

# **E-Science research focus FNWI**

Arnold W.M Smeulders

Willem Bouten

Antoine H.C van Kampen

## Nederlandse samenvatting

Een steeds groter deel van de wetenschap leunt op e-Science, het vakgebied dat zich bezig houdt met het verwerken, organiseren, modelleren en analyseren van grote hoeveelheden data. Zonder e-Science zouden veel moderne wetenschappelijke gebieden en maatschappelijke uitdagingen onontgonnen blijven, omdat ze zonder rekenkracht, grootschalige gegevensopslag en snelle toegang daartoe, niet te betreden zijn. Aan steeds meer wetenschappelijke prestaties ligt dan ook e-Science ten grondslag, zoals onder andere blijkt uit het grote aantal Spinoza prijzen dat erop steunt.

Het zwaartepunt e-Science zorgt in de breedste zin voor die ondergrond. In de fundamentele e-Science gaat het daarbij om de principes van informatieverwerking. De kernvraag is heel oud: welke principes van informatie en kennis bepalen de processen in de wereld? Wat is een goede manier om informatie en kennis te representeren, en wat is de juiste manier om ermee te redeneren? Het antwoord op deze abstracte vraag is door de computer in een stroomversnelling gekomen, maar de vraag naar de relatie tussen data, informatie en kennis is nog steeds de kern.

Bovenop de fundamentele laag ligt die van de strategische e-Science. In deze laag zitten de componenten van de kringloop die ten grondslag ligt aan de exacte wetenschappen: experimenteren, analyseren, theoretiseren, modelleren – en opnieuw experimenteren. Technieken voor het mobiliseren van data- en rekenkracht, voor netwerken, voor modelleren en simuleren, voor analyseren van data staan hier centraal. Ook het visualiseren van informatie en de integratie van al deze componenten horen in deze laag. In de gamma- en alfawetenschappen is een dergelijke kringloop ook in opkomst, zonder experimenten, want die zijn daar vaak onmogelijk. Maar dat doet niets af aan de uitdaging voor het afgraven van bestaande gegevens in dit geval vaak van verschillende oorsprong.

Deze strategische componenten worden ingezet in de toepassingsgebieden van e-Science. Astrodeeltjesfysica, cognitie en biosystemen zijn goede voorbeelden van velden waarmee het zwaartepunt e-Science intensief samenwerkt. Ook de voedsel industrie, de materiaalkunde, de epidemiologie of waterbeheer, misdadbestrijding en televisiearchieven liggen op de weg. De hele samenleving ondergaat een informatie-explosie waarin spannend inzichten verborgen liggen.

Het zwaartepunt heeft een sterke uitgangspositie in Amsterdam. Naast AMS-IX, een van de grootste internetknooppunten ter wereld, zijn in de Watergraafsmeer het CWI en het academisch rekencentrum SARA gevestigd, en de VU. Alle expertise en faciliteiten binnen handbereik voor een ambitieuze ontwikkeling van het Science Park tot een e-Science centrum van wereldformaat. Bijna geen vakgebied kan nog zonder die kennis en infrastructuur.

*Aantal medewerkers* 100 nu

*Internationale waardering* : Het Cook report waar Nederland ten voorbeeld wordt gehouden aan de USA. De internationale review van de Bsik-projecten MultimediaN, VLe en Gigaport (netwerk).

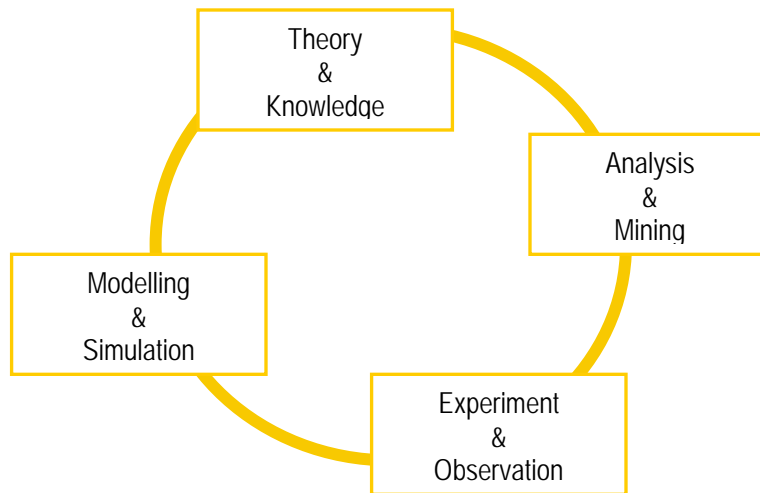
*Gewonnen prijzen* : Spinoza, VICI, VIDI, VENI.

