algorithm to infer large phylogenies by maximum-likelihood. Systematic Biology 52(5):696-704.

[Heggerness 2007] Heggerness P., Paul C., Telle J. A., Villanger Y. 2007. Interval Completion with Few Edges. STOC'07: 39th ACM Symposium on Theory of Computing, ACM New York (USA), 374-381.

[Merali 2010] Merali Z. 2010. Computational science: Error…why scientific programming does not compute. Nature 467:775-777.

[Özsu 2011] Özsu T., Valduriez P. 2011. Principles of Distributed Database Systems. 3rd edition, Springer.

Annex 1: Short Biographies of ModSysC2020 members

Nicolas Arnaud is a senior researcher at CNRS - Montpellier, and head of the Observatory for research on Mediterranean environement (OREME). He received his PhD in geochemistry and geophysics in 1992 and his HDR in 1992 both in University Blaise Pascal in Clermont-Ferrand. His is author of more than 70 international papers on the building of mountain ranges and plate tectonics implications for the deformation of continents, led with international teams and PhD students. He has been regularly involved in the leadership of national programs and has occupied scientific positions in national prospective groups and review panels at CNRS or ANR. He occupied the head of the research Federation ISTEEM in Montpellier before creating and presently heading the Observatory. This latest leadership position has led him to develop programs dedicated to a better preservation, exchange and use of extensive observation data of natural media with special attention to those of use for public policy makers.

François Briant is the WW STG Customer Centers and Center for Advanced Studies France (CAS) Chief Architect, supporting large and complex Enterprise Applications Solutions architecture design, test and delivery, as well as collaborative Research projects with French Universities. He is an IBM Distinguished Engineer since April 2008. His key competence includes IBM Server, Enterprise Applications Solutions (EAS) Infrastructure, SOA, Information Management Architectures and Smarter Cities architectures. François is also providing WW STG Customer Centers teams with technical and leadership development directions. François has a diploma of Engineer, from the Ecole Nationale Supérieure de Physique (Centrale Marseille). He is an elected member of IBM Academy of Technology and is teaching Enterprise Systems Architecture in Montpellier University.

Olivier Gascuel (http://www.lirmm.fr/gascuel) is a senior researcher (DR1) at CNRS and head of the Methods and Algorithms in Bioinformatics team at LIRMM; 115 publications in international journals, conferences and books; more than 6000 citations in international peer-reviewed journals; his publication (Guindon and Gascuel 2003) has been the Science Watch Current Classic in ecology and environment since October 2007 (i.e. most cited paper in the field); associate editor and editorial board member of five international journals on bioinformatics and evolution; CNRS Silver Medal in 2009.

Olivier Gimenez (http://www.cefe.cnrs.fr/biom/Permanents/O_Gimenez.htm) is a researcher at CNRS, Montpellier. He received an education in mathematics and statistics, his PhD in biostatistics and his HDR in in evolution and ecology. He is especially involved in the modelling of population dynamics, the estimation of demographic parameters and management of large mammals using statistical modelling. He has published 66 papers in peer-reviewed statistical and biological journals, 6 book chapters and a book "Bayesian Analysis for Population Ecology" at CRC Press. His work is at the interface of mathematics, statistics, ecology and evolution. He is an associate editor for journals in biology and statistics, including Biometrics, Wildlife Biology, Open Statistics and Probability Journal. He is in charge with others of an Interdisciplinary Master "Science et Technologies de l'Information et de la Communication pour l'Ecologie, l'Environnement et le Développement Durable" (http://www.master-stic-environnement.com) and leader of a CNRS "Réseau Thématique Pluridisciplinaire" Mathematics and Biodiversity.

