

Towards 2020 Science: A Draft Roadmap

<p style="writing-mode: vertical-rl; transform: rotate(180deg);">Goals</p>	<p>Automated remote species-identification</p> <p>Reliable global warming, natural disaster and weather-prediction models</p> <p>Keystone-species identification</p> <p>Systems approach to biology becomes mainstream</p> <p>Comprehensive biological data integration: majority of 'omes co-analysable</p> <p>Individualised medicine</p> <p>Modelling based vaccines</p> <p>Ex vivo molecular-computer diagnosis</p> <p>Systems biology emerges as new discipline</p> <p>LHC preparation</p> <p>LHC switch-on</p> <p>Finding the Higgs boson?</p>	<p>Predictive models of: effects of rainforest destruction, forest sustainability, effects of climate change on ecosystems, effects of climate change on foodwebs, restoration ecology planning, global health trends, sustainable agriculture solutions</p> <p>Full model of a single cell</p> <p>Full model of a multi-cellular tissue/organ/organism</p> <p>First purely <i>in silico</i> developed drug comes to market</p> <p>Personalised <i>in situ</i> molecular-computer 'smart drug'</p> <p>Verifiable global ecosystem models</p> <p>Verifiable global epidemic models</p> <p>Understanding complex biological systems, from cells and organisms to ecosystems</p> <p>Comprehensive codification of biological knowledge</p> <p>Predictive models (mathematically precise, definitional, verifiable) of biological systems (cells, organs, ecosystems)</p> <p>Understanding the make-up of the Universe</p>	<p>Foundational theory of Global Ecosystems</p> <p>Understanding of Earth's life support systems</p> <p>Predictive model of effects of human activities on Earth's life support systems</p>
<p style="writing-mode: vertical-rl; transform: rotate(180deg);">Scientific Challenges</p>	<p>Integration of, and access to, oceanographic data, climate data, remote sensing data, taxonomic data, demographic data, disease surveillance data, species data, metagenomics data</p> <p>New representations of biological systems as: Logical schemas (ontologies) Programs (FORET) Equations (e.g. Lotka-Volterra)</p> <p>Population dynamic-models emerging: non-local interactions, spatially explicit, non-linear, stochastic</p> <p>Statistical and correlative models and analysis. Able to deal with sparse / noisy data; niche modelling (e.g. GLIM, GAM, GARP); pattern analysis classification; cluster analysis; novel visualisation; parameter estimation</p> <p>Understanding / codification of metabolic pathways and gene regulation network</p> <p>Initial metagenomics theory and data required for theoretical underpinning of biodiversity</p> <p>Compositional virtual cell</p> <p>Autonomous observation and analysis of complex ecological data</p> <p>Computational model of carcinogenesis</p> <p>'Postdictive' modelling possible for wide range of systems</p> <p>Main disease pathways and gene networks identified</p> <p>Cell cycle model</p> <p>Cell differentiation model</p> <p>A computational theory of synthetic biology</p> <p>Molecular computers increasingly used to design alternative 'smart' medical therapies</p> <p>Biological systems abstraction, perturbation and falsification / verification</p> <p>Novel, biologically inspired, computing architectures and paradigms ('Building blocks of next century of computing')</p>	<p>Mathematical and empirical verification / proofs of computational models</p> <p>Applications of algebraic concurrency theory enabling significant breakthroughs in understanding and predicting behaviour of complex natural systems</p> <p>Autonomous experimentation - 'Artificial scientist'</p> <p>Efficient software for parallel linear and non-linear algebra applicable to full matrices of dimension - 10⁶-10⁷</p> <p>Widespread <i>de novo</i> creation of models, theories and solutions (e.g. protein-drug candidates) from data using advanced Machine Learning</p> <p>Computer science concepts now core discipline in natural sciences (as mathematics are today in physical sciences)</p>	<p>Abstraction, descriptive complexity and formal verification concepts from computer science enable both postdictive and predictive multi-scale modelling</p> <p>Abstract Machines / codification concepts and tool evolution now start to enable initial predictive modelling from molecular to population biology level (e.g. metagenomics, species-biodiversity, epidemiological level) including non-local interactions with climate and intervention data and models</p> <p>Compositional enables prediction of emergent behaviours of composite entities</p>
<p style="writing-mode: vertical-rl; transform: rotate(180deg);">Concepts/Tools from Computer Science</p>	<p>Integration of computing, computer science and mathematics, leads to integration of modelling, experimental and theoretical approaches in science</p> <p>Emergence of molecular machines</p> <p>Multi-scale modelling: micro-macro model integration; stochastic/deterministic model integration; hierarchical modelling architecture</p> <p>Scaling of Bayesian inference to very large data sets</p> <p>Active Learning techniques proliferating in science - towards autonomous experimentation (and understanding the brain)</p> <p>Integration of sensors (environmental, physiological, chemical) and machine learning and data management - towards autonomous experimentation</p> <p>Codification of Biology begins: Emergence of Abstract Machines for biology (e.g. using Process algebra [π-calculus - SPIM, PEPA])</p> <p>Emergence of Synthetic Biology (e.g. Biobricks)</p> <p>Mobile Ambients, Brane calculus</p> <p>Application of Kolmogorov complexity concepts start to enable understanding and measurement of biological complexity across scales</p> <p>Statecharts