*Deelname projecten* : BigGrid, NBIC, VanVelzen infrastructuur, e-Science Centre, VidiVideo, ViroLab en vele andere

*Samenwerking* : SARA, CWI, VU, AMOLF, SURF, AMC, Spinoza, CCCT.

# 1. Introduction

E-Science is the systematic science of collaborative compute- or data-intensive research. It is the result of the remarkable and ever increasing impact that the advent of the computer, the computer networks and the digitization of all information continue to have on science and, for that matter, the global society in general. In the new era that we are witnessing, explicit knowledge will be systematized and made accessible through digital libraries, patent-files, Wikipedia, and the Internet in general. In fact, all parts of the iterative cycle of science are affected by digitization, systemization or computation. Where compute power of a



*The cyclical flow of information in research growing quickly in volume and pace by the digitization, now touching all aspects*

single computer has doubled every two years for more than thirty years in a row, the capacity of networks has doubled every year. The capacity of storage grows even faster, only to be surpassed by the sensing capacity. As a consequence, in almost all branches, data-intensive or compute-intensive arts and sciences have spawned off.

E-Science is structured as most scientific disciplines. There is a fundamental layer

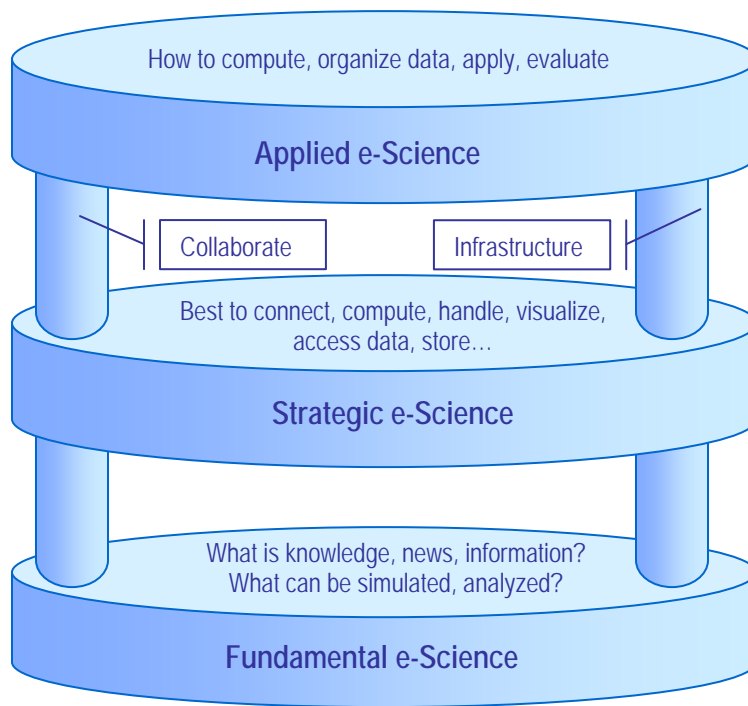
addressing fundamental questions, a strategic layer providing the tools on which e-Science applications can be build. There is nothing new here. All areas of science do so. Only the scale of impact of e-Science covers so many and so different areas of science (and arts) and society. The wide range of areas makes it an interdisciplinary science par excellence. Like in most other sciences infrastructure and collaboration are other essential ingredients, while the rapidly expanding power in computing, network bandwidth, and data storage is the driving force in the case of e-Science.

*Fundamental e-Science* is about the principles of information processing. The core questions are very old: what is information and knowledge? How do they determine the processes in the world? What are proper measures to express information in? How can we formally reasons with information? The answer to these questions has been accelerated by the advent of the computer but the core still is the essence of the interplay between data, information and knowledge, with or without a computer.

*Strategic e-Science* is about methodology. It will find answers to questions like: Can we with the aid of computer engineering unleash the compute power and the data when distributed over many connected machines? Can we store and handle all data resulting from experiments? And, can we simulate phenomena to predict the outcome of experimentation or even to predict the future (for example

meteorology)? Can we design sound methods of computation? Can we assess the information and knowledge content in the data? And can we interact effectively with the user?

*Applied e-Science* concerns the science in other disciplines than ICT that is based on a multidisciplinary approach and enhanced by information and communication technology. When counted by impact, the success of applied e-Science is so overwhelming that it has surpassed the level of simple citations. E-Science already plays a dominant role in the so-called big sciences, but it is now rapidly moving into the daily practice of many other sciences and more recently in arts as well. It enables the analyses of hierarchical systems of higher complexity and massive amounts of data.



