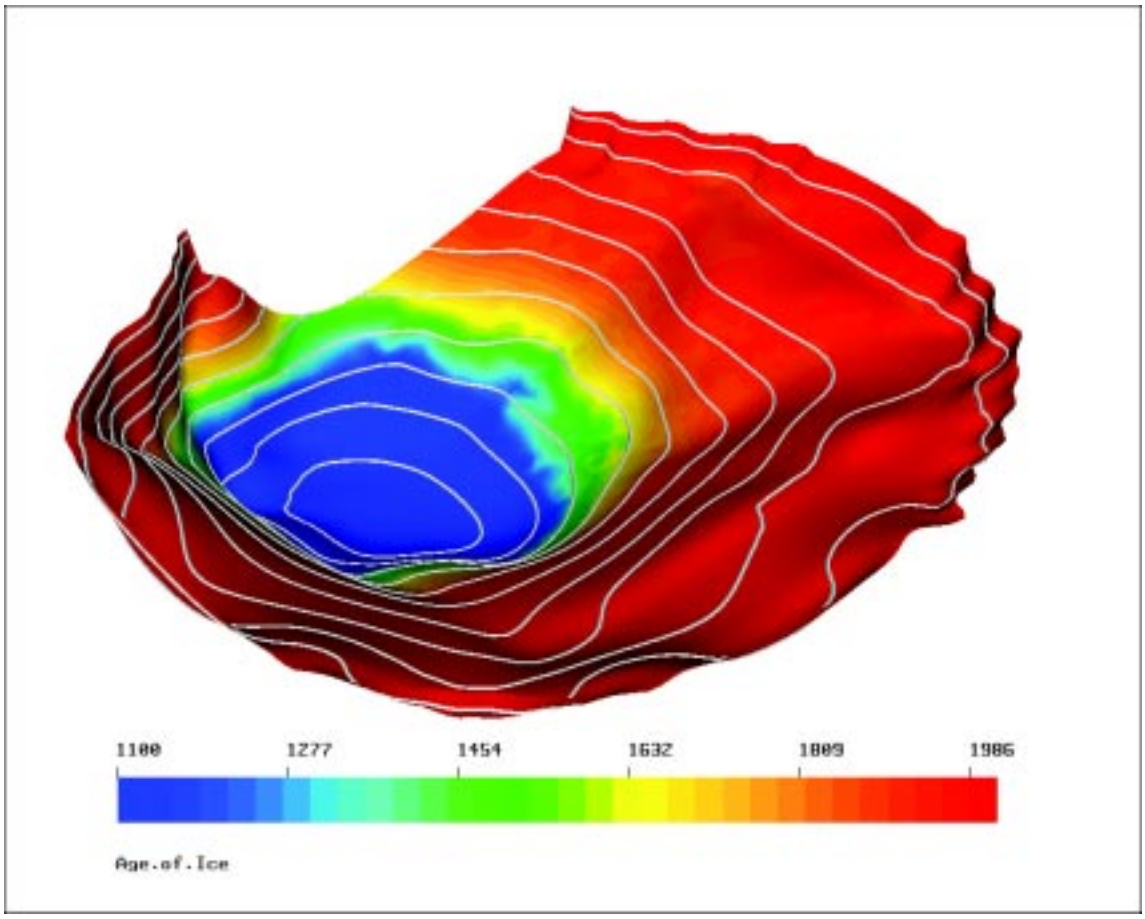
Nordic eScience

Research, Education, and Sustainable Infrastructure Services

A strategy document for the Nordic Council of Ministers

2007-07-17



Executive summary

An outstanding basis for Nordic eScience collaboration exists through the Nordic national computer networks and high-performance computing infrastructures. In addition, existing Nordic organisations such as NORDUnet and the Nordic DataGrid Facility (NDGF) prove that together the Nordic countries can act more forcefully than each country on its own. This strategy document lays out a plan for how Nordic eScience collaborations can be strengthened to provide sustainability, more efficient resource sharing, joint infrastructures and joint education programs as well as a clearer organisation for strategic planning and funding and for sharing of responsibilities.

A well designed and monitored Nordic eScience programme can in a 5 year perspective offer a well worked-out virtual system for sharing eScience resources and competences. eScience technology expertise can be visibly integrated in research groups across all fields of science, and more seamless research collaborations can be under way, which utilize a multitude of technologies across geographical borders. The competitiveness and visibility for Nordic research worldwide would be strengthened.

In a 10 year perspective Norden can be viewed as one entity and a role model internationally when it comes to virtual sharing of eScience resources. Norden can be a world leading actor in eScience infrastructure and technology development, with resulting cutting edge research in many areas of science. Nordic industry can have access to the technology experts of the 21st century, skilled with the advanced computing technologies that are the foundation for service-oriented business models and a knowledge-based industry.

eScience is about global collaboration in key areas of science, and about the next generation of infrastructures that will enable it. eScience infrastructures include computer networks, high-performance computing and visualization systems, federated databases, and networked-enabled research instrumentation. It also includes the grid - the distributed computing technology that provides access to remote resources and enables collaborations among distributed virtual organizations. Such collaborations, which require very large data collections, very large-scale computing resources and high performance visualisation facilities, will revolutionise the way science is undertaken in the twenty-first century.

eScience is about global collaboration in key areas of science

eScience involves all disciplines, from physics and astronomy to the medical sciences,

earth sciences, biology and chemistry, social sciences and humanities. For the latter, eScience opens up new avenues for collaborative and large-scale research with a global perspective. Aspects of eScience typically arise in all stages of the research cycle, emphasizing the need for close collaboration between eScience and subject matter scientists.

