I-Science

Computer science
Research programme cluster
by
NWO Physical Sciences

GLANCE – GLobAl computer scieNCE
VIEW – Visual Interactive Effective Worlds
STARE – STAR E-science

NWO
Netherlands Organisation for Scientific Research

Contents

Preface

The programme cluster I-Science aims for reinforcing the research discipline computer science in the Netherlands and creating clusters of expertise. Another aim of this programme cluster is to contribute to preservation and improvement of the prominent position of the Netherlands in the international developments involving e-science. The programme cluster consists of three research programmes: GLANCE, VIEW and STARE. GLANCE focuses on large-scale parallel and distributed systems, VIEW focuses on generic visualization techniques, and STARE is a research programme stimulating multidisciplinary research on the borderline between computer science and astronomy. The first chapter of this text concerns the general background of I-science. Chapters 2, 3 and 4 explain the scientific aims and content of each of the programmes that are part of the programme cluster. In Chapters 5 and 6 the organization of the programmes and the assessment and schedule are explained respectively.

Chapter 1 – Background

1.1 NWO Physical Sciences Strategy

In the previous years, the Physical Sciences council of the Netherlands Organization for Scientific Research (NWO) has worked towards a coherent reinforcement of research in computer science. This policy is based on the National Research agenda for computer science[1] (NOAG-i), which identifies seven priority research themes. Strengthened by a growing budget, the same policy will be continued in the upcoming years, focusing on three strategies; <u>Deepening strategy:</u> Continuing the reinforcement of the discipline computer science, and its research groups; <u>Cooperation strategy:</u> Intensifying the cooperation with other sciences, social sectors and business. <u>Investment strategy:</u> Extension of the national computer infrastructure and utilization of the facilities.

The NWO Physical Sciences strategy for the future of Computer Science is visualized in Figure 1. The Computer Science discipline (inside ring) is being reinforced and profiled. Starting from the inside ring, the cooperation with other science disciplines is stimulated. From this point of view the NWO theme *Digitization and Information Technology* is implemented. This theme is one of the nine trans-disciplinary themes selected by NWO, themes which are of importance for society, into which NWO is channeling substantial investments and will continue to do so. For computer science the extension to other science areas within the theme Digitization and Information Technology is a logical step since new computer science techniques are essential for their progress.

[1] Nationale Onderzoeksagenda Informatica 2001-2005 "NOAG-i", 2001

1.2 The NOAG-i themes

The NOAG-i identifies seven scientifically challenging and socially relevant research themes. These themes are:

- Parallel and Distributed Computing (PDC)
- Embedded Systems (ES)
- Software Engineering (SE)
- Multimedia (MM)
- Modeling, Simulation and Visualization (MSV)
- Intelligent Systems (IS)
- Algorithms and Formal Methods (AFM)

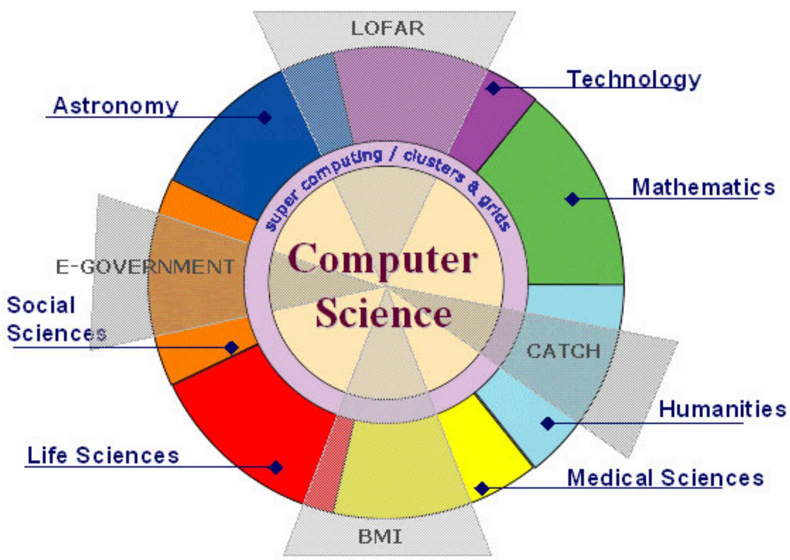

Figure 1: The physical sciences council strategy

The implementation of the NOAG-i themes by the Physical Sciences Council is as follows:

Theme	PDC	ES	SE	MM	MSV	IS	AFM
Program-mes (*= new)	GLANCE* BRICKS*	PROGRESS (via STW)	JACQUARD	CATCH*	VIEW* BRICKS*	ToKeN CATCH* BRICKS*	BRICKS*
	Open Competition						

The new programmes GLANCE (GLobAl computer scieNCE) and VIEW (Visual Interactive Effective Worlds) are part of the I-Science programme cluster and will focus on fundamental research within the themes PDC and MSV. The thematic content of these programmes will be described in more detail in Chapters 2 and 3. The I-Science cluster will subsidize relatively large projects (3-4 executors). Small projects within the same theme can be submitted in the Open Competition. These projects will be involved in the programme activities of the I-Science cluster. The Bsik research programme BRICKS, which is enforced by the CWI, will allocate part of its budget to researchers through a competition organized by NWO, called FOCUS. FOCUS is directed at providing encouragement for individual researchers, in the same style as NWO's Innovational Research Incentives Scheme[2]. I-Science however is directed at strengthening research groups that are active within the relevant themes, by means of relatively large grants.

1.3 I-Science, the programme cluster

The programme cluster I-Science (eye-science) consists of the three programmes VIEW, GLANCE and STARE. GLANCE and VIEW result directly from the NOAG-i themes as described in the previous section. STARE is a multidisciplinary programme on astronomy and computer science. This programme is described in Chapter 4. By subsidizing relatively large projects, I-Science aims to contribute to the formation of clusters of expertise within computer science research in the Netherlands.

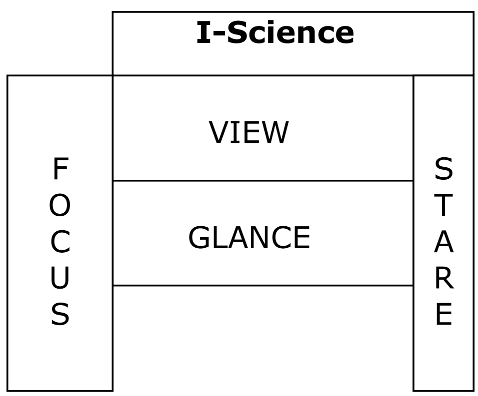

Figure 2: The connection between the new computer science programmes at NWO

Figure 2 shows the connection between the new computer science programmes at NWO. The programmes VIEW and GLANCE occupy the largest part of the I-Science programme cluster. The programme STARE, also part of the cluster, is linked to VIEW and GLANCE through research questions from astronomy in for example the area of visualization, or dealing with large datasets characteristic for astronomy. The programme STARE is however not limited to

[2] For details see http://www.nwo.nl/vernieuwingsimpuls

research in the area of Parallel and Distributed Computing and Modeling, Simulation and Visualization. Not part of the cluster, but also active in the area of PDC and MSV, is FOCUS, which was also mentioned in the previous section.