Christophe Godin (http://www-sop.inria.fr/virtualplants/wiki/doku.php?id=team:godin) is a senior researcher at INRIA, Montpellier. He is an engineer from the School of Civil Engineering in Paris (1986), and received his PhD degree in complex systems control from the Technological University of Compiègne, France (1990), and the Habilitation degree in computer science from the University of Montpellier, France (2003). He did a postdoctoral fellowship in robotics from the Marcoussis Laboratories of Alcatel-Alsthom Research (1989-1991) and was a project leader at Cap Gemini Innovation, Paris, in language processing (1991-1992) and at AMAP, CIRAD-Montpellier, (1993-2003) in plant growth modeling and analysis. In 2003, he joined the French National Institute for Research in Computer Science and Control (INRIA) as a research director. His research interests include modeling and simulation in developmental biology, modeling of plant development, graph theory, stochastic processes, multiscale modeling, 3D modeling of organs and branching systems, dynamic systems with dynamic structures, fractals, models of tissue development, regulatory networks, L-systems, and the design of computer platforms for simulation in biology. He authored and co-authored more than 70 scientific papers and book chapters. He has been associate editor of several journals, including Frontiers in Technological Advance in Plant Sciences. He is was a co-founder of the series of international conferences FSPM on plant modeling and is a permanent member of the steering board. He is the coordinator of the open software platform for plant modeling OpenAlea. He has been an invited speaker to international workshops and conferences (6-8/year on average). He is also a referee for French regional, ministerial, and English, European funding agencies. He is currently involved in several national and international projects

Hervé Jourde is an assistant professor in Hydrological and Environmental Sciences at University Montpellier 2, in the HSM Laboratory where he coordinates the team "Karst and Heterogeneous Media - Hydrogeology, Hydraulics and Transfers". He joined this laboratory, first as a lecturer in 2001, and obtained his HDR in 2007. Previously, he was a research associate (1999-2000) at Stanford University, California, where he specialized in fluid flow modeling and scale transfer modeling in the Dept of Geological and Environmental Sciences. He received his Ph. D. degree in Hydrological Sciences from University of Montpellier 2 in 1998. His research activities focus on characterization and modeling of underground transfers, with particular attention to groundwater/surface water interactions and to water resource evolution in response to climatic and anthropogenic forcing. This implies field observation, which resulted in the development of various experimental sites (www.medycyss.org). He has served as SC member of various international conferences on Karst Hydrology and occupied scientific positions in review panels at ANR. He co-authored book chapters and published more than 30 technical papers. He works in close partnerships with private companies and Universities in Europe, USA and South America.

Pascal Kosuth is a senior researcher at IRSTEA, Montpellier (www.tetis.teledetection.fr). He received an education in mathematics and physics and in Water resources and environmental sciences (Ecole Polytechnique; Ecole Nationale du Génie Rural, des Eaux et des Forêts). His PhD is in automatics and architecture of complex systems (INSA-CNRS). His research records deal with water resource, aquatic ecosystems, irrigation, water management, large tropical rivers dynamics, ecosystems and land dynamics. He has worked for more than 7 years in tropical countries (Haïti, Brazil, Senegal). Since 2005 he has been specially involved in the development of satellite Earth Observation techniques for the monitoring, understanding and management of agriculture, environment and land dynamics. Director of the TETIS laboratory (AgroParisTech, IRSTEA, Cirad) (2005-2010) he is the Director of Maison de la Télédétection since 2006. Member of the European Space Agency Earth Science Advisory Committee (2003-2007) and of the French Space Agency Scientific Committee (2009-now) he is the coordinator of the GEOSUD initiative and EquipEx (www.geosud.teledetection.fr).

Pascal Neveu (http://www.montpellier.inra.fr/umr_asb/umr.php?page=neveu) is a researcher at INRA. He is the leader of the GAMMA team (Gestion Analyse et Modèles pour les Masses de données en Agronomie) of the MISTEA research unit. He has been involved for many years in the conception of scientific Information Systems. He has worked with a lot of teams of food science, plant study and environmental fields. He is also involved in the set up of ontologies and in the methodological development of distributed system. He is leader of a work package dedicated to knowledge representation and data management in the EU project CAFÉ (Computer Aided Food processes for control Engineering). He has participated in several international projects such like Stic AmSud MIFIMA: Mathematics, Informatics and fisheries management, 2007-2009 or Telemac EC Project Telemonitoring and advanced telecontrol of high yield wastewater treatment plants.