In the later part of 2006, The Nordic Council of Ministers formed an eScience workgroup assigned to propose a common Nordic eScience strategy. In response to this request, the workgroup emphasizes *the need for long term sustainable collaborations to maintain and develop Nordic eScience infrastructures such as computer networks, High Performance Computing (HPC) resources and Grid infrastructure*. The workgroup further proposes to establish three new programmes to promote Nordic eScience research and infrastructures and two new Nordic eScience education programmes:

- *A programme for resourcing computational "grand challenge" research.*
- *A programme for establishing a Nordic infrastructure for databases and data repositories.*
- *A programme for Nordic eScience collaboration with The Baltics, Europe and beyond.*
- *Nordic Master and PhD education programmes on eScience technologies.*

In order to provide direction and funding, the workgroup suggests that a new Nordic strategic eScience committee is established, with members that are closely associated with the national research councils and funding bodies and which represent all research fields. The committee should be placed under an existing suitable organization, such as NordForsk.

eScience involves all disciplines, from physics and astronomy to the medical sciences, earth sciences, biology and chemistry, social sciences and humanities

Content

1 Introduction

eScience supporting a knowledge-based society

The society in the 21st century is facing an ever-increasing globalization and pressure on economic competitiveness, resulting in a move from a more physically-based to a more knowledge-based economy. A well educated population, strong research environments, ever ingenious innovation, and a highly advanced information and communication technology (ICT) infrastructure, with accompanying applications, are crucial to this shift. Besides serving as a driving force for the welfare of the society in general, the research communities have to be competitive actors on the global research market.

eScience is about global collaboration in key areas of science, and about the next generation of ICT infrastructures that will enable it. Such collaborations will require advanced ICT infrastructures, access to very large data collections, very large-scale computing resources and high performance visualisation facilities. eScience will change the dynamics of the way science is undertaken in the twenty-first century. The future challenges lie in multidisciplinary areas and eScience is the natural enabler for this.

The Nordic dimension

In a global perspective the Nordic countries share a common cultural heritage and a high awareness of the need for technology development and access to good infrastructures. There is a general understanding that investment in technology and infrastructures, properly planned and pulled off, is cost-effective. Each country on its own has limited resources to tackle the challenges that face science and the society, but together the Nordic countries have a potential to forcefully define and drive the emerging field of eScience, to the benefit of research in all research fields.

Together the Nordic countries have a potential to forcefully define and drive the emerging field of eScience, to the benefit of research in all fields of science.

It is our ability to efficiently identify and transfer the new eScience tools to relevant applications, both short and long term, that will determine the competitive advantage for the Nordic countries in creating and adapting new technologies and software. Examples range from R&D intensive grid services linking together research infrastructure/facilities (telescopes, archives, databases etc), sharing ICT processing and storage capacity on a global scale, across to wider societal ICT applications like telemedicine, digital libraries, distance education, eGovernment services etc. At the same time the common Nordic knowledge, generated through common research activities will spur industry on to create new products for the global market.

A Nordic eScience strategy

In the later part of 2006, The Nordic Council of Ministers (NCM) formed an eScience workgroup assigned to propose a common Nordic eScience strategy. The group members are presented at the end of the document, and the NCM letter initiating the work is attached. This document is the workgroup's response to the initiative. The process of establishing the strategy has included 4 full-day meetings, one telephone conference and email communication. A scientific secretary to the group was appointed early in the process. In summary, the workgroup emphasizes the need *for long term sustainable collaborations to maintain and develop Nordic eScience infrastructures, such as computer networks, High Performance Computing (HPC) resources and Grid infrastructure.* The workgroup further proposes to establish three new programmes to promote Nordic eScience research and infrastructures and two new Nordic eScience education programmes. We briefly introduce each programme in this section, and return in sections 4 and 5 to existing infrastructures and to more detailed descriptions of the programmes.

> eScience opens up new arenas for research and development, such as telemedicine, digital libraries, distance education, eGovernment services etc.

- *A programme for resourcing computational "grand challenge" research*. A Nordic Computational Grand Challenge Survey has recently identified important research areas with extraordinary needs for IT-resources. Sample areas include physics, astrophysics, climate computing, molecular modelling, atmosphere modelling, tsunami propagation, protein structure refinement, biological simulation, oceanography, simulations of complex liquids and computations for nano-particles and nano-material. The new programme proposed here should open up for broader eScience grand challenges, which cover the full research cycle, and which gain added value from Nordic collaboration and coordination. Examples include access to sufficient computing resources, development of new models, algorithms, and software tools to solve increasingly complex general and application-specific problems, solutions to a range of data locality and parallelisation problems, development of application-oriented grid tools, etc.