The programmes VIEW, GLANCE and STARE will be implemented parallel to each other, they share the same grant structure and communication plan. Their budgets and scientific contents however are clearly separated. The reason for this is that on the one hand the three programmes all contribute to the same aims of I-Science, and on top of that it is organizationally clearer to give the programmes the same layout. On the other hand, the research community in the Netherlands indicated a preference for separated budgets and contents in order to guard the reinforcement of each separate theme. The scientific content of each programme within I-Science will be explained in chapters 2, 3 and 4.

1.4 National and international developments

In the Netherlands, it is well-known that ICT developments have a large influence on both social and economical aspects of society. In this view, IT research was recently given an extra push through the Bsik grants. VL-e (Virtual Laboratory e-science), BRICKS, MultiMedian and LOFAR (LOw Frequency ARray) are examples of such Bsik programmes. The programme BRICKS stimulates basic research in computer science, and the scientific contents of BRICKS and I-Science are complementary.

The European Commission has made IST (Information Society Technology) the number one priority of the 6th framework programme. In Lisbon 2000 a strategy was adopted for an accelerated transition to a competitive and dynamic knowledge economy. This requires wider adoption, broader availability and an extension of IST applications and services in all economic and public sectors and in society as a whole. IST are the key underlying technologies for easier and efficient knowledge creation, sharing and exploitation.

An eye-catching recent international development in computer science is e-science. "*E-science is about global collaboration in key areas of science, and the next generation of infrastructure that will enable it*."[3] E-science involves many computer science disciplines, such as grid computing, programming, and data and visualization technologies. In the UK and USA large national e-science programmes have recently been deployed. Even though such a programme does not exist in the Netherlands, scientists and institutions from the Netherlands play an important role in the e-science developments. Through the I-science programme cluster we make a contribution to e-science, focusing on generic research in the area of large-scale parallel and distributed systems and on visualization. Complementary to these themes is the programme STARE, which focuses on challenging Computer Science research triggered by astronomical problems.

(Inter)national developments concerning the individual programmes will be highlighted in the following chapters.

[3] John Taylor, OST, Director General of the UK Research Councils

Chapter 2 – GLANCE: GLobAl computer scieNCE

Fundamental research on large-scale parallel and distributed systems

2.1 Background

In 1969, Leonard Kleinrock, one of the chief scientists of the original ARPA project which started the Internet, wrote: "*As of now, computer networks are still in their infancy, but as they grow up and become more sophisticated, we will probably see the spread of "computer utilities", which, like present electric and telephone utilities, will service individual homes and offices across the country.*" Despite major advances in both hardware and software, thirty-five years later this may be still as true as it was in 1969, for how far away are we still from computing as a utility? The GLANCE programme is designed to foster research of a fundamental nature in large-scale parallel and distributed systems, in order to make another step in this direction. The programme is developed in cooperation with NCF (National Computer Facility).

Parallel and distributed computing deal with the structure and the use of collections of network-connected processors and computers. For parallel computing, such a collection is typically homogeneous, has very fast interconnections, is managed by a single resource manager, and runs applications with a regular structure - in short, such a collection is a parallel system. For distributed computing, such a collection is typically heterogeneous with both fast and slow (and possibly wired and wireless) interconnections, is managed by the separate resource managers of the constituent parts, and is used for applications with components with different functionalities - in short, such a collection is a distributed system.

Over the last two decades, distributed computer systems have become the rule rather than the exception. With the proliferation of the Internet and the Web, which is one of the prime examples of a distributed system, it is probably very difficult to point at any two computers in the world that are not in some way connected and that cannot in some way communicate. Two more recent developments in distributed systems are grids for compute-intensive and data-intensive scientific and engineering applications, and overlay networks (peer-to-peer systems), which are so far mainly used for file exchange. Relatively homogeneous distributed systems made up of standard components are increasingly used for demanding parallel applications. However, customized parallel machines are still being designed and built to deal with the seemingly insatiable need for high-performance computing in science and engineering. The latter use custom interconnects and sometimes custom processors.

Two important characteristics of distributed systems are autonomy and cooperation. On the one hand, the constituent components of a distributed system should enjoy some level of autonomy in that they manage their own resources, on the other hand they have to cooperate to achieve common goals, to the point that they present themselves to their users and the applications as single entities. Autonomy is taken to its extreme in overlay networks, with all peers cooperating on a voluntary basis. The most important characteristic of parallel systems is efficiency. In particular, with the increasing use of clusters for parallel applications, the raison d'être for true parallel systems is to obtain a high performance, which scales with the number of processors used.

2.2 Aim

GLANCE aims at strengthening fundamental research on large-scale parallel and distributed systems in the Netherlands. Tremendous advances in computer hardware (processors, networks, disks) have led to parallel and distributed systems of unprecedented scales, as exemplified by IBM's Blue Gene with projected systems with up to 64K nodes, and on the Web with around a billion computers. It is this scale and the ensuing scalability issues that form the focus of the programme GLANCE. In addition, as there is already a substantial application-oriented research programme on grids in the Netherlands (e.g., in the Virtual Laboratories for e-Science programme), research in GLANCE should be of a fundamental nature. Thus, project proposals in GLANCE should address fundamental research questions in one or more of the following themes in large-scale parallel and distributed systems:

1. Management and analysis
2. Data acquisition and data processing
3. Autonomous configuration, composition, and coordination

Below is a description of each of these themes with some relevant research questions.

2.3 Research themes

Theme 1: Management and analysis

Managing large-scale parallel and distributed systems has long since been recognized as a major problem, but relatively little fundamental research has been devoted to it. Management includes various aspects, such as complexity, resource management, fault tolerance, and performance analysis, as described in more detail below.

* *Reducing the complexity of management*

A major problem in global-scale computing is the tremendous complexity of all the systems combined. Management and behavior analysis of both the system resources and the applications have become challenging tasks. Therefore, it is important to simplify these tasks for application developers and users alike. Self-managing systems are now recognized as a promising approach to reducing the complexity of management, control and analysis of large-scale systems. Self-managing systems can be broadly categorized as self-configuring, self-optimizing (performance), and self-healing (fault tolerance). Self-managing systems have to analyze the available resources and configure themselves to execute accordingly in a dynamic environment, enabling applications to run, optimize their performance, and adapt to changing conditions. For this purpose, they need appropriate mechanisms.

The users of large-scale systems need to be provided with adequate mechanisms and interfaces to analyze and control the behavior of their applications and identify and isolate error conditions. For this purpose, access via commodity devices ranging from PC’s to PDA’s and mobile phones is likely to become necessary, and so the management framework should be flexible and adaptive to the communication means employed by the user. When there are multiple options, this framework needs to set up the best possible connection to the user, selecting appropriate encryption mechanisms, adjusting the volume of management information that is transmitted (which may include compression), presenting the appropriate control mechanisms to the user, etc. Therefore, in addition to the self-management mechanisms of large-scale systems, dynamically adjusting the configuration of the control framework is an important second component of self-managing systems.