*Infrastructure and collaborative E-Science* enables large scale science and distributed global collaborations through the Internet<sup>1</sup>. A typical feature of such collaborative scientific enterprises is that they require access to very large data collections, computational methodologies, very large scale computing resources and high performance visualization in the interaction with the individual scientists.

*On the basis of its impact,* e-Science has been named spearhead by the dean of the Faculty of Science and accepted by

the president of the University of Amsterdam as part of the strategic plan of the University in 2008. The mission of the spearhead is to perform fundamental, strategic, and applied e-Science research and to initiate or intensify e-Science in collaboration with the other sciences and arts. It is based on the vision that in this era e-Science will be a determinant factor in the scientific progress of many disciplines. It is also based on the conviction that these developments will have big societal effects and will find their way in many sectors of the economy. Whether it is knowledge intensive industries (e.g. in the life science sector), the financial service industry, or the retail industry, e-Science methodologies will have a huge impact.

In the sequel of this report a summary is given of this new emerging field, bridging computational science and a broad range of other disciplines, in the sciences, the humanities and the social and behavioural sciences. It is written from the perspective of the FNWI, where e-Science next to system biology and astroparticle physics has

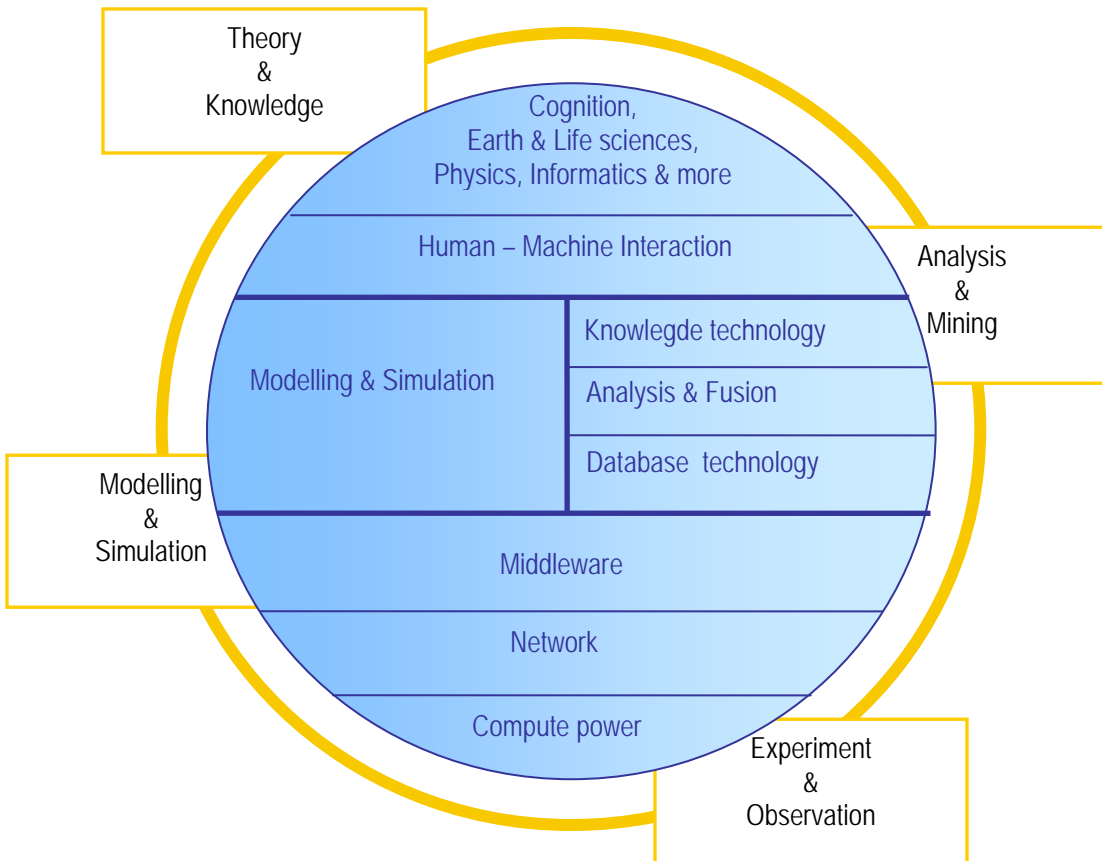
<sup>1</sup> (<http://www.nesc.ac.uk/nesc/define.html>)

been selected as faculty spearhead, with e-Science applications an important element of the latter two as well. The e-Science activities, however, should be seen in a broader framework in Amsterdam. Amsterdam has become an internationally recognized e-Science centre, because of its contributions to the computational e-Science foundations, the infrastructure and the way in which e-Science has merged with other disciplines. NWO institutes in Science Park Amsterdam, like CWI, AMOLF, NIKHEF, the Dutch High Performance Computing and Networking Centre SARA and the VU all contribute to this internationally strong position.