- *A programme for establishing a Nordic infrastructure for databases and data repositories.* A unique feature for the Nordic countries is the comprehensive databases and data repositories carried by national authorities for administrative or monitoring purposes. These unique data structures are currently underused for research. Enabling secure and efficient modes for accessing and processing these data sources, and for merging them with research data from other sources, opens up for new and unique research data infrastructures in the life sciences, earth sciences, social sciences and medicine - research areas which have not traditionally been technology driven. Besides implementation and development of technology for federated databases and data repositories, huge efforts are needed to set up policy frameworks for standards, data harmonisation and access, including administrative and legal components. Harmonisation on the Nordic level has a clear advantage.

- *A programme for Nordic eScience collaboration with The Baltics, Europe and*

beyond. The development of eScience research and infrastructures is not only a national or a Nordic matter, but a global issue requiring a substantial degree of international collaboration. An illustrative example is the development of the grid as a key enabling technology for large-scale resource sharing. For the purpose of interoperability and sustainability, any joint Nordic effort must also be coordinated with European and other international initiatives. This programme provides an interface between the Nordic eScience development and similar ventures within Europe and beyond. The infrastructures proposed in the ESFRI roadmap and related research challenges in the EU framework programmes serve as concrete examples.

- *Nordic Master and PhD education programmes on eScience technologies.* The proposal is to establish Master and PhD programmes, which build on current successful national initiatives. These education programmes are best administered within the framework of the Nordic Master Programme, hosted by the Nordic Council of Ministers and the NordForsk research schools, respectively. Comprehensive Nordic eScience education programmes give Nordic research a broad competitive advantage and provide Nordic industry with the technology experts of the 21st century, skilled with the advanced computing technologies that are the foundation for emerging service oriented business models and knowledge-based industry.

By coordinated initiatives, the Nordic countries have a unique opportunity to gain comprehensive coverage of vital eScience topics and attract a critical mass of researchers. Coordination across research fields and national borders maximizes the speed by which new methodologies can be transferred to a broad range of eScience applications. Cross-national resource sharing will allow more resources to individual research projects, without imposing on other needs. The increased international visibility and increased total resource volume will give Nordic researchers added opportunity to take on leading roles in international high-profile collaborations. Comprehensive eScience education programmes provides for a broad competitive advantage on the eScience arena for Nordic research.

Nordic eScience needs sustainability, more efficient resource sharing, joint infrastructures, joint education programs and a clearer organisation and division of responsibility.

2 Driving forces for eScience research and infrastructure

Common for the whole field of eScience is the demand for new and improved high-quality software including user-friendly application software, software libraries, middleware, and user environments. Below, we describe key forces that drive the need for new research and infrastructure in the field of eScience.

Changing needs in the broad spectrum of eScience application areas. The progress in scientific areas with a longer tradition of using eScience technology puts demands on new models, algorithms, and technology to solve even more complex problems. The Nordic Grand Challenge Survey [9] has identified important research areas with extraordinary needs for computing resources. Moreover, the recent move towards using eScience techniques in research areas with little previous experience of large-scale eScience infrastructures represents a new and important challenge. These application areas, for example in the life sciences, social sciences, and humanities, require tools to solve new computational, data management, and visualization problems.

The development of computing technologies. Trends in the evolution of basic computing technology have significant impact on eScience software development. One trend is the rapid increase in the number of processors in state-of-the-art computer systems. This trend is currently driven by the cost-effective development of high-end computer systems built from commodity components. It is expected to accelerate by the use of so called multi-core technologies, as computer developers strive to find new ways to increase computer performance as clock frequencies are becoming hard to increase. As a consequence eScience researchers will soon face desktop machines with hundreds of processors and high-end systems with tens or hundreds of thousands of processors. In addition, the increasing gap between computational speed and memory performance forces computer vendors to build deeper and more complex memory hierarchies in order to exploit the performance of the computers.

Globalisation of research. The development of international grid-based infrastructures, applications, and collaboration environments goes hand in hand with the internationalization of research. While the prerequisites for globalisation are provided by improved connectivity and network performance, and by the global infrastructure development, it is the research needs that trigger the move towards large-scale virtual research environments. This development is still in its infancy and current grid production environments are targeted to a small number of research fields and applications with particular characteristics. The need to develop the grid to meet the needs of more general and global computing infrastructures requires major research undertakings.

3 Worldwide recognition of the importance of eScience

The importance of eScience for research in the 21st century has been recognized by numerous reports worldwide. With these as a point of departure we put the Nordic strategy in an international context. The overall message conveyed by these reports is that eScience technology is of outermost importance for the further development of research across virtually all disciplines, and that the eScience field is of paramount interest to the whole society. Note that this short section has no ambition to be comprehensive.

Through its ambitious eScience programme, started in 2001 [1][6][7][12][18], the UK has lead the international promotion and development of eScience technologies over the whole eScience field:

> *"In the future, e-Science will refer to the large scale science that will increasingly be carried out through distributed global collaborations enabled by the Internet. Typically, a feature of such collaborative scientific enterprises is that they will require access to very large data collections, very large scale computing resources and high performance visualisation back to the individual user scientists."*

The UK eScience programme has not only coined the name "eScience", but also proposed the grid as the architecture to bring eScience components together.