Andrea Parmeggiani got his PhD in Physics (Paris VII) in 2000 working at the Curie Institute (Paris). He was post-doctoral fellow in Theoretical Physics at the Technische Universität Munich (Germany), Harvard University (USA) and Hahn-Meitner Institute, Berlin (Germany). In the end of 2003 he became Assistant Professor in Theoretical Biological Physics (HDR in 2008) at the University of Montpellier 2, Biology Department. Since 2010 he is Principal Investigator at the "Biological Physics and Systems Biology" modeling team in the Biology Lab DIMNP, UMR5235 CNRS/UM2/UM1. He shares his interests between teaching and research activities, promoting interdisciplinary approaches to all student levels (L, M, D) via disciplines like stochastic processes, nonlinear systems, statistical mechanics and modeling of biological systems at molecular and cellular scales. He is responsible of the Master specialty "Molecular and Cellular Biophysics" (Master BioMed Recherche, Biology Department) and of the Master "Physics and Engineering of Living Matter" (Physics Department). Member of the University Council for Student Life and Studies (CEVU), he was invited professor in Italy, Slovenia, Germany and Japan to teach nanobiophysics, motor proteins stochastic modeling and collective phenomena in intracellular processes. His research interests in theoretical physics and biological physics concern the study of the (multiscale) physical principles of intracellular organization in eukaryotic and prokaryotic cells via statistical mechanics and stochastic processes methods.

Esther Pacitti (http://www.lirmm.fr/~pacitti/) is a professor of CS at University Montpellier 2 and the head of the Zenith team for UM2 at LIRMM. Previously, she has been an assistant professor at University of Nantes (2002-2009), a member of the Atlas INRIA team (2003-2010), an assistant professor at University Paris 5 (2000-2002), and a researcher at Nucleo de Computacao Eletronica, UFRJ, Rio de Janeiro (1985-2000). She obtained her PhD degree in CS in 1998 from PUC-Rio, Rio de Janeiro, and her HDR from University of Nantes in 2008 on the topic of data replication in different contexts (data warehouse, cluster and P2P). Her research now focuses on distributed data management in large-scale systems (P2P, cluster, grid, cloud), in particular, P2P data sharing and data-centric scientific workflow management. She has published more than 70 technical papers. She has been serving as PC member of major international conferences including SIGMOD, VLDB, ICDE, ICDCS, CIKM, EDBT, EuroPar, etc. She was the general chair of HPDgrid08, the PC vice-chair of the Parallel and Distributed Database track at Euro-Par08, and the PC chair of DAMAP08 and HPDgrid06.

Patrick Valduriez (http://www-sop.inria.fr/members/Patrick.Valduriez) is a senior researcher (DR0) at INRIA, and the head of the Zenith team for INRIA at LIRMM. He has also been a professor of CS at University Paris 6 (1999-2002) and a researcher at Microelectronics and Computer Technology Corp. in Austin, Texas (1985-1989). He received his Ph. D. degree and Doctorat d'Etat in CS from University Paris 6 in 1981 and 1985, respectively. His research focuses on data management in large-scale distributed and parallel systems (P2P, cluster, grid, cloud), in particular, scientific data management. He has authored and co-authored over 200 technical papers and several textbooks, among which "Principles of Distributed Database Systems". He has been a member of the ACM SIGMOD board, a trustee of the VLDB endowment and an associate editor of several journals, including ACM TODS, the VLDB Journal, Distributed and Parallel Databases, and Internet and Databases. He has served as PC chair of major conferences such as PDIS93, SIGMOD97, VLDB98 (industrial chair), VLDB99 (European chair). He was the general chair of SIGMOD04, EDBT08 and VLDB09. He was the recipient of the 1993 IBM scientific prize in Computer Science in France. He obtained the best paper award at VLDB00.