## 2. Evolving perspectives on e-Science

*Hard infrastructure.* E-Science started as the computationally intensive part of science that is carried out in distributed network environments, or science that uses immense data sets that require grid computing. From the compute point of view, e-Science merely is the infrastructure of computers to deliver the compute power to those who wish to compute. We have a broader vision where also critical components of the computational methods and software are part of the infrastructure to enable reuse from application to application. Typical components which one often finds in e-Science applications are covering advances in methodology in middleware, data storage, data analysis, measurement methods, statistics, visualisation, interaction, data mining, modelling and simulation. We adhere to the full circle vision of e-Science.

*From the point of view of the sensory data,* on the one side there is the tendency of generating as many data as one can for future evaluation". Very large repositories have been formed in this way. With the advent of digital sensing devices, networks and storage capacities it is only tempting to generate more. The more data one has the more important it is to have available generic methods for documenting the data.



*Tools of e-Science in the information circle of science*

On the other side, specific modelling for one specific experiment make the evaluation very costly and reuse impossible. Some sensors are unique (LOFAR, CERN) by

necessity, some are more widely available (fMRI instruments). In all cases, the more idiosyncratic the equipment or the data are, the more expensive it will be to evaluate the data and the larger the necessity to reinvent the methodological wheel for the purpose at hand. Standardization of data formats and software methods pays off.

From the perspective of *applied e-Science*, there is a independent methodologies of - computer science to specific scientific questions in an application area of e-Science. E-Science aims to fulfil the needs of application domains on the one hand and at the same time should bring innovation and fundamental research in computer science on the other. Technological developments now enable multi-disciplinary participative science that can build upon sharing data of increasing scale and diversity, sharing methodologies, tools and sharing the infrastructure, and enabling collaboration in virtual organizations. The progress is so powerful that e-Science has spawned new research fields with their own set of methods, demands and best-practices.

Apart from the hard infrastructure also the *soft infrastructure* is important. It ranges from the compute layers, to the network and the set of methods up to interaction. An intrinsic part of the infrastructure is remote communication in virtual organizations in a multi-disciplinary environment. The best chance on success is to bring together teams who are able to identify innovative technological or methodological developments. The teams have a capacity to judge how these achievements can contribute to solving our problems implementing the tools in their daily routine. The teams would typically be multi-disciplinary and cooperating on a daily basis, preferably on one place and with ample time for knowledge communication and education. There is much to be gained along these lines of cooperation but such requires investments of time and money.

*The value of corporation.* The smaller the gap between a traditional scientific domain and information and computer science, the closer e-Science is to the daily routine. Particle physics, astronomy and meteorology are used to handle far larger amounts of data than biology or history. Other sciences like food science profit from the economic impact of innovation that facilitates the change. *The bigger the gap, the more to gain* would be an easy saying. However, the community may be too small and journals may not be too keen to accept a different approach. And, it is difficult to ask new scientific questions when unaware of the possibilities. The goal should be to explore the value of cooperation. We endorse a program where fundamental, strategic and applied e-Science are in close contact to mutual benefits.

### 3. Advances and challenges for e-Science

All science cases described below require a rich landscape of (integrated) resources, consisting of networks, computing, and storage and computational methods.

#### ***The domain of informatics***

*Computation - fundamental.* E-Science as a name and a science started from computation. Computation as needed in processing the endless streams of experimental data, or computation as needed in the running of model simulations. Hence, computation is at the core of the methods and the many applications of e-Science. One way to expand the possibilities is through GRID computing, distributing the work-load over whatever computer is available in the world-wide GRID networks without any effort to the user or even the programmer, a strong topic at the VU.

*Computation – strategic.* The public – private programs VLe and BigGrid have delivered infrastructural layers for computation as well as the actual machines to make large scale computations. SARA delivers large scale computation. An important cluster for distributed processing is located in the Free University, with whom there is a standing cooperation around the parallel compute machines DAS3, DAS4 of the ASCI research school.