Computational science is a key part of eScience, today widely recognized as the third pillar of the scientific method, complementing the classical pair: theory and experiment. The following statement from the often cited report *Computational Science: Ensuring America's Competitiveness* produced by the US President's Information Technology Advisory Committee (PITAC) [20] is easily generalizable to eScience *per se*:

> *"While it is itself a discipline, computational science serves to advance all of science. The most scientifically important and economically promising research frontiers in the 21st century will be conquered by those most skilled with advanced computing technologies and computational science applications."*

The PITAC report is only one of several reports that advise strong commitments to eScience research and infrastructure. Some other US examples are produced by The Executive Office of the President [5], The Office of Science [11] and The National Science Foundation (NSF) [13], The NSF Cyberinfrastructure Council [10] and The Towards 2020 Science Group [19] (the latter actually performed in the UK, but for Microsoft Inc.).

On a European level, both the e-Infrastructure Reflection Group (e-IRG) [2] and the HPC Europe Task Force (HET) show commitment to international collaboration for coordinated international e-Infrastructures. e-Infrastructures and the global dimension of research have also been highlighted in the recent roadmap produced by the European Strategy Forum on Research Infrastructures (ESFRI) [3]. It is worth emphasizing that the 35 projects put forward in the ESFRI roadmap cover all areas of science, and most of them explicitly or implicitly include eScience components. In order for the development of pan-European infrastructures, as laid out in the ESFRI roadmap, to be efficient, a broad unified conceptual framework and action plan regarding eScience infrastructures, education, research and applications would be needed. Notably, the Nordic countries are well represented in several of the FP7 infrastructure planning proposals currently being prepared.

The Nordic countries have, individually, addressed the importance of the eScience area in a number of recent initiatives, e.g., the eScience programme in Finland 2007 [14], the eVITA programme in Norway 2006 [4][8], and the Swedish Research Council's formation of an eScience infrastructure group 2005 [17], and a proposed research programme by the Swedish Infrastructure for Computing (SNIC) 2007 [15][16]. While the Nordic countries have a long history of successful collaboration in computer networking, grid software development and more recently in challenges concerning high performance computing, it is only now that the discussions start on how to meaningfully coordinate key features of the scientifically broader Nordic national eScience initiatives and on how to guarantee long-term sustainability.

It is only now that the discussions start on how to meaningfully coordinate key features of scientifically broader Nordic eScience initiatives and on how to guarantee long-term sustainability.

4 Sustainability for existing Nordic eScience infrastructures

In this section, we briefly present the major eScience infrastructures currently available in the Nordic countries and we point out opportunities for intensified Nordic collaboration. The need for sustainability is emphasized, since long-term availability is fundamental for any large-scale and long-term research

4.1 Computer networks

The Nordic national University networks provide high-quality and reliable computer network services that are fundamental to the eScience development. The availability of networks with internationally competitive performance is vital, and recent initiatives for offering dedicated so called lambda connections are of interest to the eScience community.

The Nordic coordination of networks and network-related activities is already well organized, mainly though the NORDUnet collaboration. The fibre optical NORDUnet infrastructure currently links four of the five Nordic countries. An upgrade of the currently rather unstable and insufficient network links to Iceland are, however, needed, and appropriate connectivity to global partners like Russia, The Baltics, China, Japan and others needs to be continuously ensured through the NORDUnet collaboration.

4.2 HPC resources

The five Nordic countries have organized their main High Performance Computing (HPC) resources in slightly different ways. Finland has one main national center in the Finnish IT Center for Science (CSC). Denmark, Norway and Sweden have different types of decentralized national organizations, represented by the Danish Center for Scientific Computing (DCSC), the Norwegian Notur and the Swedish National Infrastructure for Computing (SNIC), respectively. The Icelandic HPC resources are maintained by the University of Iceland. Notably, although all countries have adopted different organizations, there is a well-defined contact point per country, e.g., for Nordic, European and global collaborations.

Well-organized cross-national resource exchange with appropriate funding allocation between the Nordic countries will allow more efficient resource utilization for the HPC organizations, and individual users will get access to a larger and more diverse set of resources. Moreover, by acting as a coherent virtual body, the Nordic HPC resources and competences

will have increased visibility internationally and will be very well equipped to take part in large-scale European (and otherwise international) collaborations.

We propose a substantial degree of collaboration between the Nordic HPC organizations, including technology and policy development to set up procedures for cross-national resource allocation and accounting.

4.3 Grid infrastructure

Grid infrastructures[1] are now available in all five Nordic countries. In Denmark, the Danish Center for Grid Computing (DCGC) was established in August 2003 with a three year grant for a series of research and development projects. There are currently plans to make DCGC a sustainable part of the Danish HPC organization, DCSC. The Finnish Material Sciences National Grid Infrastructure (M-grid) is a joint project between the Finnish HPC Center, CSC, seven universities and the Helsinki Institute of Physics, giving access to a distributed set of computers. In Iceland, a grid infrastructure is under construction as part of the University of Iceland's participation in Nordic grid activities. In February 2007, the Research Council of Norway granted funding for the establishment of a national grid infrastructure (NorGrid). SweGrid is a Swedish national grid infrastructure project hosted by the Swedish HPC organization, SNIC. The long-term plan for SweGrid is to give grid access to all national resources at SNIC centers.

