INTRODUCTION
64 CEO welcoming remarks
65 Taming the Big-Data Beast - Together
66 eScience Engineers = Digital Scientists able to bridge Science and ICT
67 eScience Integrators = Exploiting the transformative potential of eScience

INTERVIEW
68 Amandus Lundqvist and Jos Engelen: The Time is Ripe for eScience
71 Barend Mons: eScience – Beyond Individual Knowledge
74 Marco de Vos: To the stars and beyond with eScience
77 Louis Vertegaal: Big-Data: a desert of information
80 Frank Seinstra: The new scientific paradigm.

SCIENTIFIC CASE STUDIES
82 Introduction
85 Enhanced Science = Big-Data for the Big Bang
86 Enhanced Science = Understanding the effects of Climate Change
88 Enhanced Science = A deeper understanding of living systems
90 Enhanced Science = Turning observation into knowledge
92 Enhanced Science = Unlocking our Hidden History
94 Enhanced Science = Managing our Place in the World
96 Enhanced Science = Transferring medical models into clinical care
98 Enhanced Science = Uncovering hidden knowledge
100 Enhanced Science = Putting Big-Data to Work

PATH FINDING PROJECTS
104 Introduction
108 Enhanced Science = Insight into the Origin of the Universe
110 Enhanced Science = A deeper understanding of Murray
112 Enhanced Science = Techniques to understand complex systems

TECHNOLOGY DEVELOPMENT
116 Introduction
120 Enhanced Science = Methods and Tools for Optimised Scientific Discovery
122 Enhanced Science = Tools to exploit the potential of Big-Data

PORTFOLIO & PARTNERSHIPS
126 New Projects
128 Project Network
129 Project Calls 2013
131 The Collaboratorium
133 Organisation of the Netherlands eScience Center
136 Data Stewardship
INTRODUCTION

the Netherlands eScience Center

It has been less than two years since we launched the Netherlands eScience Center (NLeSC) and only a year since the foundation of its own institute. In those first twelve months, the NLeSC has achieved a great deal, building valuable partnerships and contributors from across the globe. The NLeSC is present to ensure a well-organised and resourced environment that will work hard to develop national and international visibility by undertaking scientific research using advanced research computing and networking.

All NLeSC researchers are engaged in collaborative and interdisciplinary scientific research. NLeSC researchers focus on developing and evaluating scientific software and eScience components that support the implementation of scientific processes, while working closely with other teams and institutions. The NLeSC team consists of digital scientists, mostly PhDs, engineers, and a dedicated team of broadly-cross-disciplinary researchers. NLeSC stimulates creative data analysis possible across multiple scientific disciplines. NLeSC also supports the continuing advancement of the scientific research teams, with the aim to change scientific practice by making large-scale collaborative research. It has been less than two years since we launched the NLeSC, and we are looking forward to extending this team in the following years.

The NLeSC team consists of the following disciplines: Life Sciences, Chemistry & Materials, Social Sciences & Humanities, Earth Sciences, and Advanced Cyber Infrastructures. NLeSC has developed a multidisciplinary research team to explore the potential of eScience engineering, through the creation of eScience research teams, in collaboration with the scientific disciplines. NLeSC has recruited a team of eScience Engineers, who are dedicated to developing and implementing eScience solutions for the scientific community. The NLeSC is committed to supporting the continued advancement and professional development of its eScience research team, with the aim to build a strong research community.

The NLeSC is committed to supporting the continued advancement and professional development of its eScience research teams, while working closely with other teams and institutions. NLeSC is developing a multidisciplinary research team to explore the potential of eScience engineering, through the creation of eScience research teams, in collaboration with the scientific disciplines. NLeSC has recruited a team of eScience Engineers, who are dedicated to developing and implementing eScience solutions for the scientific community. The NLeSC is committed to supporting the continued advancement and professional development of its eScience research team, with the aim to build a strong research community.

The NLeSC is committed to supporting the continued advancement and professional development of its eScience research teams, while working closely with other teams and institutions. NLeSC is developing a multidisciplinary research team to explore the potential of eScience engineering, through the creation of eScience research teams, in collaboration with the scientific disciplines. NLeSC has recruited a team of eScience Engineers, who are dedicated to developing and implementing eScience solutions for the scientific community. The NLeSC is committed to supporting the continued advancement and professional development of its eScience research team, with the aim to build a strong research community.

The NLeSC is committed to supporting the continued advancement and professional development of its eScience research teams, while working closely with other teams and institutions. NLeSC is developing a multidisciplinary research team to explore the potential of eScience engineering, through the creation of eScience research teams, in collaboration with the scientific disciplines. NLeSC has recruited a team of eScience Engineers, who are dedicated to developing and implementing eScience solutions for the scientific community. The NLeSC is committed to supporting the continued advancement and professional development of its eScience research team, with the aim to build a strong research community.

The NLeSC is committed to supporting the continued advancement and professional development of its eScience research teams, while working closely with other teams and institutions. NLeSC is developing a multidisciplinary research team to explore the potential of eScience engineering, through the creation of eScience research teams, in collaboration with the scientific disciplines. NLeSC has recruited a team of eScience Engineers, who are dedicated to developing and implementing eScience solutions for the scientific community. The NLeSC is committed to supporting the continued advancement and professional development of its eScience research team, with the aim to build a strong research community.
EXCISE – ENHANCED SCIENCE

New technologies, both natural and science-based, offer the promise of new opportunities to exploit this resource, leading many to describe the current landscape, from academia to private industry, as the ‘data-driven’ society. In this landscape, the value of data is recognized, and research not only relies on accurate and up-to-date data, but also on the ability to access, analyze and visualize data in order to generate new ideas and insights.

Researchers today are looking for innovative ways to transform data into knowledge and new insights into action. Data science, the science of data, is gaining prominence as a field of study and a discipline in its own right. This trend is not limited to academic institutions but is also driving innovation in the private sector.

EXCISE ENSURE A ‘DATA-DRIVEN’ MINDSET

The Netherlands eScience Center (NLeSC) was launched in 2011 as a unique collaboration between the principle Dutch Scientific and Academic institutions: ANW, NWO, Dutch University and NWO, Dutch higher education and research partnership for ICT. The center was instigated in a direct response to the need for a sustainable and coherent e-infrastructure across all scientific disciplines.

AN EXPERT CENTER FOR BIG DATA-SCIENCE

The advent of big data is changing the way we do research. The current scale of data is changing our understanding of big data. The increase in data volume, velocity, variety, veracity, and value has led to a new way of doing science. In this landscape, the value of data is recognized, and research not only relies on accurate and up-to-date data, but also on the ability to access, analyze and visualize data in order to generate new ideas and insights.

New technologies, both natural and science-based, offer the promise of new opportunities to exploit this resource, leading many to describe the current landscape, from academia to private industry, as the ‘data-driven’ society. In this landscape, the value of data is recognized, and research not only relies on accurate and up-to-date data, but also on the ability to access, analyze and visualize data in order to generate new ideas and insights.

Researchers today are looking for innovative ways to transform data into knowledge and new insights into action. Data science, the science of data, is gaining prominence as a field of study and a discipline in its own right. This trend is not limited to academic institutions but is also driving innovation in the private sector.

EXCISE ENGINEERS

The eScience engineer concept has been developed in response to the operational needs of NLeSC and reflects the growing requirement for data scientists throughout the economy. It is already widely recognized throughout science that data-driven approaches are making an impact in a wide range of disciplines. As scientists noise that optimising the digital landscape is at the interface of computing, sciences that we can provide the best of both worlds. We can seamlessly integrate data scientists with limited informatics experience and ICT professionals without specific scientific expertise require access to a virtual team of experts from industry, academic and commercial partners to translate the imaginative combination of existing data management, data analysis and ICT to improve experimental design, communication.

The ultimate goal of eScience is to enable data-intensive science and sustainability, Environment & Technology. New, NLeSC has the capability to collaborate with the growing requirement for data and ICT to improve experimental design, communication.

As a problem driven organisation, NLeSC has prioritised the following five domains; Chemistry & Materials, Life & Agriculture, Physical Sciences & Technology, Information & Communication and Sustainability & Environment. Various partners collaborating in public-private partnerships to bridge these two worlds.

As scientists noise that optimising the digital landscape is at the interface of disciplines by bringing together scientists with limited informatics experience and ICT professionals without specific scientific expertise require access to a virtual team of experts from industry, academic and commercial partners to translate the imaginative combination of existing data management, data analysis and ICT to improve experimental design, communication.

EXCISE ENGINEERS

The eScience engineer concept has been developed in response to the operational needs of NLeSC and reflects the growing requirement for data scientists throughout the economy. It is already widely recognized throughout science that data-driven approaches are making an impact in a wide range of disciplines. As scientists noise that optimising the digital landscape is at the interface of computing, sciences that we can provide the best of both worlds. We can seamlessly integrate data scientists with limited informatics experience and ICT professionals without specific scientific expertise require access to a virtual team of experts from industry, academic and commercial partners to translate the imaginative combination of existing data management, data analysis and ICT to improve experimental design, communication.