*Systems and Networking – fundamental.* Computation and data exchange occur through the network. For such reasons, advancing e-Science implies advancing the speed and tailoring of networks and its data description and protection layers. The Informatics Institute at the UvA as part of SurfNet and Gigaport research, has a good reputation on fundamental network research achieving very high bandwidths. In the Watergraafsmeer, the AEX internet exchange is one of the biggest in the world. Advance networks are strong at the UvA.

*Systems and Networking – strategic.* Newly emerging e-Science applications need advanced services from the network. This requires that both application and network interface more closely to fulfil the computation needs at hand. DataGRID is emerging to deal with data growing in isolation separated physically, by different sensors, by different data formats, or by different taxonomies. The purpose here is to make the data universally available.

*Data storage and exploration – fundamental* Another component of e-Science is data storage. As storage has grown even faster than compute power, the balance of computer systems now tips towards keeping all data. This raises an enormous problem under the name of the data explosion. Data warehouses, databases for diverse datasets and federated databases may offer solutions. On the topic of data exploration, text analysis and picture or video search engines cannot be missed in retrieving targets from very large collections.

*Data storage and exploration – strategic* In such application areas as geo-sciences, earth observations, astronomy, high-energy physics, television science, historic sciences, and cognition data sets are growing and growing beyond comprehension. And in effect, many new experiments may result in a new insight how to interpret or how to process old data. Hence, databases are an essential element of e-Science where the challenge is how to integrate dispersed and heterogeneously data dictionaries grown in the past under dynamic conditions. Fundamental and strategic

research in data storage is a strong topic at CWI and UvA, and in data exploration at UvA.

*Simulation – fundamental.* Combines mathematics fused with numerical methods and (distributed) computational techniques to model systems mostly at one level to obtain insights of systems at a higher level of aggregation. Often, this requires heavy computation calling for GRID-computing, workload management or distributed processing. There are applications in genomics and cellular dynamics and environmental modelling.

*Simulation – strategic.* On the strategic side, there is a connection with modelling the history of stars and galaxies in astronomy. There is participation in VL-e, as well as the leadership in a few EU-projects in this area, including one on the prediction of the spread of viral diseases. In addition, another part of the VL-e project has landed in distributed computing solutions. Fundamental and strategic research in simulation is a strong topic at UvA.

*Human computer interfacing.* Is becoming increasingly important. With the load of data, the bottleneck of systems will be visualizing the results, be it of the data analysis (multi-dimensional data visualization), of the knowledge (knowledge tree visualizations), a virtual reality (for example how molecules see space when they lock in chemical reaction), or visualization on the real world. Experiences and usability are increasingly important in the human computer interface. Research in human-computer interfacing is a strong topic at CWI and UvA. Virtual reality visualization is strong at SARA and UvA, and data visualization is important in the UvA as well.

### **The domain of biology**

*Systems biology – strategic e-Science.* Is another spearhead of the faculty. It is a new but rapidly developing paradigm in life sciences. 'Omics' technologies and ICT developments of the past 10 years allowed us to get a nearly complete overview of all components and processes in multi-scale biological systems, ranging from the molecular level to ecosystems. The functioning of these systems and the prediction of their behaviour in specific situations is difficult to establish from this information due to the complexity of these systems. The big challenges in the field of systems biology are at the borders of biology where tight integration is required with chemistry, physics, engineering, mathematics and informatics. E-Science will contribute to systems biology by providing massive data handling, much of the methods for data analysis, advanced information management, and the compute power for multi-scale modelling.

*Molecular biology – applied e-Science.* Molecular biology is increasingly using high-throughput genomics technologies that produce large amounts of data. The management and processing of these data are increasingly difficult challenges and, consequently, there is a large need for state-of-the-art methods, compute and storage facilities, and (virtual) environments that integrate these methods and facilities to support knowledge extraction from the data. This would benefit not only life sciences research at the UvA but also many Dutch consortia such as the genomics centers of the Netherlands Genomics Initiative (e.g., CMSB, CBSG, CGC, Kluyver, NMC, NPC), the national bio banking project String of Pearls (collaboration between all academic hospitals) and the Centre for Translational Molecular Medicine (CTMM) and Top Institute Pharma (TIPharma). E-Science research can provide many solutions to the life sciences community.