Joint Nordic initiatives include the recently initiated Nordic DataGrid Facility (NDGF), which is a collaboration between Denmark, Finland, Norway, and Sweden. NDGF is a production grid facility that leverages existing, national computational resources and grid infrastructures. NDGF currently coordinates the establishment of a so-called Tier-1 for storing and processing large amount of data produced at CERN.

In order to meet expanding needs for new grid applications, the Nordic countries will have to provide a grid infrastructure which is sustainable both in terms of technology and capacity. It is imperative that the development is made through Nordic and international collaboration, in harmony with international standardization initiatives. The collaborations between major Nordic grid infrastructure providers (e.g., national HPC organizations, NDGF, national research database organizations, etc) should be continued and enforced, with the goal to provide transparent and secure access to all eScience resources, including capacity and capability computers, storage, software, databases, collaboration environments, online instrumentations, etc, addressing the differing needs of all eScience application areas.

[1] The grid infrastructure provides access to distributed IT resources and enables resource sharing across traditional organizational borders. Ideally, the grid will present the distributed IT resources to the user via a single system image, and provide the user with transparent access to all resources that he or she is entitled to use. The concept "grid" is analogous to "power grid", in an attempt to compare access to IT resources with today's easy access to electricity. The user does not need to know where and how electricity is produced. Such a computational grid infrastructure requires, for example, distributed security mechanisms, resource management and resource brokering tools, data management tools, information systems and support for establishing resource sharing policies. The grid of today is still in many aspects experimental, and its use is dominated by applications which require very large computational resources. Substantial developments are needed to explore the full potential of grid technologies in all eScience application fields.

5 Proposed new Nordic eScience programmes

The workgroup proposes to establish three new programmes to promote Nordic eScience research and infrastructure and two new education programmes. The maintenance and development of sustainable infrastructure collaborations that were discussed in Section 4 will in some parts feature also in this section.

5.1 Resourcing computational "grand challenge" research

A Nordic Computational Grand Challenge Survey [9] has recently identified important research areas with extraordinary needs for IT-resources. Sample areas for this first grand challenge initiative include physics, astrophysics, climate computing, molecular modelling, atmosphere modelling, tsunami propagation, protein structure refinement, biological simulation, oceanography, simulations of complex liquids and computations for nano-particles and nano-material. The programme proposed here should open up for broader eScience grand challenges, which cover the full research cycle, and for which Nordic collaboration and coordination provides added value.

Computational "grand challenge" research requires access to sufficient computing resources, development of new models, algorithms and software tools, solutions to a range of data locality and parallelisation problems, development of application oriented grid tools, etc. Nordic collaboration allows cost-efficient development of technologies and use of resources.

In order to address computational grand challenge problems, it is necessary to provide resources and tools for solving more demanding computational problems. Besides access to tools for storage, computing and visualization, grand challenge problems will generate a number of generic and application-oriented challenges, broadly exemplified by the following items:

- Scientific computing research, e.g., including the development of new algorithms for managing more complex computational geometries and for performing simulations on more complex computational models, development of efficient mathematical software libraries and environments for performing fundamental computational operations, and research on the efficient use of programming languages. The development of computer technology also leads to a range of data locality and parallelisation problems that need to be solved in order to fully exploit the potential of the future computing platforms.

- General grid computing research and development of application-oriented grid tools. This includes, for example, the development of application portals and problem solving environments, application-aware workflow systems and workflow repository development, support for interactive grid applications as well as applications with need for specific online instrumentation.

- Efficient utilization of eScience database infrastructures presents a broad spectrum of challenges, including research on federated database systems, technologies for secure access and data sharing and efficient procedures for structuring and annotating databases with content of widely differing character. Notice that the technical tools referred to here complement the administrative Nordic database infrastructure programme discussed in the next subsection.

The joint Nordic programme for resourcing computational “grand challenge” research brings added value in the form of:

- Extended coverage and critical mass to pursue vital eScience research topics. Solving large-scale grand challenge problems are best done by cross-national research projects.
- Better targeted interactions with application areas.
- Cost-effectiveness in utilizing computational resources through cross-national resource sharing.
- Improved international visibility.

5.2 Establishing a Nordic infrastructure for databases and data repositories

Well-maintained databases, satisfying strict requirements on quality, documentation, standards, meta-data definitions, accessibility, security, etc, will be the principal eScience resources for many application areas. In the same spirit as molecular technology has transformed biology into a computational science, ICT technology and software can pave the way for large conceptual changes in areas of science that rely on complex non-experimental data-structures. While large-scale standardized and quality controlled data infrastructures are beginning to emerge in biology, such infrastructures do not exist on any large and comprehensive scale in the health sciences, earth sciences, social sciences and humanities. In order to build these infrastructures, the proposal is to establish a programme focused on the establishment of a Nordic database and data repository infrastructure.