The ultimate goal of eScience is to enable data-intensive science and sustainability, Environment & Technology. New, NLeSC has the capability to collaborate with the growing requirement for data and ICT to improve experimental design, communication.

As a problem driven organisation, NLeSC has prioritised the following five domains; Chemistry & Materials, Life & Agriculture, Physical Sciences & Technology, Information & Communication and Sustainability & Environment. Various partners collaborating in public-private partnerships to bridge these two worlds.

The eScience engineer concept has been developed in response to the operational needs of NLeSC and reflects the growing requirement for data scientists throughout the economy. It is already widely recognized throughout science that data-driven approaches are making an impact in a wide range of disciplines. As scientists noise that optimising the digital landscape is at the interface of computing, sciences that we can provide the best of both worlds. We can seamlessly integrate data scientists with limited informatics experience and ICT professionals without specific scientific expertise require access to a virtual team of experts from industry, academic and commercial partners to translate the imaginative combination of existing data management, data analysis and ICT to improve experimental design, communication.

The ultimate goal of eScience is to enable data-intensive science and sustainability, Environment & Technology. New, NLeSC has the capability to collaborate with the growing requirement for data and ICT to improve experimental design, communication.

As a problem driven organisation, NLeSC has prioritised the following five domains; Chemistry & Materials, Life & Agriculture, Physical Sciences & Technology, Information & Communication and Sustainability & Environment. Various partners collaborating in public-private partnerships to bridge these two worlds.

The eScience engineer concept has been developed in response to the operational needs of NLeSC and reflects the growing requirement for data scientists throughout the economy. It is already widely recognized throughout science that data-driven approaches are making an impact in a wide range of disciplines. As scientists noise that optimising the digital landscape is at the interface of computing, sciences that we can provide the best of both worlds. We can seamlessly integrate data scientists with limited informatics experience and ICT professionals without specific scientific expertise require access to a virtual team of experts from industry, academic and commercial partners to translate the imaginative combination of existing data management, data analysis and ICT to improve experimental design, communication.

The ultimate goal of eScience is to enable data-intensive science and sustainability, Environment & Technology. New, NLeSC has the capability to collaborate with the growing requirement for data and ICT to improve experimental design, communication.

As a problem driven organisation, NLeSC has prioritised the following five domains; Chemistry & Materials, Life & Agriculture, Physical Sciences & Technology, Information & Communication and Sustainability & Environment. Various partners collaborating in public-private partnerships to bridge these two worlds. The eScience engineer concept has been developed in response to the operational needs of NLeSC and reflects the growing requirement for data scientists throughout the economy. It is already widely recognized throughout science that data-driven approaches are making an impact in a wide range of disciplines. As scientists noise that optimising the digital landscape is at the interface of computing, sciences that we can provide the best of both worlds. We can seamlessly integrate data scientists with limited informatics experience and ICT professionals without specific scientific expertise require access to a virtual team of experts from industry, academic and commercial partners to translate the imaginative combination of existing data management, data analysis and ICT to improve experimental design, communication.

The ultimate goal of eScience is to enable data-intensive science and sustainability, Environment & Technology. New, NLeSC has the capability to collaborate with the growing requirement for data and ICT to improve experimental design, communication.

As a problem driven organisation, NLeSC has prioritised the following five domains; Chemistry & Materials, Life & Agriculture, Physical Sciences & Technology, Information & Communication and Sustainability & Environment. Various partners collaborating in public-private partnerships to bridge these two worlds. The eScience engineer concept has been developed in response to the operational needs of NLeSC and reflects the growing requirement for data scientists throughout the economy. It is already widely recognized throughout science that data-driven approaches are making an impact in a wide range of disciplines. As scientists noise that optimising the digital landscape is at the interface of computing, sciences that we can provide the best of both worlds. We can seamlessly integrate data scientists with limited informatics experience and ICT professionals without specific scientific expertise require access to a virtual team of experts from industry, academic and commercial partners to translate the imaginative combination of existing data management, data analysis and ICT to improve experimental design, communication.

The ultimate goal of eScience is to enable data-intensive science and sustainability, Environment & Technology. New, NLeSC has the capability to collaborate with the growing requirement for data and ICT to improve experimental design, communication.

As a problem driven organisation, NLeSC has prioritised the following five domains; Chemistry & Materials, Life & Agriculture, Physical Sciences & Technology, Information & Communication and Sustainability & Environment. Various partners collaborating in public-private partnerships to bridge these two worlds. The eScience engineer concept has been developed in response to the operational needs of NLeSC and reflects the growing requirement for data scientists throughout the economy. It is already widely recognized throughout science that data-driven approaches are making an impact in a wide range of disciplines. As scientists noise that optimising the digital landscape is at the interface of computing, sciences that we can provide the best of both worlds. We can seamlessly integrate data scientists with limited informatics experience and ICT professionals without specific scientific expertise require access to a virtual team of experts from industry, academic and commercial partners to translate the imaginative combination of existing data management, data analysis and ICT to improve experimental design, communication.

The ultimate goal of eScience is to enable data-intensive science and sustainability, Environment & Technology. New, NLeSC has the capability to collaborate with the growing requirement for data and ICT to improve experimental design, communication.
At the core of NLeSC are a dedicated team of scientists able to work at the interface of their own scientific disciplines and dividing their time between NLeSC to develop generic solutions and approaches, and the project engineers working within our collaborative teams of multidisciplinary research teams. On the one hand the engineers ensure that domain researchers can make use of appropriate research tools. On the other hand, the engineers are constantly on the lookout for generalizations – i.e. research questions posed by domain experts as such that an effective mapping onto the fundamental physics of the universe. Astrophysics, we are now working within our collaborative teams of multidisciplinary research teams. On the one hand the engineers ensure that domain researchers can make use of appropriate research tools. On the other hand, the engineers are constantly on the lookout for generalizations – i.e. research questions posed by domain experts as such that an effective mapping onto the fundamental physics of the universe.
**eSCIENCE INTEGRATORS**

**Exploiting the transformative potential of eScience**

**INTRODUCTION**

The eScience Integrators are a number of leading scientists who have been identified and invited to become key additional value. What they all share is a strong commitment to their specific domain and included researchers, one of the original eleven NLeSC priority domains that would be impossible for NLeSC to develop alone. The Integrators also have an important role to play in introducing scientists to disciplines that they would otherwise be unaware of and in helping to foster new collaborations.

**DATA-STEWARDSHIP**

A crucial component of any capital investment in science & equipment is the management, funding, and/or as advisors, during an externally reviewed selected call. The selected call was initiated to help develop NLeSC’s initial project portfolio, either directly as project leaders or to provide a key additional value.

**CONVERGING TECHNOLOGIES**

In the autumn of 2011 the first large meeting hosted at the newly delivered NLeSC facility at the Science Park in Amsterdam brought together each of the eScience Integrator teams to share their experiences and ideas on the development of their projects. The meeting was the joint recognition of a regular series of meetings where each of the team share ideas, discuss issues related to science and help develop the future direction of eScience and enhanced services in the Netherlands. The Integrators also have an important role representing the needs and opinions of the scientific discipline within eScience. They can often be dominated by voices from the world of computer science. In fact, that joint voice is often more valuable and powerful, as a result of being shared, within the scientific community.

**EVOLVING TEAM**

In the autumn of 2013 NLeSC identified a number of key additional value to the eScience Integrator team. Each representative selected has an extensive history in a number of key additional values. The regular Integrator meetings help promote the ‘cross-disciplinary’ thinking and generate new collaborations. NLeSC is able to support these new collaborations in a number of ways including the development and funding of PathFinding projects which are more fully introduced later in this document. Currently two of the three FuturePathFinding projects are led by eScience Integrators.
Amandus Lundqvist and Jos Engelen, the interviewers, sit at a large conference table at the Netherlands eScience Center. They discuss the rise of eScience, the Center’s mission, and its role in enhancing scientific collaboration and innovation. Engelen mentions that the mission of the Netherlands eScience Center is to enhance scientific enhancement and innovation in science. Lundqvist also mentions a study by the Advisory Council for Science and Technology Policy (AWT) of the top priority areas in eScience, which include systems biology, emerging fields, and modern eScience concepts. He believes that the data generated by these projects and partnerships will help researchers develop new strategies. Engelen and Lundqvist believe that eScience offers many new opportunities and will help researchers develop new strategies and technologies.
Beyond Individual Knowledge

"Nowadays, hardly anyone wins a Nobel Prize for his own work alone. As one of the integrators appointed at the Dublin conference last October, I feel the impact of eScience. It signifies a paradigm shift in the way knowledge is created and shared." — Barend Mons, Director of the Netherlands eScience Center (NLeSC), Professor of the University of Leiden and NLeSC Integrator for the Life Sciences

The Three E’s of eScience

Life sciences is becoming an integrative academic field. Mons also subscribes to a broad definition of eScience: "The complexity of today’s discipline is one factor. The complexity of today’s culture is another. So, the field itself is undergoing a fundamental shift."