*Biodiversity – applied e-Science.* All over the world, organizations collect samples of species and make observations of their occurrences and abundances. These heterogeneous data become more and more available through distributed databases. This is not only scientifically exciting but also socially very relevant in the context of global change, emerging diseases, decreasing biodiversity and waning resources. The e-Science challenge is to cope with this wealth of information and to enable data federation and collaboration within the ecological sciences to obtain new ecological insights. EcoGRID, under construction within the VL-e project, is building the National Distributed Database of Flora & Fauna which contains all the data of private organizations that collect data, allied in the Association for Research of Flora & Fauna. It covers already more than 90% of all national spatio-temporal observations of species (over 30 million records). EcoGRID philosophies now play an important role in the design of LifeWatch, the ESFRI biodiversity research infrastructure. Another player is NLBIF, the Dutch node of the international Global Biodiversity Information Facility (GBIF).

*Bioinformatics and biostatistics.* Currently, the organization, integration, annotation and analysis of large and heterogeneous datasets poses a challenge. In addition, the need to develop approaches that allow efficient dissemination, sharing, re-use and integration of methods. Bioinformatics and biostatistics at the FNWI and AMC addresses these challenges and are important partners of the Netherlands Bioinformatics Centre (NBIC; NGI centre), which runs large research, support, and education programs. E-Science methodologies and approaches play a crucial role in the NBIC support program for which strategic collaborations with VL-e and BigGrid were established. NBIC participates in ESFRI ELIXIR (European Life Sciences Infrastructure for Biological Information) and in ESFRI EATRIS (European Advanced Translational Research InfraStructure in Medicine). The e-Science spearhead is expected to make significant contributions to bioinformatics and statistics at the UvA, in the Netherlands and at the European level.

### **The domain of earth and environmental sciences**

*Earth & environmental – strategic e-Science* are becoming ever more predictive sciences, where quantitative data analysis and numerical modeling play a most important role. Satellites with increasing resolution, GPS-technology, and extending sensor networks, are generating massive amounts of data about a great variety of biotic and abiotic phenomena. The e-Science goal is to manage and interpret these data, and to develop quantitative models that use these data to understand, reproduce and predict the complex biological and physical processes in geo-ecological systems. The e-Science challenge is to develop generic components for virtual laboratories that enable visualization and analyses of these massive amounts of observations integrated with even larger data cubes produced by compute intensive, multi-scale spatial-explicit dynamic simulation models. Inverse modeling and data assimilation will generate the feed-back that is needed to improve our theories and models.

### **The domain of physics & astronomy**

*Particle physics* poses large problems in term of the huge data volumes handled by thousands of scientists scattered around the world to analyze the data. A multi-tier hierarchical computing grid is in place: a Tier-0 centre at CERN connected to about ten Tier-1 centers scattered all over the world; the Netherlands (Nikhef / SARA) is one of them, as part of the BiG Grid project. This will ensure continued international

prominence for the Dutch high-energy physics community during the LHC era. This position has already started to show spin-offs in related areas, for example the Dutch leadership in the IT developments for the new field of neutrino telescopes in high-energy astrophysics (KM3NeT), which is one of the priorities on the Dutch ESFRI list.

*Astronomy* also has started to generate huge amounts of precisely timed data. LOFAR aims to sense the sky by 25.000 antennas scattered over a larger part of North Western Europe. The electronic signals from the antennas are digitized, transported to a central digital processor, and combined in software to emulate multiple conventional antennas, effectively creating a telescope looking simultaneously in eight different directions. To make radio pictures of the sky with adequate sharpness, these antennas are to be arranged in clusters that are spread out over an area of ultimately 350 km in diameter. Data transport requirements are in the range of many terabits/s and the processing power needed is tens of teraflop/s, delivered by the Blue Gene supercomputer 'Stella'. LOFAR can also be used as a more generic Wide Area Sensor Network. Sensors for geophysical research and studies in precision agriculture have been incorporated in LOFAR already. Several more applications are being considered, given the increasing interest in sensor networks that 'bring the environment on-line'.

*Computational physics.* The Computational Physics Group at AMOLF studies the phase behaviour, dynamics of colloidal and (bio) polymeric systems and the kinetics of first-order phase transformations in such systems. Computer simulations with numerical models based on interactions between atoms and molecules are used to perform virtual experiments. Comparisons with true experiments are used to improve the theory. New simulation algorithms are developed to increase the efficiency of parallel simulations. The focus of the research is on numerical simulations of crystal nucleation, the competition between crystallization and liquid-liquid de-mixing in polymer systems, and the phase behaviour of colloidal particles. AMOLF has a track record on this fundamental research, proved by the Spinoza prize that was awarded. In near future, the focus of the research will shift towards coarse grained simulations of organization and transport of bio-molecules and materials containing bio-molecular building blocks.