The software managing large-scale systems is increasingly based on open standards designed to support distributed environments, such as web services in combination with XML. These technologies are fairly new and the development of the appropriate management information models, as well as the realization of the protocols to support the framework described above, will be an important step. Part of the challenge is to manage the complexity of protocols that interoperate and have subtle cross-layer inter-dependencies. Especially in the face of adaptive configurations, the implications of the dependencies need to be explored.

* *Resource management*

Large-scale parallel and distributed systems potentially consisting of tens of thousands of hardware components and millions of software components have to be managed by resource managers. If a single resource manager would have to deal with this number of resources, it would be very complex, but if there are many different resource managers, their cooperation becomes complex. For instance, a single scheduler of a parallel machine with tens of thousands of nodes must be very sophisticated to make optimal use of the (expensive) system. The schedulers in grids made up of many separate subsystems may employ sophisticated negotiation schemes for balancing the load. But when some system’s capacity is sacrificed, mutual peer-to-peer negotiations may be a very simple paradigm for reaching resource management decisions.

In a system composed of independent administrative domains, resource usage policies induce additional complications. Such policies are generally dynamic in nature, and the actual policy and current system state may not be available in full to entities outside the resource's home organization. Resource metrics (based on self-discovery) that differentiate between policy for specific groups must be derived to enable inter-organizational scheduling - not only for compute resources, but for storage and bandwidth as well.

Research questions may be found in areas such as resource discovery (information dissemination, multilevel peer-to-peer systems), resource selection (negotiations, optimization, matching resource and application characteristics), and resource access (prefetching executables and files, adapting the environment to applications).

* *Fault tolerance*

Hardware and software faults will occur frequently in large-scale systems, and both resource managers and applications should be able to deal with this. Several techniques have been studied in the past to discover or handle faults, such as monitoring (discovering faults), replication (of files and computations), and checkpointing (storing the state of a distributed program or recording intermediate results). The scalability of many existing techniques, however, is poor. Also, it is unclear yet whether such techniques can be made transparent to the application (as several parallel systems have tried with limited success) or need to be driven by the application.

** Performance analysis*

Performance is of paramount importance and is one of the prime reasons for adopting parallel or distributed systems in the first place. The true challenge is in the scale of the systems. For instance, much previous research in load sharing in distributed systems was performed with simulations involving only tens of processors. In this theme there is ample room for all three classical methods of performance analysis, viz. measurements, simulations, and mathematical techniques.

Measurements are indispensable for understanding and explaining the behavior of large systems. For instance, our understanding of the actual behavior of the Internet and of P2P systems (self-similar behavior, heavy-tailed distributions, flash-crowd effects, power laws) and of the behavior of parallel applications in wide-area distributed systems has only come about from measurements. Many resource management mechanisms and policies are far too complicated to be amenable to mathematical analysis, but an experimental evaluation may not reveal their fundamental behavior due to "contaminating" details of actual systems. In such circumstances, simulations are called for. Mathematical techniques have proven useful for analyzing phenomena in large-scale systems. Examples are differential equations (for epidemic protocols), queuing theory and fluid models (for P2P systems), and statistics (for workload modeling).

Theme 2: Data acquisition and data processing

In many areas of science, industry, and business, huge data streams are generated and huge data sets are maintained and processed, which creates large requirements for bandwidth, storage, and processing capacity. Although in many applications these data sets have only a limited number of sources, increasingly, (massively) distributed data-acquisition systems are built for all kinds of scientific experiments and monitoring applications. For example, in the LOw Frequency ARray (LOFAR) astronomy project which is currently under development in the Netherlands, a radio telescope will be built consisting of tens of thousands of sensors spread across an area of hundreds of square kilometers. Because this system produces continuous data streams, there are also real-time constraints, and because of the sheer size of the data streams, a supercomputer is needed to process them.

Important problems in this area are data acquisition, reducing the data to manageable sizes, and distributing, replicating and storing the data.

** Data acquisition*

An important problem in large-scale distributed data acquisition systems which may consist of tens of thousands of sensors is fault tolerance and robustness. In addition to "ordinary" hardware and software failures, an aspect that is unique to data-acquisition systems is that the signal from a sensor may not simply fail in an on/off fashion, but rather can be corrupted by some noise source (e.g., by radio-frequency interference in the case of LOFAR). Although at some level of corruption it is better to discard the signal from the sensor in question altogether, there is a "grey area" in which the signal is still useful, albeit with a lower weight.

Logging the performance of each sensor in order to correct the results later may be impractical because of storage problems, but more importantly, because the type of processing involved makes later correction impossible. Therefore, data processing and management strategies must not only be robust against the complete failure of components, but must also be able to simultaneously deal with processing the performance-level data. Given the hierarchical nature of several (proposed) systems in this field, the performance data must be correctly transferred through all levels of the hierarchy for optimal quality of the final result. Ideally, such strategies for dealing with the performance data should be transparent to the applications, but it is an open question as to how this should be achieved. Although some forms of noise and other corruption problems might be modeled mathematically, simulation studies are probably needed in more general cases.

Another robustness issue in data acquisition systems is the heterogeneity of sensors in a single system. Although some systems are homogeneous regarding the sensor type and type of signal produced, others (e.g., for the global acquisition of meteorological data) have highly heterogeneous sensor types, which leads to problems of configuring and coordinating the system.

** Data reduction*

In large-scale distributed systems in which large data sets are generated and maintained, there is often a trade-off between processing capacity on the one hand and bandwidth and storage capacity on the other by reducing or compressing the data. For instance, the raw data streams from sensors may first be processed and combined (at or close to the sensors), leading to a reduction of their overall size, before being transmitted to the storage locations. Similarly, multimedia data may be stored in compressed formats, if they can be recreated for replay within certain time constraints. There are already various efforts in the area of data grids to tackle distributed storage problems, but few of these include time-constrained data acquisition and reduction from very large numbers of data sources. Although the data reduction and data management strategies will be application dependent, the development of more-or-less generic methods for their design and implementation, and studies of their scalability and performance are of great interest.

The hierarchical architecture of many large-scale systems creates challenges in areas of system architecture and bandwidth. Besides, hierarchical algorithms, in which there are parallelization problems that have not been solved yet, are crucial in this respect.

This applies both to computational aspects and I/O aspects: the available communication bandwidth may dictate the type of algorithms that usefully can be employed. Also, at present parallel filesystems are far from standardized and also need further research to meet the large data management requirements sketched above.

* *Data distribution and data analysis topologies*

The hierarchical architecture is not necessarily limited to systems in which data acquisition itself is distributed, and decoupling the analysis topology from the acquisition topology gives the option to share the workload over a larger number of organizations. Distributing the raw data, possibly in real-time, to many different locations also ensures that subsequent access to the data does not suffer from bottlenecks. At the same time it leaves the possibility to store far more raw data and thus incrementally improve on data reduction algorithms, or to evaluate multiple reduction algorithms - possibly by different independent organizations.

Large scale distribution or replication of data sets reveals the "missing link" between storage technology and high-bandwidth networking. Although both areas have experienced significant performance improvements, successfully storing large data volumes fast is still an open problem. With network bandwidths of 10 Gbit/s and more, storage systems can no longer cope with the data rate, and reliable novel distributed storage techniques are needed.