Interdisciplinary approaches, therefore, are essential. "To properly convey this complexity, the eScience Center gives all researchers a different perspective on the Life Sciences. It’s a common view, is bridging gaps between diverse disciplines. "Simply to process that data, data generated within all the different disciplines. "Simply to process that data, you need a conductor. And if, on top of that, you also need to perform on the top international stages, then you’ll also need some big name producers to back you." — Barend Mons, Director of the Netherlands eScience Center (NLeSC)

The interview begins by explaining the first dimension of eScience: "The first dimension has to do with high-quality data generated within all the different disciplines. "There’s a great variety of data producers to back you."

ESCIENCE EVANGELISM

The interview then moves on to the second dimension of eScience: "The second dimension is about these fundamental changes in the way scientific research is performed. "I am highly motivated. I’m in a phase in my career that I don’t need to direct all my energies on publication. "As the eScience Center’s Integrator, I feel the impact of eScience. It signifies a paradigm shift in the way knowledge is created and shared."

FROM "READING" TO "CONSULTING"

The interview then moves on to the third dimension of eScience: "The third dimension is about these fundamental changes in the way scientific research is performed. "As the eScience Center’s Integrator, I feel the impact of eScience. It signifies a paradigm shift in the way knowledge is created and shared."

Expert Finder in Life Sciences

Barend Mons’ big dream for this form of science is to "profile all of them and fit them together — and all the schemata they’re making into a complex process." — Barend Mons, Director of the Netherlands eScience Center (NLeSC)
“We have to pioneer, otherwise we can’t get where we want to go.”

The Netherlands is among the best countries in the world to work in astronomy. One of the main reasons is that we are willing to look beyond borders,” tells De Vos. “To be a good researcher, you have to be in the first to start with radio astronomy. In the Netherlands we believe that devices are necessary, so we became one of the first to start with radio astronomy. The Low Frequency Array-radio telescope (LOFAR) helped us a lot, since its technology is not there yet. We worked solely with professionals educated at ASTRON, but later we hired people with a strong mathematical background in the field of algorithms. The LOFAR helped us to create new possibilities and opportunites.”

According to De Vos, the introduction of a new way of thinking is in the past very successful with professional researchers at ASTRON, but we are also starting working with high performance computer experts who had a less or no education in ICT. We also hired people with a strong mathematical background in the field of algorithms. The LOFAR helped us to create new possibilities and opportunities.

Astronomers are the leading customer within NLeSC. It’s fantastic. We benefit a lot from the multidisciplinary focus, although there are large differences between different disciplines,” says De Vos. “In some fields of sciences, for example genomics, these differences between different disciplines are eliminated at NLeSC.”

Despite these and other differences, De Vos does not see every strong synergy we are making. “Every eScience Integrator has his or her own way of doing that, I think. I stimulate people to provoke those knowledge sharing of ideas as a large European project. That’s how I try to create support for eScience.”

Working with Big Data and the corresponding software demands different approaches than fundamental sciences. De Vos is in a project that is associated with the Hanze Institute of Technology as an applied researcher. “Making Big Data requires a good mix of researchers and applied professionals if you manage to create that kind of mix, there are enormous possibilities on the labor market that could give the Netherlands a huge competitive advantage.”

“Every eScience Integrator has his or her own way of doing that, I think. I stimulate people to provoke those knowledge sharing of ideas as a large European project. That’s how I try to create support for eScience.”

“Every eScience Integrator has his or her own way of doing that, I think. I stimulate people to provoke those knowledge sharing of ideas as a large European project. That’s how I try to create support for eScience.”
Working with Big-Data is one of the big challenges for science, says Louis Vertegaal from the Netherlands Organization for Scientific Research (NWO). “It’s like sand in a big desert. You have to know what to extract and how to work with it, in order to achieve new ways of science.” In short: all disciplines in which working with vast amounts and diversity of data is common ground. “Especially astronomy is a pioneering discipline”, Vertegaal says.

Vertegaal’s story resembles that of ASTRON-director Marco de Vos. The universe in all its complexity inhibits a gigantic amount of scientific data, which makes working in a certain discipline a very complex task. Vertegaal: “An example of this is the development of the Square Kilometer Array (SKA) in South-Africa and Australia, the new radio telescope they are about to install. The amount of data traffic which will be generated is enormous.”

ORDER IN THE DATA-CHAOS

According to Vertegaal it’s the informaticians who are now creating order in the data chaos. The Netherlands eScience Center is founded to provide for interdisciplinary interaction. “I’ve known the director of the eScience Center, Jacob de Vlieg, for a long time. We worked together years ago. What they are doing at the eScience Center is helping scientists to cooperate with businesses on a demand-driven basis.”

One example of such cooperation is the development of the Square Kilometer Array (SKA). In Westerbork (NL) a new radio telescope they are about to install. The amount of data traffic which will be generated is enormous.

WANTED: EXCELLENT INFORMATICIANS

“We cover the excellent Big-Data users in the Netherlands, so what we now need is excellent informaticians who are ready to work with that data”, says Vertegaal. “We have to develop extremely fast and energy efficient computer systems. The SKA is a good example of this. It’s a gigantic project with the help of large super computers.”

One example of such cooperation is the development of the Square Kilometer Array (SKA). In Westerbork (NL) a new radio telescope they are about to install.

NEW FUNDAMENTAL QUESTIONS ARISE

At the same time Vertegaal witnesses a shift in the way we are doing science, now that the computational power is growing and Big-Data is coming up: “The new fundamental questions we can ask ourselves are so different that a new type of science might arise. A digital cross section of philosophical books and writings enable us to ask totally different questions. It’s hard to predict where this will lead, but I am certain that it will be possible to edit data. That one can manage things one could not have been doing before. Undoubtedly, this will raise new questions we never would have thought of earlier.”

Dr. Louis Vertegaal

Louis Vertegaal, director at NWO, responsible for the disciplines of astronomy, chemistry, computer science and mathematics speaks about the growing challenge of Big-Data and eScience.

“Dr. Louis Vertegaal

It’s like sand in a big desert. You have to know what to extract and how to work with it, in order to achieve new ways of science.”

Louis Vertegaal, director at NWO, responsible for the disciplines of astronomy, chemistry, computer science and mathematics speaks about the growing challenge of Big-Data and eScience.

A DESERT OF INFORMATION

Vertegaal’s story resembles that of ASTRON-director Marco de Vos. The universe in all its complexity inhibits a gigantic amount of scientific data, which makes working in a certain discipline a very complex task.

A DESERT OF INFORMATION

One example of such cooperation is the development of the Square Kilometer Array (SKA). In Westerbork (NL) a new radio telescope they are about to install. The amount of data traffic which will be generated is enormous.

WANTED: EXCELLENT INFORMATICIANS

“The new fundamental questions we can ask ourselves are so different that a new type of science might arise. A digital cross section of philosophical books and writings enable us to ask totally different questions. It’s hard to predict where this will lead, but I am certain that it will be possible to edit data. That one can manage things one could not have been doing before. Undoubtedly, this will raise new questions we never would have thought of earlier.”

Louis Vertegaal, director at NWO, responsible for the disciplines of astronomy, chemistry, computer science and mathematics speaks about the growing challenge of Big-Data and eScience.

“Dr. Louis Vertegaal

It’s like sand in a big desert. You have to know what to extract and how to work with it, in order to achieve new ways of science.”

Louis Vertegaal, director at NWO, responsible for the disciplines of astronomy, chemistry, computer science and mathematics speaks about the growing challenge of Big-Data and eScience.

A DESERT OF INFORMATION

Vertegaal’s story resembles that of ASTRON-director Marco de Vos. The universe in all its complexity inhibits a gigantic amount of scientific data, which makes working in a certain discipline a very complex task.

A DESERT OF INFORMATION

One example of such cooperation is the development of the Square Kilometer Array (SKA). In Westerbork (NL) a new radio telescope they are about to install. The amount of data traffic which will be generated is enormous.

WANTED: EXCELLENT INFORMATICIANS

“The new fundamental questions we can ask ourselves are so different that a new type of science might arise. A digital cross section of philosophical books and writings enable us to ask totally different questions. It’s hard to predict where this will lead, but I am certain that it will be possible to edit data. That one can manage things one could not have been doing before. Undoubtedly, this will raise new questions we never would have thought of earlier.”

Louis Vertegaal, director at NWO, responsible for the disciplines of astronomy, chemistry, computer science and mathematics speaks about the growing challenge of Big-Data and eScience.

“Dr. Louis Vertegaal

It’s like sand in a big desert. You have to know what to extract and how to work with it, in order to achieve new ways of science.”

Louis Vertegaal, director at NWO, responsible for the disciplines of astronomy, chemistry, computer science and mathematics speaks about the growing challenge of Big-Data and eScience.

A DESERT OF INFORMATION

Vertegaal’s story resembles that of ASTRON-director Marco de Vos. The universe in all its complexity inhibits a gigantic amount of scientific data, which makes working in a certain discipline a very complex task.

A DESERT OF INFORMATION

One example of such cooperation is the development of the Square Kilometer Array (SKA). In Westerbork (NL) a new radio telescope they are about to install. The amount of data traffic which will be generated is enormous.