### **The domain of computational chemistry**

*Molecular simulation* has the potential to play a central role in the design of new materials and processes, and in the modelling of biological processes. However, even if we allow for a thousand fold increase in computing power, there remains a huge gap between the length and time scales that can be studied in simulation and those that are relevant for most industrial and biological processes. Novel computational techniques will have to be developed to bridge this gap. To this end atomistic simulations and coarse-grained mesoscopic models that work on length and time scales that are larger by three to ten orders of magnitude are integrated.

*Catalysis for energy conservation.* On a global scale, energy conservation and ecology-friendly industrial production processes have captured much attention. Examples are the development of effective detergents at low temperatures and the development of low-energy processes in the oil industry. An important area in chemical science copes with these phenomena and problems. The work on heterogeneous catalysis targets the interaction between molecules in a gas or liquid with molecules on a (solid) surface. The goal is to develop and to construct surface materials such that the desired chemical reactions will be pursued efficiently, with minimal energy requirements, i.e. at low temperatures. Experimentally, it is very

difficult to manage these processes, since they execute at molecular and even atomic levels. Large computer simulations are the only way to model the reactions involved.

*Fluid dynamics.* Fluid dynamics is the scientific discipline that covers all kinds of flow phenomena: hydrodynamics, aerodynamics, climate research, meteorology, combustion processes, etc. With the presence of turbulence, flow processes in general are extremely complex. They contain many different time and length scales, as we can see for instance in stormy weather or in burning flames. Direct Numerical Simulation is needed to describe these detailed processes accurately. Progress in understanding turbulence will need both access to the fastest low-latency computers, as well as continuous effort in numerical methods and implementation software.

### **The domain of cognition**

*Cognition.* In the experimental science of cognition, typically large amounts of fMRI and EEG-scans are being generated from a patient or a subject. In one typical strand of experiments, an individual is shown a picture for 20 msec and the responses of the brain are being measured. The typical data volume is MBytes per EEG or fMRI recording. The subsequent processing is laborious, and involves advances stochastic processing of the brain data including gauge recordings, modelling on the basis of a priori knowledge and with minor question-specific alterations in the computational methods. The number of images per experiments is currently a thousand. Both the computational complexity and the number of data are expected to grow in the next year due to the new 7 Tesla scanner as it becomes available for research. There is a strong link with the University of Amsterdam general spearhead for cognition, the ACCS and the Institute of Informatics on the topic of human and computer cognitive vision.

### **The domain of content and technology science**

*In communication sciences,* there is a developing link on the topic of media and political decision making as studied by ASCOR and topic of an ERC grant - and the Institute of Informatics studying search engines for Internet information. They are both members of the newly found multidisciplinary Centre for Creation Content and Technology, CCCT, at the University of Amsterdam.

*Television and media sciences* are developing into a digital domain as well. Currently television science is mostly analogous but quickly the consequences of the digitization kick in here as well. At the same time, the top-performers over the last five years in the world-wide competitions on video search engines, moving from manual collection of impressions to searching digital ones. At the same time, the Beeld & Geluid national archives turns digital not only for the incoming streams of video but also for their archives. Future collaboration between the "arts" and the "sciences" is expected to lead to further innovations in the humanities and social sciences.

*Historic sciences.* Large digital collections of annotated speech and other language data (text, audio and video), data bases on population characteristics (census data, population registers), economic time series, and geospatial data require new approaches in fields such as phonetics, sociology, econometrics, demography, history, geography and archaeology. DANS, the national center for permanent access to research data for the humanities and social sciences, is exploring the potential of the data grid and the linkage of heterogeneous data resources.

*The alpha and gamma domains* in general have quickly growing demands for data storage. They run easily into the Peta and Exa bytes. The corresponding computation needs develop similarly. Also research for the development of text or video search engines have not yet reached the half way either. Ample opportunities are here to develop e-Science. As the University of Amsterdam takes key positions here on the application side as well as on the methodological and computational side, good progress to an international leading position is foreseeable.

## 4. Education

To boost e-Sciences the faculty requires the facilitation of cross-disciplinary communication and cooperation. To improve communication and cooperation we also need to incorporate the e-Science approach in our disciplinary education programs both at BSc and MSc level. These investments in the community will be hard to measure.