Theme 3: Autonomous configuration, composition, and coordination

The increasing power of processing and information resources available today is a trend that is likely to intensify in the foreseeable future. The decreasing costs and higher speeds of processors, now with chip-level concurrency, larger memories, and connections with higher bandwidth, provide more raw processing power and interconnection possibilities for global and ubiquitous computing than ever. This increases the significance of concurrency and its relevance in systems on both micro and macro scales. Modern systems will exploit concurrency at many different levels: at the instruction level and data level to exploit the massive on-chip concurrency of future microprocessors, and at the system level through process or component coordination.

Although the forms and the technical details of concurrency at micro and macro levels are very different, they now both share a common context of utilization in modern computing environments: massively concurrent, heterogeneous, distributed, dynamic systems. This gives rise to a number of non-trivial conceptual and technical issues other than what has already been addressed in the classical studies of concurrency. Before demanding new emerging applications in e-science and e-business can put the raw power of modern computing and network facilities to good practical use, challenging issues in the areas of (1) coordination, (2) composition, (3) deployment, and (4) distributed intelligence to support them must be resolved.

** Coordination models and languages*

In contrast to middleware software such as MPI and CORBA, coordination models and languages are meant to offer a systematic means to close the conceptual gap between the cooperation model of an application and the lower-level communication model used in its implementation. The inability to deal with the cooperation model of a concurrent application in an explicit form contributes to the difficulty of developing working concurrent applications that contain large numbers of active entities with non-trivial cooperation protocols. A complex protocol is typically not recognized as a "commodity" in its own right, because it is only implicit in the behavior of the rest of the concurrent software. This makes maintenance and modification of the cooperation protocols difficult, and their reuse next to impossible. Coordination languages are the linguistic counterpart of the ad hoc platforms which offer middleware support for software composition.

The popular shared data space coordination models offer a metaphor that directly matches only a particular class of applications, where entities share and exchange data through a common, public medium with one another. Catering for other classes of applications often involves implementing non-trivial communication and coordination protocols on top of shared data space models. Furthermore, realizations of such shared data space models, at least in their current implementations, do not scale well over large distributed systems. Challenging and open problems in this area include (1) non-shared-data-space models and languages, especially to accommodate point-to-point and multi-party interactions; (2) coordination by third parties (exogenous coordination); (3) support for mobility and dynamic reconfiguration; (4) models and languages that espouse both synchronous and asynchronous primitives as well as their combinations to construct coordination protocol; and (5) models and implementations that are scalable over large heterogeneous distributed platforms.

** Composition*

Component-based software engineering advocates the construction of software systems through the composition of coordinated autonomous components. Its benefits include software reuse, simpler and faster construction, enhanced reliability, and a dramatic reduction in the complexity of construction of provably correct critical systems, many of which involve real-time concerns. Effective and flexible component composition is still a challenge today, and real-time constraints may make it even more demanding.

In service-oriented computing, a web service can represent a unit of business that an organization exposes to other organizations on the World Wide Web. The recent Web Services standard governs how one defines, advertises, and uses web services. Composition of primitive web services into complex ones is a next challenge. Existing proposals for languages for service composition (also called choreography of web services) often lack foundations in theoretical computer science and possibilities to address composition and coordination from a more general perspective than business process and workflow applications only. Composition of software components or web services to build complex functionality or services is a research area that was recently recognized to be of great importance to the Internet (Grid and Web) research community.

Current models and standards for components and web services incorporate quite inflexible mechanisms for their composition, where each entity imposes severe restrictions on how it can be composed with others. A flexible composition paradigm wherein the same set of components or services can be composed in different ways to yield systems with different emergent behavior constitutes a prominent challenge in this area.

* *Software deployment*

Large distributed applications are increasingly constructed as complex systems by integrating heterogeneous collections of software components (programs, databases, web services, agents, etc.), which may have been constructed by different developers. Moreover, such components can change rapidly and independently, and may be added to a system while it remains in operation. As a consequence, it may be difficult to deploy and manage the whole system. Therefore, software deployment becomes a crucial step in the software life cycle, leading to new challenges involving languages, models, and systems for its support and automation. Software deployment involves all activities related to the release, the installation, the activation, the update, the deactivation, and the removal of components, as well as the configuration of the resources to satisfy the performance requirements of the whole system. In order to support deployment, concurrency must become the primary typing information, which must be quantifiable and must be available dynamically as some kind of meta-data to the entities that aggregate, coordinate, monitor, and reconfigure future distributed applications and systems. For the deployment of such distributed applications, (new versions of) software components and web services must be uploaded to (remote) resources, must be (automatically) installed, and must be configured for their tasks.

* *Distributed intelligence*

Along with the advantages of distributed components such as efficient construction of systems out of reusable third-party and off-the-shelf software, come challenges such as coping with changes mandated by the independent evolutions of a system's constituent components and protocols that happen at the discretion of their respective autonomous providers and custodians. Composing distributed components and services, therefore requires advanced automated processing to give semantic meaning to software components before they can be used as building blocks to realize new complex functionality. Distributed software components, web services, and resources are described by, or annotated with, metadata to give a well-defined meaning to each entity. This distributed metadata allows software agents, users, and programs to interoperate, dynamically discover and use resources, and extract knowledge to solve complex problems.

Open systems as envisioned by the Internet community need metadata descriptions of the entities in the system, and need to share a common formalism and rules for inferring new metadata and knowledge through ontologies. Services are needed for reasoning about and querying metadata and ontologies. The challenge is to provide the management functionality for (1) knowledge discovery and applications and systems management; (2) semantic modeling of application tasks and requirements, services, data sources, and computing platforms; and (3) self-configuration and autonomous management where entities manage themselves according to their observations and interactions with other entities. The required scalability and the complexity of systems calls for this management functionality to be autonomous and intelligent.

2.4 (Inter)national developments

The scientific content of GLANCE was designed in consideration of other existing research programmes in the Netherlands, such as SENTINELS, BRICKS and VL-e. The programme SENTINELS focuses on security in ICT, networks and information systems. For this reason, security, which is a very relevant theme for large-scale parallel and distributed systems, is not specifically addressed in GLANCE. One of the themes in the Bsik programme BRICKS is Parallel and Distributed Computing. This theme focuses on Security, identification and authentication, Quality of Service in communication networks, and Distributed systems for streaming multimedia. The scientific contents of the Parallel and Distributed Computing theme in BRICKS and GLANCE are designed to be complementary. The VL-e programme aims at creating an e-science environment and doing research on methodologies, while the content of GLANCE is of a much more fundamental nature.