WANTED: EXCELLENT INFORMATICIANS

“The new fundamental questions we can ask ourselves are so different that a new type of science might arise. A digital cross section of philosophical books and writings enable us to ask totally different questions. It’s hard to predict where this will lead, but I am certain that it will be possible to edit data. That one can manage things one could not have been doing before. Undoubtedly, this will raise new questions we never would have thought of earlier.”
EvolutioN of Star ClustersOver Time

Evolution of star clusters from disciplinary interests and desire to deliver a new type of scientific output after finding advantages hard to sustain. Seinstra replies, "There were lots of great opportunities for me as my scientific and technological skills. New partnerships with chemists, plant biologists and astronomers as well as multi-media engineers as well as multi-media engineers. Our rapidly improving understanding of climate research changing the way researchers approach their task.

Supercomputers facilitate a new paradigm in science, where we look at the evolutionary history of star clusters. Our work is focused on bringing these models together in a way that transcends the domain. The Netherlands eScience Center offers a great possibility to bring scientists from around the world together. That is a great step forward.

"Frankly, enough climate research is being conducted in the same way. We need separate models for the ocean, atmosphere and sea ice. The models are focused on bringing these models together in a way which transcends the domain. That's a development we are extensively more and more aware of at the Netherlands eScience Center."

"Our work is focused on bringing these models together in a way which transcends the domain."
The Netherlands eScience Center has an annual budget of M€6 and collaborates with scientific research groups from both academia and industry to conduct funded projects, balancing short-term eScience results and the development of a longer-term eScience strategy. The Center focuses on multi-disciplinary research collaboration and the clever and innovative combination of eScience methods that can be used in the hard sciences, the humanities, and the social sciences. We aim to achieve cross-fertilization between disciplines and between science/scholarship and ICT, and to become a coherent long-term, cost-effective eScience environment for sciences and humanities.

Each project, resulting from a variety of different funding calls, is supported to the value of K€500 and include the deployment of an eScience engineer. Project proposals are prioritized by NLeSC’s scientific advisory committee and finally sanctioned by NLeSC’s board of directors.

The projects involve mainly multidisciplinary, data-intensive research and concern the development of new and innovative eScience methodologies, techniques, and tools. The tools should have potential for broad application/deployment and align with the key NLeSC scientific themes.
The most advanced modes of large radio/millimeter telescopes like ALMA and LOFAR require structural collaboration between astronomers and eScience experts. Our Astronomy project addresses the issues related to the huge size of the datasets produced by the World’s largest radio telescopes. Areas of potential optimization include interoperability between existing packages, advanced processing platforms, distributed user support, extremely large databases, and streaming processing pipelines.

Pulsars are rapidly rotating neutron stars whose signal is received on earth periodically. Their big mass and precise period can be used to probe space and gravitation. The discovery of the first binary pulsar by Hulse and Taylor in 1973 has been so important for verifying general relativity and gravitation that they won the 1993 Nobel Prize for physics.

Finding new pulsars is difficult: it involves a brute-force search over many parameter combinations, because we do not know the position, distance, and period of the pulsars. Searching for pulsars is a Big-Data problem: typical observations produce hundreds of terabytes, and petabytes of intermediate results are produced and analyzed. Moreover, the pulsar signal is faint and can be completely covered by Radio Frequency Interference (RFI), which has to be analyzed and removed from the signal.

In the eScience project, we are greatly speeding up the search for new pulsars by using Graphics Processing Units (GPUs) for the complex computations. We are developing a complete real-time pulsar detection system using a GPU cluster, and are testing it with two Dutch radio telescopes, LOFAR and Apertif. LOFAR currently is the largest radio telescope in the world. The results of this work are extremely important for developing hardware and software on GPUs to process the raw data from large radio telescopes in real time, and for developing new techniques for pulsar detection.

In the project, we are developing a software pipeline for pulsar searches on GPUs. The pipeline includes algorithms for detecting signals, removing RFI, and identifying pulsars. The pipeline is designed to be efficient and scalable, and it is being tested on two Dutch radio telescopes, LOFAR and Apertif.

In the project, we are exploring a new development model for astronomical processing software where expertise from a variety of disciplines is combined, including mathematics for the foundations of new algorithms and computer science to optimize for high-performance platforms. Optimized software and demonstrators will be developed that can be re-used in a variety of contexts, not just for radio/millimeter astronomy, but also in other areas where large data streams are collected.

Co-financing for this work is obtained through the recently approved Hilado project, a Joint Research Activity in the EC FP7 Integrating Activity RadioNet-3. Hilado is led by ASTRON, JIVE is a major partner.

In the project, we are exploring a new development model for astronomical processing software where expertise from a variety of disciplines is combined, including mathematics for the foundations of new algorithms and computer science to optimize for high-performance platforms. Optimized software and demonstrators will be developed that can be re-used in a variety of contexts, not just for radio/millimeter astronomy, but also in other areas where large data streams are collected.

Co-financing for this work is obtained through the recently approved Hilado project, a Joint Research Activity in the EC FP7 Integrating Activity RadioNet-3. Hilado is led by ASTRON, JIVE is a major partner.

The consortium behind this project, including the Netherlands’ premier institute for radio astronomy, is building on this and other initiatives, including eScience Engineer Dr. Rob V. Nieuwpoort, Project Leader Dr. Marco de Vos, Managing Director ASTRON.
The goal of the eSALSA project is to increase our understanding of the regional effects caused by changes in the Atlantic Meridional Overturning Circulation (AMOC) by utilizing an eddy-resolving ocean model, the Parallel Ocean Model (POP), to determine sea-level changes with an unprecedented level of detail (10 km resolution).

There is still much uncertainty about the effects of climate change on the AMOC, which includes the Gulf Stream that originates at the tip of Florida. As a result of this AMOC, the climate in Western Europe is significantly milder than can be expected for its geographic location.

The AMOC is sensitive to freshwater input into the North Atlantic, for example due to melting of the Greenland Ice Sheet. Key changes in the AMOC will affect surface temperatures over Western Europe and cause regional changes in the Atlantic. With increased uncertainties in estimating the effects of climate change on the North Atlantic ocean circulation, the technical challenge is to integrate the ocean model under a range of different atmospheric forcing conditions for about 100 years such that also the uncertainties and extremes in sea-level changes can be determined.

To be able to perform the computations within the duration of the project it is essential to use advanced distributed computing techniques and state-of-the-art accelerators such as GPU’s. For efficient analysis of the hundreds of terabytes of output we are developing new visualization techniques and remote collaboration tools. This allows leading climate researchers worldwide to discuss simulation results as soon as they become available.

Dr. Jason Maassen
Project Leader
Prof. Dr. Henk Dijkstra,
Utrecht University

Dr. Jason Maassen

Understanding the effects of Climate Change

eSALSA - Predicting Local Sea-Level Changes

Advanced visualization techniques are used to analyse the huge amounts of data produced by climate simulations.
Neuroscientists analyze neuroimaging data to find biomarkers that predict who is at risk for developing an illness. The larger the sample size, the better the predictive value of the biomarker, which can be achieved by combining datasets obtained by different medical centers. This requires an eScience infrastructure for standardized image analysis, and the exchange of various meta-data and analysis results.

The aim of the Biomarker Boosting project is to build a reusable platform for sharing patient (subject) imaging data among hospitals to run a common analysis pipeline. This considerably increases the number of datasets available for analysis, thereby greatly improving the statistical value of the results.

The multidisciplinary team is developing tools and services for cooperative research in studying brain disease and demonstrates its potential by applying it to four different patient cohorts to study dementia and mild cognitive impairment (MCI). For the purpose, a common catalog (meta-data in which definitions of database objects are stored) is designed to allow for data exchange using web services, together with a researcher data sharing agreement to allow access to research-related clinical data. The second focus is to implement a standardized image analysis pipeline to calculate biomarker features. The product of the analysis pipeline is a set of low-dimensional properties derived from the high-dimensional images. These properties are called biomarkers, with volume of the hippocampi as an example. The improved statistical power is called ‘boosting’, hence the title ‘Biomarker Boosting’.

This project will demonstrate progress in clinical neuroscience through the use of eScience technologies to merge highly heterogeneous distributed data resources. It allows for secured and regulated access to the data and its derivatives, establishing a framework for biomarker extraction.

The enhanced science platform is extending the reach of eScience to neuroimaging data.
Metabolomics, the technology to comprehensively measure (changes in) the metabolites in a biological sample, has great potential to impact on our understanding of biological systems and processes at a chemical level. Full exploitation of metabolomics data is currently limited by the complexity of the datasets generated within current platforms which are difficult to manage by human experts alone. eScience technology is therefore required to play a crucial role in mining and interpreting complex metabolomics data.

In this project we are developing a computational workflow to improve and accelerate metabolite identification and biochemical pathway reconstruction in metabolomics studies. A key step in the workflow is generating an in silico metabolite network on the basis of empirically derived reaction rules that delivers candidate structures for unknown metabolites in a metabolomics experiment. This will allow more systematic and automated structure elucidation on the basis of empirically derived reaction rules that deliver candidate structures for unknown metabolites in a metabolomics experiment.