The extent to which (applied) e-Science is incorporated in education varies from discipline to discipline. On the one hand, the curriculum of physics is hard to imagine without e-Science while on the other hand e-Science currently hardly plays any role in the curriculum of biology. If we want to profit from e-Science developments, we have to invest in training students in the appreciation for the new opportunities. As a consequence, we need to develop dedicated e-Science programs or tracks at the Master and PhD-levels. Teaching young researchers to properly apply innovative techniques is essential if we want to further catalyze the knowledge discovery process. e-Science tracks in (existing) master programs should therefore aim at preparing students for a PhD in one of the e-Science application domains. We expect them to initiate new directions in research at the interface of their own discipline and e-Science. Therefore, this type of master programs should be available for students with a background in computer science, physics, chemistry, earth sciences, biology, or applied mathematics with an intrinsic motivation to work across the borders of traditional disciplines.

## 5. Means, output and quality

### Milestones and signs of quality

*Peer reviewed papers*

*Examples of collaboration*

*Example(s) of ICT infrastructure*

*Tokens of honour*

*Valorisation*

### Means

*Professorships*

*PhD students*

*Dissemination and conferences*

*Valorisation*

*Exchange of personnel*

*Exchange of software*

*Education program*

*Management*

### Budget

<i>Professorships &amp; senior staff program</i>	150 Keuro per year
<i>PhD students e-Science</i>	150 Keuro per year
<i>Dissemination and conferences</i>	30 Keuro per year
<i>Valorisation support</i>	10 Keuro per year
<i>Exchange of personnel program</i>	50 Keuro per year
<i>Exchange of software</i>	80 Keuro per year
<i>Education program support</i>	40 Keuro per year
<i>Management</i>	60 Keuro per year
<i>Total</i>	670 Keuro per year

in the usual metric in the understanding that even Van Raan finds it impossible to measure citations in computer science and all related fields;

in multi-disciplinary virtual collaboration; and supported applications facilitating collaboration;

both in fundamental, strategic and applied e-Science;

as part of public private projects such as the bsik-programs or possibly in cooperation with the e-Science centre.

Part-time and/or temporary appointments of professors or key researchers to boost directions and cooperation. Priorities are in the area of:

- Spatial data mining
- Cognitive data processing
- Information management;

PhD students on the basis of a plan and competition. Some emphasis on impact in an other discipline.

of scientific results;

into commercial or non-profit exploitation;

between groups in the spearhead to achieve progress in multi-disciplinarily;

by support through a software engineer;

initiation of e-Science courses in biology and computer science or AI;

10% of the budget for support, dissemination and promotion activities (folders, distinguished lectures)

## 6. Organisation

The e-Science focus of the faculty strives to be an active cooperation with the national e-Science Centre, currently being defined by ICT Regie.

The e-Science spearhead is governed by Arnold Smeulders (chair), Antoine van Kampen, and Willem Bouten by appointment of the dean. Funding will be assigned by the Dean to the respected institutes on the basis of a proposal by the spearhead committee. Funds will be administered content-wise by the committee and financially by the allotted institute provided they have received the yearly go-ahead by the committee. The project will report on the basis of a concise yearly plan and a concise quarterly report submitted to the spearhead committee.

### List of most directly affected chairs

<i>Hamideh Afsarmanesh</i>	UvA	Federated Databases
<i>Frank Baas</i>	AMC/ UvA	Molecular Biology/genomics
<i>Henri Bal</i>	VU	Distributed Computing
<i>Guus Schreiber</i>	VU	Large Scale Ontologies
<i>Peter Bolhuis</i>	UvA	Computational Chemistry
<i>Willem Bouten</i>	UvA	Computational Geo-ecology
<i>Roel van Driel</i>	UvA	Systems Biology
<i>Daan Frenkel</i>	AMOLF	Computational Physics
<i>Jef Huisman</i>	UvA	Aquatic Ecology
<i>Antoine van Kampen</i>	AMC/UvA	Medical Bioinformatics
<i>Cees de Laat</i>	UvA	Advanced Networks
<i>Victor Lamme</i>	UvA/FMG	Cognition Data Computation
<i>Robert van Liere</i>	CWI	Visualization
<i>Martin Kersten</i>	CWI & UvA	Large Databases
<i>Maarten de Rijke</i>	UvA	Text Retrieval Engines CCCT
<i>Peter Sloot</i>	UvA	Simulation, Astronomy, Medicine
<i>Arnold Smeulders</i>	UvA	Multimedia Analysis CCCT
<i>Ed Tan</i>	UvA/FMG	Communication Sciences CCCT
<i>Ronald Wanders</i>	AMC/UvA	Biochemistry
<i>Lynda Hardman</i>	CWI/UvA	Interactive Information Access
<i>André de Roos</i>	UvA	Theoretical Ecology
<i>Richard Rogers</i>	UvA/FGW	Media Sciences