According to the NOAG-i[1], central points of activity in the area of Parallel and Distributed Computing in 2001 were Delft University of Technology, Vrije Universiteit Amsterdam, Universiteit van Amsterdam and the University of Groningen. The NOAG-I shows that the Universities at Eindhoven and Twente as well as the CWI had promising groups in 2001, though somewhat smaller. In the last years there has been a growth at the technical universities (Delft, Eindhoven and Twente) in the area of networking/telematics. This growth can be related to the Freeband research programme. All named universities, as well as Leiden University and University Utrecht participate in the ASCI research school. High performance computing and large scale distributed information systems are two of the main target areas of ASCI. The DAS-1 and DAS-2 systems built by the ASCI research school and funded by NWO have obtained a high visibility in the international systems community, especially as DAS-1 (built in 1997) was one of the first systems of this type. Since 1997, research on large scale distributed systems and Grids has increased dramatically, all over the world. Many large scale platforms have been built for research on Grids (e.g. the US TeraGrid, UK Grid) and distributed systems (e.g. Planetlab).

At European level, within the IST priority, there is special attention to large scale parallel and distributed systems in the programme 'Global Computing'. Examples of European Networks of Excellence in the area of networking (distributed systems) are E-Next and EuroNGI. Besides the huge investments in grid technology in the USA (NSF is advising an investment of a billion dollars), much work related to the GLANCE programme is being done in the context of the NSF programme Distributed Systems.

Chapter 3 – VIEW: Visual Interactive Effective Worlds

3.1 Background

In 1965 Ivan Sutherland wrote in a visionary article, entitled "The Ultimate Display": *One must look at a display screen as a window through which one beholds a virtual world. The challenge to computer graphics is to make the picture in the window look real, sound real and the objects act real.* Forty years later an enormous amount of progress has been made. On each desktop PC realistic animations of scenes with hundreds of thousands of polygons can be shown, thanks to the developments in software and especially hardware.

The driving force for this success is simply that vision is our most important sense. The use of images and animations enables us to communicate with the computer via a highly sophisticated, parallel high-bandwidth connection, it pays of to exploit this route. The range of applications is large. The main categories are:

- Entertainment. Computer games are a mass market, computer generated imagery is used routinely for special effects and complete movies.
- Data visualization. Visualization is indispensable in science and engineering to understand the large amounts of measured and simulated data.
- Training and simulation. Simulators are a cost-effective means for training (for instance pilots), virtual reality simulations help to remedy phobia.
- Design. High quality renderings of new products are generated without making a prototype; the quality of new buildings can be judged with virtual walkthroughs.

Visualization concerns the interactive synthesis of images and animations in order to provide insight and experiences. It is a broad field, which covers a large number of aspects and integrates a variety of disciplines. Computer science is central, but disciplines such as mathematics, physics, and psychology provide foundations. More narrow, computer graphics is the central discipline, but a variety of other disciplines are strongly related, such as human-computer interaction, image processing, multi-media, and computer aided geometric design.

It might seem that the aim of Sutherland has already been reached. However, there is a need for more fundamental insight. Many of the current methods and techniques are *ad hoc.* The development of new applications is often a matter of trial and error, and the production of content is hard labor. The insight which methods under which circumstances are effective and efficient is limited. Furthermore, a number of problems are unsolved.

3.2 Aim

The aim of VIEW is to stimulate research on visualization, including modeling and simulation to support this. The current base in the Netherlands is narrow. Five relatively small research groups in the Netherlands focus explicitly on visualization and/or computer graphics, about fifteen other groups carry out research in related fields. VIEW aims at strengthening the visualization community in the Netherlands, starting from the current interests and strengths, and setting out new challenges. These challenges are characterized by the keywords *generic* and *effective*, as main themes *Interactive Data Visualization* and *Interactive Virtual Worlds* are adopted.

VIEW asks for *generic* methods and techniques. New paradigms are requested that integrate and unify so far unrelated ad hoc methods, that solve open problems, and open up new application domains. Solid foundations on theory, probably from disciplines such as mathematics or psychophysics, are needed, which are used to derive insights and algorithms that are fruitful for many applications. Typically, a model based approach is used. A model serves two purposes here. It is a compact description that provides insight into a class of problems, and also it provides a base for the automatic generation of solutions.

VIEW explicitly asks for *effective* methods. Specifically, it has to be shown that the application of new concepts solves real-world problems, and leads to significant improvements in performance. A variety of measures can be employed here, such as simply the speed of image generation and the size of geometric models that can be handled, but also for instance the quality of the generated experience or the ease with which users can complete their tasks. Measuring such aspects is a difficult problem on its own, the development of methodologies for evaluating and comparing the effectiveness of existing and new methods is needed.

There is a delicate interplay between genericity and effectiveness. To obtain an optimal result for a specific application, the shortest path is to develop a custom solution. Also, to evaluate the ultimate effectiveness, the performance of real users using real applications has to be measured. However, the spin-off of such efforts is limited to narrow use cases and scenarios, generic methods have the potential to bring a major step forward from which many applications can benefit. How to resolve this? VIEW challenges the research community to come up with proposals to shed more light on this.

Research projects must address one of the two themes: *Interactive Data Visualization* and *Interactive Virtual Worlds*. These are elaborated in more detail in the next sections, where a description is given and a number of challenges are raised. These challenges should be viewed as inspiration, rather than as an exclusive list.

The type of problems addressed differs per theme, but there is also a considerable overlap. As expressed in their titles, *interaction* is a vital element in both themes, for instance for navigation through space, selection of objects and settings of attributes. There is a strong need for generic models and insight in the effectiveness of solutions for user interaction. *Realism* is also an issue in both, although a more controversial one. What contributes most to realism? How can we measure it? And also, for what applications do we need it, and when is a non-realistic presentation more effective?
VIEW focuses mainly on visual communication, but also audio and haptic feedback can be employed. Research proposals using these modalities are acceptable, provided they address one of the themes and that these modalities are used in combination with visual communication.

3.3 Research themes

Theme 1: Interactive Data Visualization

Data visualization concerns the use of interactive computer graphics to obtain insight in data. In many occupations professionals are confronted with enormous amounts of data. These data come from a variety of sources. Modern measurement equipment, such as used in medical imaging, satellite imaging, high energy physics, microscopy provide large streams of 2D and 3D time dependent data-sets. The same goes for supercomputer simulations, for instance for climate and biochemical research. As an example, a grand-challenge climate modeling project created 8 Terabytes of time dependent 3D data. Furthermore, in a field like genomics large existing distributed databases are used. Standard 2D visualization methods for the interpretation of these data do not scale properly.

The field of data visualization can be subdivided in several ways. A common subdivision is in *scientific visualization* and *information visualization,* where the former is concerned with spatial, continuous data, and the latter with abstract, discrete data. The names of these fields are somewhat misleading, but they have become standard. Other subdivisions are according to the type of data concerned (e.g. vector field visualization and graph visualization), or the application area (e.g. medical visualization and flow visualization).

The central problem of visualization is *how to visualize large data sets.* The following challenges can be identified.

* *Large data sets*

Standard algorithms, such as for instance for iso-surfacing, fall short when really large data sets have to be handled. Both the extraction and also the display of millions of triangles are too time consuming. Much work has already been done in those areas, but more efficient methods are still welcome.

* *Automated analysis and visualization*

Statistics and also other fields provide a wide variety of methods and techniques to analyze and reduce large data sets, such as clustering methods and principal component analysis. Most of these methods make assumptions on the underlying data, whereas in visualization it is up to the user to identify patterns and trends. How can we take advantage of such methods, for instance by simultaneous visualization of the raw data and condensed results?