The workflow is built on a flexible data infrastructure, efficient and parallelized computational algorithms and visualization of complex data. The result will be a practical toolbox that will be integrated with existing workflows for metabolomics data analysis.

Chemical informatics for metabolite identification and biochemical network reconstruction

Deeper understanding of living systems
Ecology is evolving into a data and computationally intensive science with the amount of (heterogeneous) data included for analysis increasing rapidly with time because of 1) worldwide exchange of data through the Internet, 2) the use of sensors that produce massive amounts of data (remote sensing or sensor networks), and 3) the incorporation of the environmental data (often spatiotemporal gridded data).

Traditionally, ecologists are not trained in coping with the massive amounts of data that result from data sharing, sensor networks and the incorporation of environmental data into ecological research. Generally, the methodologies ecologists use for management, visualization, exploration, analysis of data are often not suited to cope with large datasets. The challenge of the e-Ecology project is to bridge the gap between the worlds of ecology and technology. Virtual Labs (VL’s) will help to bridge this gap as they support scientific (multidisciplinary) collaboration by facilitating data access, data integrity and quality control, data post-processing, data storage and backup, data merging, data sharing, interactive data visualizations, and data analysis.

Most efforts have gone into the Bird Movement Modeling VL (www.UvA-BiTS.nl) which has a growing international user community. Users are mainly field biologists who track individual birds. A GPS-tracker is attached to a bird and used to record the GPS coordinates and 3D accelerometer values of the bird. The track data can then be combined with landscape data, weather and tidal data to gain new insights into the influence of the environment on the bird’s behavior. There are plans to use the experience with UvA-BiTS to build 2-3 other VL’s:

- A VL-EcoGrid which builds on the existing Dutch National Database for Flora and Fauna (NDFF).
- A VL-ENRAM (European Network for Radar surveillance of Animal Migration) for a multidisciplinary network of radar-biologists researchers.
- A VL-Wadden for research on the Wadden Sea where data types vary from visual observations to video data for automated classification.

Enhanced Science

Virtual Laboratories for inspiration and discovery in Ecology

The mapping of multi-scale movements of gulls from Texel.
Unlocking our Hidden History

BiographyNed: Extracting relations between people and events

The Biography Portal of the Netherlands holds a wide variety of Dutch online reference works and databases, written in different times, from different perspectives, through a limited number of metadata. This project will build a semantic layer on top of the current Biography portal in order to enrich our sources and analytical tools for history writing. This project aims to enhance the potential for historical research on the portal’s virtual community of the more than 100,000 Dutch people mentioned in the various linked databases by transforming the available data into a semantic knowledge base and through the creation of a demonstrator.

BiographyNed is a multidisciplinary project that combines expertise from historians, computer science and computational linguistics. This requires the tackling of many problems at the intersection of history writing, computational linguistics and linked data research related to various types of event extraction and network analysis.

The lead questions for the design of a semantic demonstrator are which relations can be revealed between people and events, geographical movements and networks between people? What do they tell historians about the formation of Dutch society and the foundations of the Netherlands?

The current search engine lacks the analytic tools to show interconnections, trends, geographical maps, time lines, etc. This project aims to strengthen the value of the portal and comparable biographical datasets for historical research by improving the search options and the presentation of its outcomes, starting from the Simple Event Model. The demonstrator will add a semantic layer on top of the current Biography Portal. The demonstrator can also include information from external resources, such as museum objects or Wikipedia. Ultimately, the project may help to uncover new relations between people and events by linking data that has mainly been studied in isolation so far. The pilot will focus on a qualitative selection of links, relevant to the National Portrait Gallery, currently being developed by the Rijksmuseum and eventually make these tools available to the eHumanities community at large.

The BiographyNed project aims to create a semantic knowledge base by extracting relations between people and historic events, in the Biography Portal.
The Dutch have a historical expertise and interest in water management, and in the eWaterCycle project. Researchers from the Technical University of Delft, Utrecht University, and NLeSC are cooperating to build a high-resolution, realistic model of the World’s supply of fresh water. Using this model, it will not only be possible to build a flood early warning system, but also predict the effect of unsustainable water supply usage, provide support to local governments in making decisions on water protection measures, and provide other information vital to the World’s population.

With climate change, growing population and increasing pressures on land usage, water management is quickly becoming one of the major problems of the world. Increased urbanisation of delta areas particularly in making more and more people vulnerable to flooding. The development of a high resolution global hydrological model has recently been put forward as Grand Challenge for the hydrological community. So far, the hydrological community has not yet made full use of the possibilities that recent trends in the availability of computational power offer. Current global models lack the resolution required to make accurate predictions and adequately support decision-making. Even more challenging than the refinement of the grid will be the assimilation of the massive amounts of available weather data. In order to ensure proper parameterisation of such a model, massive assimilation of massive amounts of data is needed. All of this data needs to be incorporated into a very high resolution simulation, enabling it to make decisions, not only run on very large supercomputers. In this project, we aim to exploit high performance computing infrastructures to run these simulations. As allowing the team to take into account the massive amounts of available earth observations will be a major computational challenge the demands close cooperation between ICT and hydrology. Ultimately, these models will help in making the world a safer place, as we try both lives and money.

Managing our Place in the World

enhanced Science

Data-intensive modeling of the global water cycle

The Dutch have a historical expertise and interest in water management, and in the eWaterCycle project. Researchers from the Technical University of Delft, Utrecht University, and NLeSC are cooperating to build a high-resolution, realistic model of the World’s supply of fresh water. Using this model, it will not only be possible to build a flood early warning system, but also predict the effect of unsustainable water supply usage, provide support to local governments in making decisions on water protection measures, and provide other information vital to the World’s population.

With climate change, growing population and increasing pressures on land usage, water management is quickly becoming one of the major problems of the world. Increased urbanisation of delta areas particularly in making more and more people vulnerable to flooding. The development of a high resolution global hydrological model has recently been put forward as Grand Challenge for the hydrological community. So far, the hydrological community has not yet made full use of the possibilities that recent trends in the availability of computational power offer. Current global models lack the resolution required to make accurate predictions and adequately support decision-making. Even more challenging than the refinement of the grid will be the assimilation of the massive amounts of available weather data. In order to ensure proper parameterisation of such a model, massive assimilation of massive amounts of data is needed. All of this data needs to be incorporated into a very high resolution simulation, enabling it to make decisions, not only run on very large supercomputers. In this project, we aim to exploit high performance computing infrastructures to run these simulations. As allowing the team to take into account the massive amounts of available earth observations will be a major computational challenge the demands close cooperation between ICT and hydrology. Ultimately, these models will help in making the world a safer place, as we try both lives and money.
Translating medical discoveries into clinical care

Sustainable Infrastructure for Translational Medical Research

The CTMM (Center for Translational Molecular Medicine) is a Netherlands-based public-private partnership dedicated to the development of technologies that enable early diagnosis and personalized treatment. CTMM projects work mainly in the areas of cancer and cardiovascular research, but also include work on Alzheimer’s, and rheumatoid arthritis.

The CTMM Translational Research IT (TraIT) project was initiated to develop a sustainable IT infrastructure for the Netherlands that facilitates the collection, storage, analysis, archiving and sharing of translational research data. TraIT addresses these challenges by developing an IT infrastructure that facilitates translational research logistics, data management, data integration, and data analysis at a national level, and by being the Dutch hub for international networks.

The project aims to adopt and adapt existing point solutions rather than embarking on major software development projects. A shared IT infrastructure enables secure and sustainable access to research results and appropriate software tools to collect, manage, integrate, and analyze these results. Professional training and support are available to enable medical researchers to use the software tools and work with the data efficiently.

eScience Engineer
Dr. Susan E. Branchett
Project Leader
Dr. Jan-Willem Boiten, Center for Translational Molecular Medicine

Dr. Susan E. Branchett
Dr. Jan-Willem Boiten
Center for Translational Molecular Medicine

The computer has become a fundamental tool at all stages of the health care cycle.
Scientific publications contain a wealth of unexplored and unstructured data. Mining and cross-linking this information with other sources can provide valuable insight in the physiology of organisms, explain experimental data or lead to new hypotheses.

The technological developments in life science research have led to a vast increase in data that are available in public and proprietary databases. In order to efficiently capitalize on these data, dedicated vocabularies and algorithms are necessary for annotating, searching, mining and integrating data from various sources. Although a number of generic knowledge discovery, knowledge management (KDKM) and text mining (TM) tools exist, their application in life science areas, in particular food research, is limited. Once research into the absence of structured vocabularies that are of interest to specific applications in food research.

In this research project, we are developing structured vocabularies covering the food domain. These vocabularies will be incorporated in existing KDKM and TM tools to improve their performance. Using these vocabularies, we aim to generate insights into the function of bacteria, organisms involved in food processing, etc. Furthermore, we are to identify hidden relations which may lead to a better understanding of how processes work or may lead to improved products. These relations can be used to generate hypotheses addressing important areas in food research.