and visual formalisms being used to research chemical and biological challenges</p> <p>Inductive Logic Programming used to discover families of protein folds in proteomics</p> <p>Bayesian networks used to model effects of toxins in metabolic pathways</p>	<p>Information retrieval and search for novel data types</p> <p>Data and metadata availability / requirement from publishers, together with increasing online availability, starts to transform publishing model</p> <p>Large-scale sensor networks starting to proliferate</p> <p>Integration of data, metadata, open access and publishing continues to transform scientific publishing model</p> <p>'Active XML' for ubiquitous data</p> <p>Data 'behind' scientific papers available in machine readable formats</p> <p>Journal / database hybrids</p> <p>Knowledge base for molecular mechanisms and architectures</p> <p>Truly smart lab-books</p> <p>Managed platforms and 'Service-Oriented Architectures' being widely created and adopted in science and scientific computing, but also for the 'new kinds' of conceptual and technological tools in science</p> <p>Symbolic computation integrated into scientific databases and programming languages</p> <p>Advances in database technology starts to enable: Scalability, distributed optimisation, fault-tolerance, self-management and integrations of heterogeneity, metadata and schematisation.</p> <p>Concept of intelligent interaction starts to become a reality (combining data cubes, transformations in science publishing, visualisation, distributed data management). Easy to program, easy to use interfaces and smart lab-book devices/interfaces</p> <p>Ability to rapidly build complex distributed systems from shared components that are: trustworthy, predictable, reliable, scalable and extensible</p>	<p>Abstraction, descriptive complexity and formal verification concepts from computer science enable both postdictive and predictive multi-scale modelling</p> <p>Abstract Machines / codification concepts and tool evolution now start to enable initial predictive modelling from molecular to population biology level (e.g. metagenomics, species-biodiversity, epidemiological level) including non-local interactions with climate and intervention data and models</p> <p>Compositional enables prediction of emergent behaviours of composite entities</p> <p>Computer science concepts now core discipline in natural sciences (as mathematics are today in physical sciences)</p>
<p style="writing-mode: vertical-rl; transform: rotate(180deg);">Computational platforms</p>	<p>Geographic Information Systems (e.g. Google Earth, Virtual Earth, TerraServer) increasingly evolving and increasingly used in science (e.g. climatology, biodiversity informatics, epidemiology)</p> <p>Data and metadata availability / requirement from publishers, together with increasing online availability, starts to transform publishing model</p> <p>Large-scale sensor networks starting to proliferate</p> <p>Integration of data, metadata, open access and publishing continues to transform scientific publishing model</p> <p>'Active XML' for ubiquitous data</p> <p>Data 'behind' scientific papers available in machine readable formats</p> <p>Journal / database hybrids</p> <p>Knowledge base for molecular mechanisms and architectures</p> <p>Truly smart lab-books</p> <p>Managed platforms and 'Service-Oriented Architectures' being widely created and adopted in science and scientific computing, but also for the 'new kinds' of conceptual and technological tools in science</p> <p>Symbolic computation integrated into scientific databases and programming languages</p> <p>Advances in database technology starts to enable: Scalability, distributed optimisation, fault-tolerance, self-management and integrations of heterogeneity, metadata and schematisation.</p> <p>Concept of intelligent interaction starts to become a reality (combining data cubes, transformations in science publishing, visualisation, distributed data management). Easy to program, easy to use interfaces and smart lab-book devices/interfaces</p> <p>Ability to rapidly build complex distributed systems from shared components that are: trustworthy, predictable, reliable, scalable and extensible</p>	<p>Information retrieval and search for novel data types</p> <p>Data and metadata availability / requirement from publishers, together with increasing online availability, starts to transform publishing model</p> <p>Large-scale sensor networks starting to proliferate</p> <p>Integration of data, metadata, open access and publishing continues to transform scientific publishing model</p> <p>'Active XML' for ubiquitous data</p> <p>Data 'behind' scientific papers available in machine readable formats</p> <p>Journal / database hybrids</p> <p>Knowledge base for molecular mechanisms and architectures</p> <p>Truly smart lab-books</p> <p>Managed platforms and 'Service-Oriented Architectures' being widely created and adopted in science and scientific computing, but also for the 'new kinds' of conceptual and technological tools in science</p> <p>Symbolic computation integrated into scientific databases and programming languages</p> <p>Advances in database technology starts to enable: Scalability, distributed optimisation, fault-tolerance, self-management and integrations of heterogeneity, metadata and schematisation.