* *Feature based visualization*

One solution to the central problem is simply showing only the interesting features. This is easier said than done. For many application domains it is unclear how to define what the features of interest are, and also, how to define in such a precise way that they can be detected automatically. Once those features are found, the next question is how to show them and their attributes such that the viewer can understand them effortlessly.

** Perception based visualization*

The research on the human visual system is vast, yet visualization researchers rarely study or apply what is known about the visual system when designing visualization techniques. There is much to be gained by integration and translation of the knowledge on perception knowledge into the visualization pipeline.

** Integration of SciVis and InfoVis*

Traditionally, Scientific Visualization and Information Visualization are seen as disjoint but this is artificial. Discrete datasets can be converted into continuous fields, and vice versa. One can expect this to be beneficial, but there is no methodology or model available for this.

** Automated visualization*

The production of an effective visualization requires often many choices and settings of parameters. For an optimal result, manual tuning gives the best results, but for efficiency reasons models and algorithms for automatic settings would be preferable. Also from a more theoretical point of view this is a challenge, since it requires a precise answer to the question what a good visualization is.

** Multifield data*

Much effort has been spent in visualizing scalar and vector volumetric data. However, nowadays measurement equipment often produce data per voxel in different wavelengths, in chemical simulations the concentrations of many different substances are calculated. And even worse, in radiation simulations per voxel distributions of intensities are generated. It is fully unclear how to get insight in such data sets, except by selection and visualization of these components one by one.

** Multi-modal visualization*

Virtual reality techniques have made it possible to experience abstract information spaces as 3D objects, sounds and forces in virtual worlds. The use of multiple modalities, as well as the use of another modality than the obvious one for a certain phenomenon can lead to more effective visualizations. Psychology can give insights into the characteristics and interaction between human sensory modalities and on how to adapt presentations optimally to our cognitive skills.

** A theory of visualization*

The field of Visualization consists of a large number of separate ideas, methods, and techniques. Can overall theories be defined? And do these lead to useful results?

Theme 2: Interactive Virtual Worlds

The simplest form of computer graphics is to make a picture of a single, static object. Interactive Virtual Worlds can be considered as the extreme opposite. The users are immersed in a complex scene with a large number of dynamic objects, which can be explored and interacted with. To achieve full immersion, *virtual reality* techniques such as head mounted displays and head tracking have to be employed, but, as illustrated by the popularity of computer games, also on desktop systems virtual worlds can be highly attractive and compelling. Other applications are design, scientific visualization, training, and virtual communities.

Traditionally, much effort was spent to meet the high frame rates required for immersion and the minimization of lag. For large scenes, these are still relevant issues, but there are many other open issues as well. The following gives a number of examples.

* *Manipulating a Virtual World*

In virtual worlds users should be enabled to manipulate objects. For example they should be enabled to open doors, pickup objects, deform them, etc. Such manipulation should look realistic and the system should assist the user in these tasks, for example by automatically computing the correct way of holding an object and the corresponding internal motions of the avatar executing the task.

* *Effective Navigation in Virtual Worlds*

When inspecting a virtual world the user must navigate through it. Current practice leaves this under full control of the user, using e.g. a joystick or the keyboard. Such navigation is difficult, leads to ugly motions and takes the attention of the user away from the tasks to be performed. New navigation methods are required to assist the user in moving through the environment with minimal effort. This also involves tools to inform the user about where he is and how to get to other places of interest.

* *Level of detail in interaction*

Humans, when interacting with each other and with their surroundings, adjust their interaction according to the availability of communication channels and manipulation tools, the characteristics of their interlocutor and the overall goal of the interaction. Present-day virtual reality systems do not allow for this flexibility. It is a challenge to provide a general model of interaction, where communication functions can be realized by different, computer-related and/or human-specific means, allowing different levels of detail of communication.

* *Individual and adaptive embodied conversational agents*

The individualism of embodied agents has become a hot topic. Particular demonstrators have been around to show individualism in motion characteristics, or verbal or non-verbal style. However, it is still a challenge to develop synthetic characters that are individual along all relevant dimensions (outer look, gesturing, speech, dialog strategy,...) in a consistent way, and which adapt to the characteristics of the human interlocutor.

* *Affective Appraisal of Virtual Worlds*

Virtual worlds are often used for inspecting real-life locations without the need of being present. E.g. historical sites, models of new buildings, plans for changing landscapes. It is important that the virtual model gives a similar experience as the real location. The affective appraisal of virtual worlds is heavily influences by the modeling, lighting, sound, and behavior in the world. The challenge is to create the virtual worlds in such a way that the affective appraisal matches that in the real world.

* *Shared Virtual Worlds*

Given multimodal records of a world from many different perspectives, how can we synthesize and integrate these into a global virtual model to be shared in collaborative settings, spanning both time and space? Application domains vary from the virtualization of archived personal experiences to the effective operation of remote crisis management teams.

* *Realistic Motion in Virtual Worlds*

Now that virtual worlds as used, e.g., in games have become visually very realistic, users also expect the entities in these worlds to behave in realistic ways. One basic operation for entities is that they have to move in a natural way between places. This involves questions like: what is a natural route through an environment, how to incorporate constraints (e.g. of cars), how to combine internal body motion (animation) with external navigation, and how groups of entities behave (e.g. crowd modeling).

* *Real-time Physical Simulation*

In virtual worlds physical processes, like falling and bouncing blocks, floating objects in water, car behavior, etc. need to be modeled and computed in real time. Standard physical simulation techniques are too slow for this. The goal though is not accurate behavior but realistically looking behavior (compare the lighting models used in graphics that are not based on the actual physical phenomena). This requires new ways of modeling such processes and efficient algorithms to compute them.

* *Modeling Worlds*

The definition of accurate and detailed geometric models of real-world scenes and objects remains tedious and costly work, despite all methods that have been developed and tools that are available. Directly sampling the real world, via laser scanners or by image based techniques, is an attractive alternative, but still requires often much manual effort and is by not suitable for imaginary worlds.

** Rendering*

Realistic rendering of large models, with different materials under various lighting conditions is a classic topic in computer graphics. Virtual worlds pose two extra challenges. A real-time performance is required, and also, dynamic scenes are often more critical to render realistically, because short-cuts and simplifications that are fine for still images become apparent when viewed dynamically.

3.4 (Inter)national developments

As stated before, the current research base concerned with visualization in the Netherlands is narrow. The research groups focusing explicitly on visualization and/or computer graphics are at Delft University of Technology, CWI, University Utrecht, Technische Universiteit Eindhoven and University of Groningen. According to the NOAG-i[1], there are a few other groups carrying out research in related fields, such as University of Twente, Vrije Universiteit Amsterdam, Universiteit van Amsterdam and University of Leiden. VIEW aims at strengthening the visualization research community in the Netherlands.

One of the themes in the Bsik programme BRICKS is Modeling, Simulation and Visualization. The largest part of this theme is occupied by the sub-theme Scientific Computing, which is not addressed in VIEW. A smaller sub-theme is Interactive Virtual Environments, which is closely related to the themes covered in VIEW. The researchers involved in this sub-theme will be involved in the community of VIEW.