The ontologies and related (web) services will be evaluated in two ways. Firstly, the ontology and associated services will be validated by measuring the quality of semi-automatic annotations and by demonstrating improved integration of food research data. Secondly, the hypotheses, generated with the above computational methods, will be validated in experiments in which the effects of probiotics and neutraceuticals are measured in in-vitro and in-vivo models for health. Visualizations of terms often co-occurring with bacteria do already summarize the main applications of those bacteria by a single mouse click. By the introduction of the food concepts we will be able to tag these terms which will enrich the set of terms with useful concepts which enables discovery of new relations between bacteria and concepts.
Genomics-based technologies in life sciences offer many opportunities for academic and industrial innovation in the green life-sciences domain, which will lead to completely new approaches and challenges in plant research and breeding. This project addresses challenges in data handling, high-performance computing and data management which demand adequate eScience environments for the retrieval, analysis, manipulation and use of data.

Genomics-based technologies in life sciences are revolutionizing academic and industrial innovation in the agro-food/green life-sciences domain. Our understanding of the genetic basis of important plant traits is radically changing and will lead to completely new approaches in plant research and breeding. These new technologies create challenges in data production but foremost in data handling, high-performance computing, data management, standardization, statistics, design for experimentation, visualization, and multidisciplinary collaboration.

This project will build upon an existing public-private project called Virtual Lab for Plant Breeding, and is a combined NLeSC and Technology Top Institute Green Genetics (TTI-GG) project involving important parties relevant for the Dutch Green Genetics community.

To effectively utilize these technologies in plant science and plant breeding, adequate eScience environments for storage, retrieval, security, analysis, manipulation and use of data are needed. Although most issues have a large technical component, there are profound conceptual, methodological and even social components to consider. These challenges can hardly be met by any individual company or organization, hence in 2009 a pre-competitive initiative, the Virtual lab for Plant Breeding (VLPB) was initiated. Starting with 4 companies and 3 academic institutes, now over 15 organizations and companies are participating.

VLPB tackles eScience challenges in the green life sciences across several key areas: proper design for experimentation, good bioinformatics methodology, functional problem-solving environments, reliable e-infrastructure and adequate e-bioscience support. In addition to this, much attention is paid to support the VLPB community and secure its continuity.

The Virtual Laboratory for Plant Breeding
In addition to the major projects, NLeSC sponsors smaller eScience initiatives, referred to as "Path Finding" projects, which are intended to provide NLeSC with the opportunity to rapidly meet short-term scientific challenges, address immediate technological goals or investigate the potential to initiate full projects. Each project receives €50 funding or FTE 0.5 eScience engineer support.

Path Finding projects should deliver one of the following:

- Develop a generic eScience solution, from initiation to roll-out, with potential utility in a number of areas
- Demonstrate the scientific or technological viability of a novel and challenging eScience approach with the intention to use this proof-of-concept to support a full project proposal
- Provide previously unforeseen eScience support to existing NWO or KNAW projects with the expectation that this input will significantly improve the scientific output of the project and provide publication opportunities
- The project results should contribute to the eScience Technology Platform

NLeSC and its partners have initiated three Path Finding projects, and based on the success of these activities, we plan to launch several more in 2013.
The early evolution of star clusters is one aspect of the formation of our Universe which is not yet completely understood. Using the AMUSE Astrophysical environment, a simulation can be made that mimics the exact circumstances of such a cluster. We simulate gravity, as well as stellar evolution, and hydrodynamics of the gas surrounding the new stars. This simulation leads to new understandings of the processes involved, and their effects on the current state of our universe.

The above is only one example of the possibilities of the Astrophysical Multipurpose Software Environment, or AMUSE. AMUSE provides a software framework for large scale simulations of astrophysical systems, in which existing codes for gravitational dynamics, stellar evolution, hydrodynamics and radiative transfer can be easily coupled.

Based on the work done in the Ibis project at the VU University Amsterdam, we are building a robust, stable system to run these large scale simulations on a possibly distributed set of resources, including supercomputers. This will allow astrophysicists from around the world to scale up their simulations, increasing our understanding of the universe.

AMUSE is part of a larger goal to facilitate analysis of complex problems that can be simulated by combining multiple distinct models of physical (or other real-world) phenomena. Examples of such problems abound, including in climate research, water management, computational fluid dynamics, and self-assembling complex scien
tific systems. Generalization of the capabilities of AMUSE to such further research domains would be a prime example of accelerated eScience and domain cross-fertilization.
There is a growing interest from companies, governments and universities in the daily communication that takes place on online social media such as blogs, Facebook, and Twitter. Linguists and researchers in communication studies can use this data to study language variation and change. Companies may track the reputation of a product after its introduction. Governments may follow the spread of news messages and spot initial local reports of incidents. Police may monitor Twitter for suspicious behaviors. However, the amount of social media data is large and obtaining specific parts that are interesting for a certain purpose is not easy.

This Path Finding project aims at developing a centralized service for gathering, storing, and analyzing Twitter messages and making available derived information to a consortium of researchers in communication studies and language technology throughout the Netherlands. The service will be based on an existing platform set up at the ISLA (UvA) and the RUG with infrastructure from SURFsara. The Twitter API, providing free access to approximately 1% of all tweets worldwide, is constantly harvested and the resulting data stored. Interfaces to this data provide users with a number of analysis tools that can be run on all content and metadata.

Analysis of Social Media Messages: the Case of Twitter

The Path Finding project aims at developing a centralized service for gathering, storing, and analyzing Twitter messages and making available derived information to a consortium of researchers in communication studies and language technology throughout the Netherlands. The service will be based on an existing platform set up at the ISLA (UvA) and the RUG with infrastructure from SURFsara. The Twitter API, providing free access to approximately 1% of all tweets worldwide, is constantly harvested and the resulting data stored. Interfaces to this data provide users with a number of analysis tools that can be run on all content and metadata.

There is a growing interest from companies, governments and universities in the daily communication that takes place on online social media such as blogs, Facebook, and Twitter. Linguists and researchers in communication studies can use this data to study language variation and change. Companies may track the reputation of a product after its introduction. Governments may follow the spread of news messages and spot initial local reports of incidents. Police may monitor Twitter for suspicious behaviors. However, the amount of social media data is large and obtaining specific parts that are interesting for a certain purpose, is not easy.

This Path Finding project aims at developing a centralized service for gathering, storing, and analyzing Twitter messages and making available derived information to a consortium of researchers in communication studies and language technology throughout the Netherlands. The service will be based on an existing platform set up at the ISLA (UvA) and the RUG with infrastructure from SURFsara. The Twitter API, providing free access to approximately 1% of all tweets worldwide, is constantly harvested and the resulting data stored. Interfaces to this data provide users with a number of analysis tools that can be run on all content and metadata.
In parallel with the advancements of computing and information technologies and growing data availability, concept driven computer simulation models are becoming ever more complex. Models include increasing numbers of parameters as they contain longer chains of cause-and-effect relations and as requirements with respect to spatial and temporal resolution expand. As model parameters represent lumped system properties that can often not be measured at the scale of interest, model calibration (a subset of inverse modeling) is the only solution to parameterizing these models. It has been shown that Bayesian inverse modeling has much broader application possibilities than model calibration alone. Although the inverse modeling methodology is gaining interest, it is not yet widely used in all domains of science because it is data and compute intensive. It is envisioned that an interactive framework for inverse modeling will contribute to the acceptance of the approach.

To lower the barrier for potential users of the methodology, eSiBayes aims at the design, development, implementation and testing of an interactive framework for inverse modeling based on Bayesian statistics.

eSiBayes: eScience infrastructure for Bayesian inverse modeling

To facilitate the design of complex systems

enhanced Science

Jurriaan H. Spaaks
Project Leader

Prof. Dr. Ir. Willem Bouten,
University of Amsterdam
Whilst the focus of NLeSC is to contribute to scientific breakthroughs using advanced computing, rather than on breakthroughs in computing science, there does remain a need to develop generic technologies to underpin this work.

In many scientific disciplines, domain expertise alone is no longer sufficient for scientific progress. Researchers must also manage ‘Big-Data’, use advanced compute facilities, integrate (un)structured data of mixed type and origin, and rapidly share results. NLeSC needs to keep track of these developments, contribute to them and communicate resulting opportunities within the scientific community. Also NLeSC will continuously bridge the gap between the autonomously developing e-infrastructure in all its facets and the optimal implementations of software, applications and middleware to profit from the technology.

NLeSC is therefore funding two eScience technology development projects, one internal and one external, which work together with the aim to scout, adopt, develop, integrate, and make available a set of tools that abstract away the complexities of the underlying hardware, ranging from storage facilities and supercomputers, to high-resolution visualization systems and instruments. Driven by the demands of NLeSC application projects, the tools are then combined in complete eScience solutions. NLeSC aims to ensure exchange and re-use of best practices, to prevent fragmentation and duplication, and to enhance collaboration among very diverse areas of research and development.
An important aspect of eScience is the development of new methods and tools to support scientists to enhance the ways they conduct research and to optimize the route to new scientific discovery. The route from data to information to knowledge and insight can and should take optimal advantage of modern ICT facilities and e-infrastructures, but it requires specialist experience. Ideally, researchers should be engaged with scientific challenges rather than with ICT.