</p> <p>Concept of intelligent interaction starts to become a reality (combining data cubes, transformations in science publishing, visualisation, distributed data management). Easy to program, easy to use interfaces and smart lab-book devices/interfaces</p> <p>Ability to rapidly build complex distributed systems from shared components that are: trustworthy, predictable, reliable, scalable and extensible</p>	<p>Information retrieval and search for novel data types</p> <p>Data and metadata availability / requirement from publishers, together with increasing online availability, starts to transform publishing model</p> <p>Large-scale sensor networks starting to proliferate</p> <p>Integration of data, metadata, open access and publishing continues to transform scientific publishing model</p> <p>'Active XML' for ubiquitous data</p> <p>Data 'behind' scientific papers available in machine readable formats</p> <p>Journal / database hybrids</p> <p>Knowledge base for molecular mechanisms and architectures</p> <p>Truly smart lab-books</p> <p>Managed platforms and 'Service-Oriented Architectures' being widely created and adopted in science and scientific computing, but also for the 'new kinds' of conceptual and technological tools in science</p> <p>Symbolic computation integrated into scientific databases and programming languages</p> <p>Advances in database technology starts to enable: Scalability, distributed optimisation, fault-tolerance, self-management and integrations of heterogeneity, metadata and schematisation.</p> <p>Concept of intelligent interaction starts to become a reality (combining data cubes, transformations in science publishing, visualisation, distributed data management). Easy to program, easy to use interfaces and smart lab-book devices/interfaces</p> <p>Ability to rapidly build complex distributed systems from shared components that are: trustworthy, predictable, reliable, scalable and extensible</p>
<p style="writing-mode: vertical-rl; transform: rotate(180deg);">Issues</p>	<p>How to ensure new kinds of tools are user (scientist) friendly?</p> <p>Commercial software enterprises increasingly working with science community to freely provide code, applications and components to science community on shared-source basis</p> <p>Need for better collaboration between computer scientists and other scientists, and industry - how to build new kinds of communities</p> <p>Education - how to produce 'new kinds' of scientists now urgently needed (computationally and mathematically highly literate)</p> <p>Need to raise public awareness of science and global challenges</p> <p>Challenge of sustaining / re-enforcing the scientific principle under increasing complexity and computationalisation: falsification, repeatability, public domain, competing schools of thought</p> <p>Open Source development, investment and economic model starts to mature</p> <p>Education - how to produce 'new kinds' of scientists now urgently needed (computationally and mathematically highly literate)</p>	<p>How to ensure new kinds of tools are user (scientist) friendly?</p> <p>Commercial software enterprises increasingly working with science community to freely provide code, applications and components to science community on shared-source basis</p> <p>Need for better collaboration between computer scientists and other scientists, and industry - how to build new kinds of communities</p> <p>Education - how to produce 'new kinds' of scientists now urgently needed (computationally and mathematically highly literate)</p> <p>Need to raise public awareness of science and global challenges</p> <p>Challenge of sustaining / re-enforcing the scientific principle under increasing complexity and computationalisation: falsification, repeatability, public domain, competing schools of thought</p> <p>Open Source development, investment and economic model starts to mature</p> <p>Education - how to produce 'new kinds' of scientists now urgently needed (computationally and mathematically highly literate)</p>	<p>How to ensure new kinds of tools are user (scientist) friendly?</p> <p>Commercial software enterprises increasingly working with science community to freely provide code, applications and components to science community on shared-source basis</p> <p>Need for better collaboration between computer scientists and other scientists, and industry - how to build new kinds of communities</p> <p>Education - how to produce 'new kinds' of scientists now urgently needed (computationally and mathematically highly literate)</p> <p>Need to raise public awareness of science and global challenges</p> <p>Challenge of sustaining / re-enforcing the scientific principle under increasing complexity and computationalisation: falsification, repeatability, public domain, competing schools of thought</p> <p>Open Source development, investment and economic model starts to mature</p> <p>Education - how to produce 'new kinds' of scientists now urgently needed (computationally and mathematically highly literate)</p>
<p>2005</p>	<p>2010</p>	<p>2015</p>	<p>2020</p>

Notes:

This draft roadmap is intended to be used as a simple guide to some of the developments underway or required in science and computing towards 2020.

The items in the roadmap represent a selection of the key milestones and challenges either underway, expected or required. We have not, however, attempted to incorporate every development possible or required. That is, our draft roadmap does not represent an exhaustive set of items.

In order to keep this draft roadmap simple, we have focused mainly on the biological domain - but developments in other research areas (such as energy and particle physics) will face similar challenges and be underpinned by similar milestones.

The linkages (dependencies, relationships) between elements (key milestones or challenges) do not appear on the roadmap because of the extent and complexity of such linkages. In order to keep the roadmap visually clean and readable, even the primary linkages have been omitted.

© Microsoft Corporation, 2006