The need for more research in visualization has been recognized in other countries also. The NSF supports many different visualization projects. Within the Division of shared Cyber Infrastructure, one of the four divisions within NSF's Computer and Information Sciences programme, interactive visualization techniques take a prominent position. On European scale, within kp6, subjects like visualization, rendering, interactivity, graphics and mixed reality come up within the programme Networked Audio-Visual Systems, Home Platforms for Rich Media Communication.

Chapter 4 - STARE

4.1 Background

One of the biggest challenges in science these days is to let scientific progress keep up with the ever increasing data streams. In astronomy, the amount of data collected is doubled each year, due to ongoing advances in instrumentation. The raw datasets are now reaching sizes in the order of petabytes. Advanced computer science tools and techniques are required in order to transfer, manage, analyze and visualize this information. To enhance rapid developments, a close cooperation between astronomers and computer scientists is required. The interaction between astronomy and computer science can bring both science disciplines to a higher level. While astronomy can benefit from computer science techniques to get the maximum out of the data, the computer science discipline is challenged by the involved complicated questions from astronomy. The programme STARE is designed in order to stimulate this interaction.

VIEW and GLANCE, the other two programmes in I-Science that are described in the previous chapters, respectively aim at computer science in the area of visualization and large-scale parallel and distributed computing, by which they contribute to the foundation for e-science (enhanced science). STARE aims at innovative computer science that can enhance astronomy, meaning the research in STARE will not only contribute to the foundation for e-science, but will actually BE e-science. The research to be performed in STARE may very well be (but is not necessarily) related to the work that is being done in VIEW and in GLANCE. For the successor of the NOAG-i (see chapter 1), the NOAG-ict, nine appealing research themes were defined. STARE primarily acts in the NOAG-ict theme 'The data-explosion'[4], but also the themes 'The virtual laboratory' (e.g. distributed computing and grids), 'The networked world' (e.g. sensor and P2P networks) and 'Methods for design and building' (e.g. real-time, dependability, system design) play a role.

4.2 Aim

STARE aims at stimulating innovative research at the interface of computer science and astronomy. It does so by facilitating cooperation between computer scientists and astronomers. Research proposals can be submitted by teams consisting of (at least) one computer scientist and (at least) one astronomer. The programme is explicitly meant to support high quality research with surplus value for *both* research disciplines. This relevance for both disciplines is the essence of the programme, and is a main criterion for judging the programmatic fit of proposals with the programme. Computer science should not be confused with computational science, which is not a focus of this programme. Below some relevant research themes are given.

---

[4] See the IPN website, www.informaticaplatform.nl, for more information on the NOAG-ict themes.

4.3 Research themes

Theme 1: Technology development

** Data handling*

The first step after measuring astronomical data is to reduce the usually broad bandwidth data to manageable sizes and formats. This involves for instance calibration, filtering, noise reduction, quality control, the selection between which raw data to store and which to discard or the specific data handling in case of transient phenomena. There is a need to develop techniques for automated parallellization.

** Data storage*

Observational astronomy yields huge data streams which have to be stored in databases for longer periods of time after filtering and calibration. However, storage alone is just the first step. Model- and proximity-based query processing techniques, inspired by the astronomer's vocabulary, should be designed. There is a strong need for the further development of alternative query languages (e.g. array-comprehension-, and functional-based). Another major challenge for both computer science and astronomy is data lineage. Data lineage deals with the problem to properly describe, monitor and manage the derivation process of data. In principle, it should be possible to partly re-calculate derived information if one of the earlier steps in the data lineage path has been changed. More examples are the consistency of distributed datasets, connecting different catalogues by associated entries, database speed, the development and application of smart databases, and the developments of standards for astronomical data.

** Distributed computing and grids*

Astronomy is one of the most important applications to make use of innovative distributed computing and grid technology. This is illustrated by developments such as the Virtual Observatory and distributed telescopes such as JIVE and (in the future) SKA. Challenges lie for example in the area of scalability; How can we set up a networking software/hardware infrastructure that can that efficiently transport the tera,- (or peta)bytes of (streaming) astronomical data. How do we scale existing grid technology to this kind of data? How do we guarantee Quality of Service for these kind of operations? How does the VO connect to the grid? How does the end user get access by means of VO-like interfaces to processing power to solve his observational-research problem? How do astronomers based in one University connect to and make efficient use of scalable resources at other nodes?
Other related questions are whether we can simulate and test large distributed systems such as LOFAR on its full scale using grid technology. Can we make sure the system is robust enough to deal with, for example, failing sensors? How do we correlate data from distributed data streams? And how to solve the I/O problem, which is still the bottleneck when large data streams are transported? Web services can for example be useful for making visualizations of different databases and combining them.

Theme 2: Interactive data analysis and interpretation

The aim of measuring, transporting and storing astronomical data is to enable researchers to interpret what has been measured. Visualization, data mining and the virtual observatory enable us to access and understand the data.

** Visualization*

An important step in the data processing sequence is visualization. The main concern for the visualization of astronomical data is the development of automatic methods that can handle the enormous amounts of data and select the relevant parts of the data to be visualized. At the same time, these automated methods have to leave room for human interaction, since the analytical power of the human brain is still indispensable for a correct interpretation of the data.
Challenges lie in the area of automated feature extraction (from noisy data) and pattern recognition, visualizing 3D data cubes or high-dimensional parameter spaces, efficiency with respect to speed and memory of existing algorithms for very large data sets, integrating filtering and visualization in a problem solving environment, the tension between automation and human interaction, visualization of databases, and the intensive interaction between the database management system and the visualization tool. Also, software parallellization and graphics hardware acceleration are important issues to consider.

** Data mining*

In order to optimally exploit the huge amounts of data of different frequencies and modalities that are stored in different formats, at distributed locations, very advanced data mining techniques are called for. How can we trace what's relevant from these huge amounts of (visualizable) data?. It should be possible to search through different databases efficiently to find possible relations along several dimensions. An intuition based model is often the starting point in astronomy and demands fast and flexible access to experimental data for validation of the model. Large-scale data mining techniques are essential here. A related issue is data mining in numerical models. If something has been calculated before, there is no need to calculate it again, but this is only useful when scientists can find what has been calculated and have access to the results. The storage, query processing and data mining techniques inherently call for a parallel and distributed solution. Both at the level of system software and data exchange, e.g. P2P database technology, and in methods to automate integration of heterogeneous sources. It will become commonplace that information from different sources should seamlessly be combined to reach new scientific insights.