The goal of eSTeP is to develop tools, interfaces, and libraries to deal with and extract information from large amounts of distributed data, requiring high-speed networks, and high-resolution visualization equipment. Moreover, in many cases data and results, as well as compute kernels and full scientific workflows, are made sharable among multiple collaborating parties.

As part of our strategy we explicitly avoid ‘reinventing the wheel’ even for domain-specific eScience solutions. To this end, eSTeP follows a layered and modular approach. At the lowest layer, system-level libraries are developed (or adopted) and integrated, following five themes: ‘data’, ‘computing’, ‘networking’, ‘visualization’, and ‘other’ (e.g. reproducibility, provenance). The low-level libraries specifically aim at hiding the particular idiosyncrasies of accessing, and making optimal use of, the underlying hardware and middleware infrastructures. The highest layer offers generic and domain-specific solutions fit for optimized scientific discovery.

Project Leader
Dr. René van Schaik, NLeSC
Dr. Frank J. Seinstra, NLeSC
The project addresses several research challenges in computer science specifically needed to make breakthroughs in data-driven research. The project is focused on the data explosion problem, which is one of the most important challenges in almost all areas of science. The sheer volume and the distributed nature of many datasets lead to complicated technical problems around data transport and data processing. The data also become more complex and heterogeneous, especially when different datasets are combined from various instruments and databases, which is typical for experimental sciences. In particular, this complexity makes it a challenge to extract semantically useful information from the data.

The parallel and distributed computing environments on which applications have to run are also changing dramatically. As work, one processor core is replaced by multiple processor cores (multi-cores like GPUs) and networking (hybrid networks, sensor networks, storage networking, and middleware (virtualization and Clouds)).

The project divides focus between the volume aspect of the data explosion (Big-Data) and the complexity aspect (heterogeneous data). One postdoc works on infrastructure innovation and distributed data processing (especially on many-cores such as GPUs) and a second focuses on information management, in particular complex analysis of scientific datasets for various disciplines.

E-Science Engineer
Dr. Rob V. van Nieuwpoort
Project Leader
Prof. Dr. Henri E. Bal,
Computer Science,
Free University Amsterdam
eScience is an inherently multidisciplinary pursuit concerned with the need to bridge the gap between high-powered computing and networking on one side and data-intensive science on the other. Amongst the most crucial aspects of our eScience strategy is the exploitation of shared technologies across domains. How can we make use of tools developed in one field in other application areas? The ability to utilise existing tools to solve new scientific problems is at the core of NLeSC and our vision for enhanced Science. In this, eScience challenges isolationist approaches to research and the “not invented here” attitudes that prevent the identification of potential synergies and the sharing of technologies. NLeSC believes that there remains a huge opportunity to apply tools from one discipline within another, with the obvious benefits to the receiving party, but also ensuring the most value can be derived from the work of the donating party and expanding its impact in non-obvious directions.

Even technologies developed with very specific applications in mind may have the potential to be applied within other scientific areas, when synergies can be identified. eScience has an important role to play in introducing scientists to tools from other disciplines that they would otherwise be unaware of and in helping tailor these solutions for optimal application.

With this goal in mind, NLeSC is providing funding for new eScience projects that pioneer the application of ICT methods developed for one scientific discipline within new, application areas. For example, can we use image analysis methods developed by astrophysicists to analyse medical imaging data or use computer-assisted chemical characterisation approaches from the pharmaceutical industry in other fields requiring chemical analysis methods? The eSOCCS call is an attempt to fund researchers who share this idea that their discipline can benefit from the work undertaken in other areas.

The eSOCCS proposal received a total of fifty five pre-proposals whose primary applicant represent thirteen universities/Medical Centers and three NWO institutes. The five highest quality proposals have been prioritized by NLeSC’s external scientific advisory committee and approved for funding by our board of directors. Each project, including their project leader, is listed below. Each project will be funded to approximately half a million euros, including the collaboration of an NLeSC eScience engineer.

PORTFOLIO & PARTNERSHIPS

Continuing Development of NLeSC Project Portfolio

The Netherlands eScience Center (NLeSC) will fund five new eScience projects in early 2013 resulting from the “eScience Open Call for Converging Sciences (eSOCCS)” that was conducted during 2012.
NLeSC operates as a network institute with a physical center at the Science Park Amsterdam. NLeSC coordinates a collaborative scientific program including research groups from both academia and industry to conduct funded projects, balancing short-term eScience results and the development of a longer-term eScience strategy.

As a network institute the success of NLeSC is dependent on close collaboration with excellent scientists throughout the Dutch academic community. Without the world-leading research being conducted across the country’s universities and knowledge institutions, NLeSC would not be able to function. In our goals to add additional expertise to works being conducted at our collaborating research groups in a classic 1+1=3 model, NLeSC also share special relationships with both of its founding organisations, NWO and SURF, particularly SURFsara and SURFnet, and are grateful for their support.
NLeSC’s second call, focused on converging sciences (eSOCCS), was conducted via an open and less focussed call process to ensure wider engagement, provide support for eScience in unexpected domains and as a tool for communicating the role of NLeSC in general.

In line with the NLeSC strategy and continuity of the existing portfolio, the following granting lines will be available for 2013:

1. A joint thematic open call Data Science with NWO-EW which is linked to the Roadmap ICT Data, data, data initiative. This call focuses on new opportunities for collaboration between private companies and knowledge centers (public-private partnerships). The goal is innovation on production, management and analyses of very large datasets (Big-Data in terms of the 5V’s (volume, velocity, variety, verification and value). The call includes the Industrial Partnership Programme, Technology Area’s and Knowledge Innovation Mapping to optimally suit business demands. Total available subsidy budget approximately K€2,500, NLeSC contribution K€1,000.

2. A joint thematic open call Digging into Data with the focus on how ‘Big-Data’ changes the research landscape for the humanities and social sciences. Each project is a partnership among two or more national teams, from at least two of the four participating countries. Total call budget for the Netherlands K€550, NLeSC contribution K€100. Granted projects with a Dutch partner might receive an additional Path Finding project bonus (max. K€50) to make technology generally applicable and sustainable.

3. A selected thematic call to reinforce existing initiatives in e.g. Green Genetics, eChemistry, eScience Technology or Logistics. Discussion with eSAC on domains and call conditions planned for Q1. Total budget K€2,500.

4. A continuous selected call for Path Finding projects. These projects are intended to provide NLeSC with the opportunity to rapidly meet shorter-term scientific challenges, to address immediate technological goals as envisioned, the potential to initiate full projects. These projects have a maximum expenditure of K€50. The majority of funding should fund personnel costs and the maximum duration is one year. The project results should contribute to the eScience Technology Platform (eSTeP). The total budget for such projects will be K€300-500.

5. Contribution as founding father (together with KNAW, UU, VU, UVA and IBM) to the Center for Humanities and Technology (CHAT). This research institute integrates humanities and natural sciences to foster innovation in the humanities and includes as incubator facility for new startups (ICT create industry). Proposed total budget K€450/yr for the next four years (2013-2016). NLeSC contribution K€250/yr. Under discussion.

6. Enlighten Your Research 4 (EYR4) call for proposals organized by SURFnet, SURFsara, NLeSC, and SURF. EYR4 will focus on the general theme ‘bring your science to the next level’ and provide a selected group of winners with access to advanced e-infrastructure (computing, storage, networking), advanced software solutions and support in mapping and solving scientific problems onto the provided e-infrastructure. The framework and contribution by NLeSC is K€50.

The NLeSC project portfolio is currently divided in the following themes: Sustainability & Environment; Chemistry & Materials; Humanities & Social Sciences; Life Sciences and eScience Methodology. NLeSC’s project portfolio began with a selected, peer reviewed call to set the scientific scope of the center’s activities, to engage key scientists and Integrators from around the country and provide a basis for the transfer of converging sciences in the future.

In line with the NLeSC strategy and continuity of the existing portfolio, the following granting lines will be available for 2013:

1. A joint thematic call Data Science with NWO-EW which is linked to the Roadmap ICT Data, data, data initiative. This call focuses on new opportunities for collaboration between private companies and knowledge centers (public-private partnerships). The goal is innovation on production, management and analyses of very large datasets (Big-Data in terms of the 5V’s (volume, velocity, variety, verification and value). The call includes the Industrial Partnership Programme, Technology Area’s and Knowledge Innovation Mapping to optimally suit business demands. Total available subsidy budget approximately K€2,500, NLeSC contribution K€1,000.

2. A joint thematic open call Digging into Data with the focus on how ‘Big-Data’ changes the research landscape for the humanities and social sciences. Each project is a partnership among two or more national teams, from at least two of the four participating countries. Total call budget for the Netherlands K€550, NLeSC contribution K€100. Granted projects with a Dutch partner might receive an additional Path Finding project bonus (max. K€50) to make technology generally applicable and sustainable.

3. A selected thematic call to reinforce existing initiatives in e.g. Green Genetics, eChemistry, eScience Technology or Logistics. Discussion with eSAC on domains and call conditions planned for Q1. Total budget K€2,500.