The comprehensive national registers and databases carried by authorities for administrative or monitoring purposes constitute a unique feature for the Nordic countries. These data structures are currently underused for research. Technical solutions for enabling secure and efficient modes for accessing and processing these data, and for merging them with other sources of research data are discussed in the previous subsection under grand challenges. Efficient utilization of these data sources, opens up for new unprecedented research opportunities in the life sciences, earth sciences, social sciences and medicine - research areas which have not traditionally been technology driven.

In an international perspective Norden has unique data sources in the form of registers and data repositories. A common Nordic data infrastructure allows for unprecedented cutting edge research in the life sciences, earth sciences, social sciences and humanities.

The work on establishing databases and data repositories as national infrastructures is currently in an early phase in the Nordic countries. When it comes to organizational issues, Sweden has probably reached the farthest by establishing a Database InfraStructure Centre (DISC) as a national infrastructure. Norway's national data archives in the social sciences, manifest in the Norwegian NSD, serve as excellent proof of concept. Moreover, Statistics Denmark and Statistics Sweden provide researchers with external access to de-identified micro-data over the Internet. This recent change from on-site to remote access is a powerful example of how new technology can facilitate the way research is carried out. A key feature in building database infrastructure is to allow quality research databases to be maintained by a number of different organizations, ranging from major national and international organizations to local research groups making their data available. The need for coordinated but not centralized solutions is crucial, with the meta-data definitions and ontologies allowing links between different data-structures on different media. Considerable effort is needed to resolve important and difficult policy issues related to administrative and legal aspects of data-sharing, to ownership of data and to secure solutions for maintaining individual integrity.

While the bulk of the work on setting up databases must be done through national efforts and funding, the added value would be dramatic for science, industry and the society at large from Nordic cooperation on data harmonization and on setting up policy frameworks that allow cross-talk over research fields and national borders. Now, before the national solutions are in place, is the right time to initiate this work.

The joint Nordic programme for establishing an infrastructure for databases and data depositories brings added value in form of:

- Clear identification of actors responsible for data-harmonisation and policy decision (ethics and legal issues, open source policy) on the national and Nordic levels, respectively.
- Increased coordination and homogenization, which facilitates the use of multiple data sources (across disciplines and national borders).
- Improved international visibility.

5.3 Nordic eScience collaboration with The Baltics, Europe and beyond

The research and development of eScience infrastructure is not only a national or a Nordic matter, but a global issue requiring a substantial degree of international collaboration. An illustrative example is the development of the grid as a key enabling technology for large-scale resource sharing. For the purpose of interoperability and sustainability, any joint Nordic efforts must also be coordinated with European and other international initiatives. This view is further established by the roadmap of the European Strategy Forum on Research Infrastructures (ESFRI) [3], which has identified 35 important research infrastructures of European interest, most of them explicitly or implicitly including eScience components.

> The Nordic countries one by one are participating in the preparatory phase of the European FP7 ESFRI research infrastructures programme. A faster and more efficient process on the national level would result if Norden acted together in Europe and beyond.

This being the motivation, the proposal is to establish a programme for Nordic collaboration on research and development of new eScience infrastructures. The collaborations in focus include both Nordic efforts and Nordic participation in Baltic, European and wider international initiatives, e.g., targeting EU-infrastructures proposed in the ESFRI roadmap and related research challenges in the EU framework programmes. It should be mentioned that although the Nordic countries are already involved in such European activities, the Nordic position can be significantly strengthened by Nordic collaboration and common Nordic initiatives. Examples of ESFRI infrastructure initiatives, with strong eScience components and with Nordic national activity in the preparatory phase, include European high-performance computing (PACE), bioinformatics (ELIXIR), biobanks and biomedical resources (BBMRI) and language technology (CLARIN). If national efforts were coordinated from the start, cost-efficiency gains are expected for the individual countries. As a proof of concept the Nordic countries have experience in joint participation in large-scale EU infrastructure projects in the past, such as the Enabling Grids for EsciencE (EGEE) project, where Denmark, Finland, Norway, and Sweden, together with the Netherlands, form a Northern node. The NDGF coordination of a Nordic Tier-1 for storing and processing data produced at CERN is another example of ongoing joint Nordic participation in large-scale EU projects.

The joint Nordic program brings added value in form of:

- A more cost-effective participation in projects funded by EU and internationally.
- Ability to participate in a larger number of projects funded by EU and internationally, by distributing the workload and cost for co-funding.
- Ability to take on more leading roles in projects funded by EU and internationally, as a stronger partner.

5.4 Nordic Master and PhD education programmes on eScience technologies

Joint Nordic eScience Master and PhD education programmes allow efficient use of resources and critical mass to target different student backgrounds and application areas.

The need for competence building in the development and use of eScience technologies is common for all Nordic countries. Comprehensive training and education programmes are the key to the long-term success of any effort to establish eScience infrastructures in broad scientific areas. These efforts should include a mix of students from technology core areas such as computer science and the mathematical sciences, as well as students from the application areas. Students from the application areas are expected to have varying background in basic computing techniques, and the education programmes must adapt to the students' differential needs. By coordinated Nordic efforts one can attract sufficiently large groups of students to offer education programmes tailored to specific backgrounds in a way that would not be possible for local or national programmes.