** Virtual Observatory*

In the Virtual Observatory (VO), all astronomy data and literature will be organized into a coherent system that can be accessed by anyone, in any form, from anywhere[5]. The VO

[5] *The world-wide telescope*, A. Szalay, J. Gray, Science, V.293, pp. 2037-2040, 2001

consists of a collection of data centers, each with unique collections of astronomical data, software systems and processing capabilities. Storage, services and processing are connected in the Virtual Observatory. Through a VO, scientists can get access to computing power and storage capabilities that exceed by far the possibilities at their local institution. Furthermore, by combining data from different surveys (e.g. with spectra at different frequencies) a deeper and better insight into the observations can be reached. Processing grid technology will in the future be an essential backbone for the virtual observatory. At the moment, the European partners gathered in the EURO-VO project are building the infrastructure for connecting archives of data-centers, allowing astronomers to publish data on VO, which in turn can be searched and datamined by the community using services that are also provided by the data providers. Obviously, to do this for all instruments and archives is an immense task. Challenges for VO-like STARE projects can be found in new innovative combinations of dedicated web-services, distributed processing, distributed archiving and federated databasing (essentially any of the above technology and interactive data analysis issues) which will help to build the VO and make it useful and accessible for astronomical research with the new generation of instruments.

The legacy problem (can we still use old software and systems for new instrumentation and developments?) is relevant for each of the previous themes.

4.4 (Inter)national developments

Many current (international) astronomical projects in which Dutch researchers are involved call for advanced computer science methods in order to handle the enormous amounts of data. Examples are ALLEGRO (ALMA Local Expertise Group, ALMA=Atacama Large Millimeter Array), OmegaCEN (supporting data center for the OmegaCAM wide-field camera), JIVE (Joint Institute for VLBI in Europe), LOFAR (LOw Frequency ARray) and 3D-radiation-hydrodynamics. JIVE is working on the e-VLBI project, in which telescopes are linked electronically in real-time, in order to enable astronomers to analyze the data as the observations are being made. Another example is LOFAR (LOw Frequency Array), a M€ 150 ICT project which is under development in the Netherlands right now. LOFAR will consist of thousands of radio sensors, an ultra-fast terabit data network and a central supercomputer, making it the largest radio telescope in the world. LOFAR is developed by ASTRON (the Netherlands foundation for research in astronomy). Another innovative instrument under development is SKA (Square Kilometer Array), which will make measurements through groups of antennas that will be distributed over different continents.

Relevant large projects in ICT research in the Netherlands are for instance VL-e and GigaPort. VL-e is the Virtual Laboratory project, that carries out research along the complete e-Science technology chain. In GigaPort, SURFnet6 is developed, an innovative network that combines a traditional network with optical network services.

The Virtual Observatory is going through rapid developments all over the world. Examples are UK's AstroGrid, the Astrophysical Virtual Observatory (AVO), funded by the European Commission, and the National Virtual Observatory in the USA (NVO, funded by NSF). Many other countries, like Russia, Australia and Germany are working on a Virtual Observatory as well.

Chapter 5 – Organization

5.1 Programme Management

GLANCE and VIEW have their own programme committee. The programme committee is responsible for the programme and its scientific coordination, and consists of experts from different research organizations and universities. The NWO Physical Science Council provides the secretary of the programme committee. Tasks of the programme committee are, among others, determining the assessment criteria, reporting on the progress of the programme and stimulating the knowledge transfer. For the members of the committees see our website: www.nwo.nl/I-Science.

5.2 Budget

The budget of I-Science is M€ 10,5, of which M€ 4,5 is allocated to VIEW, M€ 4 is allocated to GLANCE, and M€ 2 is allocated to STARE. By this budget, a total of 21 projects can be subsidized. A small part of the budget is reserved for programme activities such as the organization of symposia, workshops and other activities stimulating knowledge transfer. This part will be administered by the NWO Physical Sciences Council. Projects within the same research theme as I-Science that are granted in the Open Competition or in FOCUS will be involved in the programme activities.

The budget for VIEW and GLANCE will be allocated through two application rounds, separated by one year. The competition round for STARE will run parallel to the Open Competition round of 2005.

5.3 What can be applied for?

I-Science will subsidize relatively large projects, consisting of about 3-4 researchers. Each grant consists of k€ 500 and can be used for:

- PhD positions (four years ~165 k€, including bench fee);
- Post-doc positions (two years ~k€ 115 or three years ~k€ 171, including bench fee);
- Assistant/associate professors;
- Technical personnel and programmers;
- Instrumentation worth up to k€ 110;
- Travel costs for congress visits other than included in the bench fee;

See the call for proposals for details. Small projects within the same research theme as described in chapters 2,3 and 4 can be submitted in the Open Competition. Instrumentation worth over k€ 110 can be applied for through NWO-M (see www.nwo.nl/ew).

5.4 Who can apply?

Applications can be submitted by researchers working at Dutch universities, NWO- and KNAW institutes. A proposal can be submitted by one researcher (the main applicant) representing one research group or a consortium of research groups.

For STARE, an extra requirement is that applications can only be submitted by teams of researchers consisting of at least one astronomer and one computer scientist.

Chapter 6 – Assessment and schedule

6.1 Assessment

The proposal must concern research as described in chapter 2, 3 and 4 and must answer the formal requirements as described in sections 5.3 and 5.4. The programme committee will verify whether the proposal meets these conditions and, if it does, will consider the proposal admissible. Next, the proposal will be assessed for its scientific quality. The programme committee will determine its exact assessment strategy anticipating on the possibility that members of the programme committee might be involved in one of the proposals. The assessment strategy will be designed such that an objective assessment of the proposals is guaranteed. The proposals in STARE will be assessed by the assessment committees for computer science and astronomy in the Open Competition. The GBE will make the final decision on the allocation of the grants.

6.2 Schedule

The time schedule up to allocation of the grants in the first round of GLANCE and VIEW is as follows:

Activity	Date
Call for proposals	November 2004
Deadline for proposal submission	February 2 2005
Assessment	March/April/May 2005
Decision on allocation of the grants	June 22, 2005

A similar time schedule will be designed for the second application round, which will most likely have its deadline for proposal submission in February 2006.

The competition round for STARE will run parallel to the Open Competition round of 2005. The time schedule up to allocation of the grants in STARE is as follows:

Activity	Date
Call for proposals	June 2005
Deadline for proposal submission	September 2005
Assessment	November/December 2005
Decision on allocation of the grants	February 2006

More information on I-Science can be acquired from dr. ir. H. Poot (070-3440915; poot@nwo.nl) at the NWO physical sciences council or on our website, www.nwo.nl/I-Science.

Abbreviations

BRICKS	Basic Research in Informatics for Creating the Knowledge Society
Bsik	Besluit subsidies investeringen kennisinfrastructuur (Decision on investments in knowledge infrastructure)
CWI	Centrum voor Wiskunde en Informatica (National research institute for Mathematics and Computer Science)
FOCUS	reinFOrcing CompUter Science
GBE	Physical Sciences Council (GebiedsBestuur Exacte Wetenschappen)
GLANCE	GLobAl computer scieNCE
LOFAR	LOw Frequency ARray
MSV	Modeling, Simulation and Visualization
NCF	Nationale Computer Faciliteit (National Computer Facility)
NOAG-i	Nationale Onderzoeksagenda informatica (National Research agenda for computer science)
NWO	Nederlandse Organisatie voor Wetenschappelijk Onderzoek (Dutch Organisation for Scientific Research)
PDC	Parallel and Distributed Computing
STARE	Star E-science
VIEW	Visual Interactive Effective Worlds
VL-e	Virtual Laboratory e-science