4. A continuous selected call for Path Finding projects. These projects are intended to provide NLeSC with the opportunity to rapidly meet shorter-term scientific challenges, to address immediate technological goals as envisioned, the potential to initiate full projects. These projects have a maximum expenditure of K€50. The majority of funding should fund personnel costs and the maximum duration is one year. The project results should contribute to the eScience Technology Platform (eSTeP). The total budget for such projects will be K€300-500.

5. Contribution as founding father (together with KNAW, UU, VU, UVA and IBM) to the Center for Humanities and Technology (CHAT). This research institute integrates humanities and natural sciences to foster innovation in the humanities and includes as incubator facility for new startups (ICT create industry). Proposed total budget K€450/yr for the next four years (2013-2016). NLeSC contribution K€250/yr. Under discussion.

6. Enlighten Your Research 4 (EYR4) call for proposals organized by SURFnet, SURFsara, NLeSC, and SURF. EYR4 will focus on the general theme ‘bring your science to the next level’ and provide a selected group of winners with access to advanced e-infrastructure (computing, storage, networking), advanced software solutions and support in mapping and solving scientific problems onto the provided e-infrastructure. The framework and contribution by NLeSC is K€50.
As we develop methods and approaches to deal with the so-called data-deluge, it is imperative we remember that behind every data point is a potential story and that computing methods, however compelling, are not able to replace the creativity and imagination of scientists. Dealing with and exploiting Big-Data clearly requires the development of novel technical approaches and protocols including challenging traditional methods of humans interfacing with information.

There is a clear trend within industry to develop “war-rooms” or “cyber-commons”, providing project teams with immediate access to their data and tools for its exploration and analysis in an environment suitable for discussion and hypothesis building. As experiments become larger, more complex and increasingly multidisciplinary, we have to develop new ways to enable data-driven project work and cross-location collaboration. Cyber-commons are technology-enhanced work environments that support project dealing with Big-Data as a way of making everyday work more effective and efficient.

Cyber-commons are the modern lenses for bringing data into focus.”

“eScience is a contact sport.”

"Putting researchers in touch with their data

As an example, NLeSC, in collaboration with the Institute for Marine and Atmospheric Research (IMAU) and SURFsara, have demonstrated the potential of undertaking large-scale research across multiple disciplines and multiple sites. At the recent GLIF conference in Chicago, the NLeSC eScience integrator Prof. Henk Dijkstra of IMAU led an interactive demonstration focused on his use of eScience approaches in climatology research. Prof. Dijkstra, participated from a video conferencing suite in South Africa in partnership with Maarten van Meersbergen (NLeSC) and Michael Kliphuis (IMAU) at the Collaboratorium in Amsterdam and Frank Seinstra (NLeSC), Paul Wielinga and Tijs de Kler (SARA) at the EVL in Chicago. This eScience experiment demonstrated the importance of collaboration across disciplines and organisations for modern research, and the need for infrastructures to facilitate these interactions.

The demonstration on the same topic was given by Jason Maassen, eScience engineer, during the March 2013 visit of Prince Willem-Alexander to the Collaboratorium. Prince Alexander explained the need to undertake these types of data-driven exploration to help understand ocean flows and ultimately ensure the safety of the Netherlands.

The Collaboratorium is open for academic and commercial partnerships and a demonstration of the potential value of collaboration between NLeSC and SURFsara.
Organization of the Netherlands eScience Center

Portfolios & Partnerships

PORTFOLIO & PARTNERSHIPS
Operations & Office

EXECUTIVE COMMITTEE
The eScience center management, scientific, technical, strategic, and business development, communication and business development, communication and office management.

eSCIENCE ADVISORY COMMITTEE
The eScience Advisory Committee is an independent scientific advisory board that plans the Dutch eScience strategy. The CEC and the eScience Advisory Committee initiate and manage the process of creating a National eScience strategy.

The first meetings of the eScience Advisory Committee will be held in early 2010.

eSCIENCE ADVISORY COMMITTEE

Chair: John van der Meer, Professor of Biochemistry, University of Amsterdam

- Wim van der Linden, University of Maastricht
- Dr. Jeroen van der Meulen, University of Amsterdam
- Dr. W. B.G. Wim Liebrand, Technical University Delft
- Dr. Patrick J.C. Aerts, University of Antwerp
- Dr. Marco de Vos, University of Amsterdam
- Dr. Dr. Jan Sen, Chairman of the Netherlands eScience Center
- Dr. Scott J. Lusher, University of Oxford
- Dr. Jeff Templon, University of California

The eScience Advisory Committee will have five working groups:

- eScience Engineers
- eScience Integrators
- eScience Project Leaders
- eScience Project Portfolio
- Operations & Office

eSCIENCE ADVISORY COMMITTEE

 Operations & Office

- Dr. Paul Tymstra, Professor of Developmental Neurobiology, University of Amsterdam
- Prof. Dr. Simon Portegies Zwart, University of Leiden
- Prof. Dr. Paul Tiesinga, University of Amsterdam
- Prof. Dr. Willem Bouten, Radboud University Nijmegen
- Prof. Dr. Jan-Dax, University of Maastricht
- Prof. Dr. Scott J. Lusher, University of Oxford
- Prof. Dr. Jeff Templon, University of California
- Prof. Dr. Scott J. Lusher, University of Oxford
- Prof. Dr. Jeff Templon, University of California
- Prof. Dr. Jeff Templon, University of California
- Prof. Dr. Jeff Templon, University of California
- Prof. Dr. Jeff Templon, University of California
- Prof. Dr. Jeff Templon, University of California
- Prof. Dr. Jeff Templon, University of California
- Prof. Dr. Jeff Templon, University of California
- Prof. Dr. Jeff Templon, University of California
- Prof. Dr. Jeff Templon, University of California
- Prof. Dr. Jeff Templon, University of California
- Prof. Dr. Jeff Templon, University of California
- Prof. Dr. Jeff Templon, University of California
- Prof. Dr. Jeff Templon, University of California
- Prof. Dr. Jeff Templon, University of California
- Prof. Dr. Jeff Templon, University of California
- Prof. Dr. Jeff Templon, University of California
- Prof. Dr. Jeff Templon, University of California
- Prof. Dr. Jeff Templon, University of California
- Prof. Dr. Jeff Templon, University of California
- Prof. Dr. Jeff Templon, University of California
- Prof. Dr. Jeff Templon, University of California
- Prof. Dr. Jeff Templon, University of California
- Prof. Dr. Jeff Templon, University of California
- Prof. Dr. Jeff Templon, University of California
- Prof. Dr. Jeff Templon, University of California
- Prof. Dr. Jeff Templon, University of California
- Prof. Dr. Jeff Templon, University of California
- Prof. Dr. Jeff Templon, University of California
- Prof. Dr. Jeff Templon, University of California
- Prof. Dr. Jeff Templon, University of California
- Prof. Dr. Jeff Templon, University of California
- Prof. Dr. Jeff Templon, University of California
- Prof. Dr. Jeff Templon, University of California
- Prof. Dr. Jeff Templon, University of California
- Prof. Dr. Jeff Templon, University of California
- Prof. Dr. Jeff Templon, University of California
- Prof. Dr. Jeff Templon, University of California
- Prof. Dr. Jeff Templon, University of California
- Prof. Dr. Jeff Templon, University of California
- Prof. Dr. Jeff Templon, University of California
- Prof. Dr. Jeff Templon, University of California
Data-Stewardship in the Big-Data Era: Taking of Care Data

Netherlands eScience Center

In this document, we propose a number of actions to ensure that science has to adapt or drown in a sea of data from disparate research groups, will require continued discussion and review, but data-driven research approaches are fundamental to modern scientific disciplines and can provide a route for discovery and discovery.

Data-driven research approaches are fundamental to modern scientific disciplines and can provide a route for discovery and discovery.

Data-driven research processes, such as cross-type data integration, data-mining, visualisation and analytics, become fundamental to all scientific disciplines and data curators in publications.

NWO should require all funded projects to cite data software and data-carriers in publications in a way that the scientific, contribution of data curation, developing tools or datasets would be able to satisfy these rules and ensure the role of data curator becomes a formal responsibility.

The precise percentage of project funds for data-stewardship requires further investigation, but for making it possible for all individuals to undertake the most of scientific disciplines and data curators in publications.

Capacity to undertake data-driven research exists throughout the scientific community.

Geographical or technical barriers, such as cross-type data integration, data-mining, visualisation and analytics, become fundamental to all scientific disciplines and data curators in publications.

NWO should require all funded projects to cite data software and data-carriers in publications in a way that the scientific, contribution of data curation, developing tools or datasets would be able to satisfy these rules and ensure the role of data curator becomes a formal responsibility.

The precise percentage of project funds for data-stewardship requires further investigation, but for making it possible for all individuals to undertake the most of scientific disciplines and data curators in publications.

Capacity to undertake data-driven research exists throughout the scientific community.

Geographical or technical barriers, such as cross-type data integration, data-mining, visualisation and analytics, become fundamental to all scientific disciplines and data curators in publications.