Therefore, the proposal is to establish Master and PhD education programmes in eScience technologies. These programs would be conveniently organised within the framework of the Nordic Master Programme run by the Nordic Council of Ministers and the NordForsk research school programme, respectively. The establishment of comprehensive Nordic eScience education programmes, which build on successful national initiatives, will provide a long term competitive advantage for Nordic research at large. Examples of relevant national eScience education initiatives are the Swedish National Graduate School in Scientific Computing (NGSSC), the new eScience MSc programmes at the University of Copenhagen and the University of Iceland, and the Norwegian Winter School on eScience.

Joint Nordic Master and PhD education programmes bring added value in form of:

- A larger student basis, which makes it possible to develop eScience technology courses targeting different student backgrounds and application areas.
- Broader competence for course development and distribution of teaching load.
- Nordic collaboration among students, which builds the Nordic networks of tomorrow and paving the way for future cross-national research collaborations.

6 Industrial impact

Although academia is the main focus of the proposed eScience strategy, it has important impact both on and from industry. The immediate and most important benefit for industry, from strengthening the eScience infrastructure, research, and education is competence development. Competences in core eScience infrastructure areas are highly relevant to the industrial sector. Notably, the seventh EU framework programme has a strong focus on service-oriented infrastructures for industrial and business use. Here the solutions will to a large extent be realized by extending on grid technology originally developed by the eScience community.

The single most important factor for industry in the much needed boost of the eScience competences is the recruitment of highly educated personnel, with extraordinary skills in the use and development of eScience technologies. The broad approach to eScience taken by the working group implies that not only the industrial sector with traditionally high HPC needs will be involved, but also the whole range of emerging knowledge-based companies leveraging, e.g., eScience data-mining techniques. Moreover, competence in Grid technology will be important for the development of service oriented infrastructures for business applications.

There are also ways in which industry can play an important role for improving eScience. Once again, the most important aspect is education. One such role is to contribute with an individual industrial mentor for each PhD student. For the MSc programme, the single most important industrial impact would be from contributing with master's thesis projects and their guidance. Moreover, it should be investigated to what extent industry is willing to take part in lecturing on courses, or even in the development of complete courses tailored for industrial preparation. By such participation in the MSc and PhD educations, industry will also be better positioned for recruiting the eScience students when graduated.

In addition, joint research and development projects between industry and academia will contribute to the competence development. Such collaborations can be based on individual projects or formalized industrial-academic consortia of more long-term nature. The focus can, e.g., be on major challenges with clear business benefits or further development of academic results and prototypes into quality tools for industry. Such collaborations will also create networks that substantially strengthen the competitiveness of both academia and industry, e.g., in EC-funded framework programmes.

7 Organization

In order to establish and monitor the Nordic eScience strategy, a new *Nordic strategic eScience committee* is proposed, with members from national research councils and funding bodies. All fields of research shall be represented in the new committee. In contrast to the working group, which has had little capacity to control money flows, the new committee must commit to provide detailed direction and monitoring, and to secure funding from national and Nordic sources.

The new strategic committee should be placed under an existing suitable organization, such as NordForsk, leverage existing well-established Nordic organizations such as Høgut, NORDUnet, and NDGF, and utilize expertise from various national centers and organizations working in the HPC and grid domains. The committee should not be operational. Instead, all practical tasks should be commissioned or outsourced to the best suited Nordic institution or group of institutions.

The instruction for the committee, to be formally approved by the national research councils and funding agencies should include the following items:

1. To formulate a rolling strategy for Nordic eScience, to be formally approved by the national research councils and funding bodies and to be updated regularly.

2. To work out guidelines for
   - The division of responsibilities between national actors and actors at the Nordic level concerning programme content.
   - The division of responsibilities between national actors and actors at the Nordic level concerning principles and practices of sharing programme costs.
   - How programme research and infrastructure tasks should be allocated to existing national or Nordic bodies, and how new actors could enter the scene. As a rule a competitive process should be used when allocating resources.
   - How to monitor progress from a Nordic and a national perspective.

3. To administer
   - The formulation of calls for proposals.
   - The process of reviewing proposals and allocating resources.

The strategic committee should be set up and the first call for proposals should ideally be launched during 2008, for a period of five years, with start 2009.

8 References

[1] *Developing the UK's e-infrastructure for science and innovation.* Report from the OSI e-Infrastructure Working Group.
[2] *e-Infrastructures Roadmap*. e-Infrastructures Reflection Group (e-IRG), December, 2005. http://www.e-irg.eu.
[3] *European Roadmap for Research Infrastructures*, Report 2006. European Strategy Forum on Research Infrastructures (ESFRI), 2006.
[4] eVITA. eVITenskap og Anvendelser. Forskning i en ny epoke. May 2004.
[5] Executive Office of the President of the United States. *Federal Plan for High-End Computing*, 2004.
[6] G. Fox and D. Walker. *e-Science Gap Analysis*, The UK e-Science technical Report Series, 2003.
[7] T. Hey and A.E. Trefethen. The UK e-Science Core Programme and the Grid, *Future Generation Computing System*s, Vol 18 page 1017 (2002).
[8] *Infrastructure, Theory and Application (eVITA). Work programme.* Division for Science eScience, The Research Council of Norway, September 2005.
[9] Nordic Computational Grand Challenge Survey, April, 2007.
[10] NSF's Cyberinfrastructure Vision for the 21st Century Discovery. NSF Cyberinfrastructure Council, NSF, 2007.
[11] Office of Science. *Strategic Plan February 2004*. US Department of Energy, 2004.
[12] Research Councils UK. www.rcuk.ac.uk/escience
[13] *Simulation-Based Engineering Science. Revolutionizing Engineering Science through Simulation.* Report from the National Science Foundation Blue Ribbon Panel on Simulation-Based Engineering Science, February 2006.
[14] *Suomen eScience-ohjelma (Finlands eScience-program).* Finska undervisningsministeriet, arbetsgruppen för gridstrategin, January 2007. http://www.minedu.fi/OPM/Julkaisut/2007/Suomen_eScience_ohjelma.html
[15] *Sw-eScience – Förslag till ett nationellt forskningsprogram* för metodologiska problem och frågeställningar inom eScience i ett SNIC-perspektiv, The Swedish National Infrastructure for Computing (SNIC), April, 2007.
[16] *The Swedish HPC Landscape 2006–2009. Visions and roadmap*. The Swedish National Infrastructure for Computing (SNIC), 2006.
[17] The Swedish Research Council's Guide to Infrastructure, June, 2006.
[18] John Taylor. e-Science definition. http://www.e-science.clrc.ac.uk/
[19] *Towards 2020 Science*. Report by the Towards 2020 Science Group for the Microsoft Corporation, 2006.
[20] US President's Information Technology Advisory Committee (PITAC). *Computational Science: Ensuring America's Competitiveness*. Report to the President, June 2005.

Group members

Representatives for NOS organisations
Juni Palmgren, Professor (*Chair of the NCM eScience ad hoc workgroup*), Department of Mathematics, Stockholm University, Stockholm (NOS-N).
Hege Torp, Director, Department for the Social Sciences, The Research Council of Norway, Oslo (NOS-HS).
Kalervo Väänänen, Professor, Institute of Biomedicine, University of Turku, Helsinki (NOS-M).

Representative for Nordforsk
Torkild Vinther, Senior advisor, NordForsk, Oslo.

Representative for HØGUT
Rolf L. Larsen, Deputy Director General, Norwegian Ministry of Education and Research, Oslo.

Representative from the e-Infrastructures Reflection Group (e-IRG)
Leif Laaksonen, Development Director, Finnish IT Center for Science (CSC), Espoo.

Representatives from Industry
René Buch, CEO, NORDUnet A/S and NDGF, Kastrup (Industry representative).
Kimmo Koski, Managing Director, The Finnish IT center for science (CSC), Espoo (Industry representative).
Karsten Vandrup, Senior Research Manager (*Vice chair of the NCM eScience ad hoc workgroup*) Nokia Technology Platforms, Copenhagen.

Representatives from the Nordic Council of Ministers
Rene Belsø, Senior Adviser, Nordic Council of Ministers, Copenhagen.
Kim Bärlund, Senior Adviser, Nordic Council of Ministers, Copenhagen.
Gard Titlestad, Head of Department of Education, Research and Labour Market, Nordic Council of Ministers, Copenhagen.

Group members appointed by the NCM eScience ad hoc workgroup:
Erik Elmroth, Associate Professor (*Scientific Secretary of the NCM eScience ad hoc workgroup*), Department of Computing Science and HPC2N, Umeå University, Umeå.
John Renner Hansen, CEO, Niels Bohr Institute, Copenhagen.
Ebba Thora Hvannberg, Professor, Computer Science Department, University of Iceland, Reykjavik.

Appendix

The NCM letter initiating the eScience workgroup

Nordic Council of Ministers

Store Strandstræde 18
DK-1255 Copenhagen K
Tel +45 3396 0200
Fax +45 3396 0202
www.norden.org

NordForsk
Attn.: Director Liisa Hakamies-Blomqvist (liisa.hakamies-blomqvist@nordforsk.org); Chair Lene Lange (lla@novozymes.com)

The Joint Committee of the Nordic Natural Science Research Councils (NOS-N)
Attn.: Head of Section, Anders Kjær (AKJ@fist.dk); Chair Lars Stemmerik (ls@geus.dk)

The Joint Committee for Nordic Research Councils for the Humanities and the Social Sciences (NOS-HS)
Attn.: Susanna Sepponen (Susanna.Sepponen@aka.fi); Chair Eila Helander (eila.helander@helsinki.fi)

The Joint Committee of the Nordic Medical Research Councils (NOS-M)
Attn.: Research Officer Teresa Karlsson (teresa.karlsson@vr.se); Chair Håkan Billig (hakan.billig@vr.se)

20 September 2006
/RB

A Nordic eScience strategy

The Nordic Council of Ministers (NCM) has decided to seek to establish a common Nordic eScience strategy. If to succeed, namely in terms of implementation, it is believed that such a strategy must be established in cooperation with the major Nordic research stakeholders. These are seen to be the new Nordic research organization, "NordForsk", as well as the five national research councils, via their Nordic cooperation committees (NOS Committees)

NCM, therefore, hopes that NordForsk and the NOS committees will cooperate with NCM in drafting a Nordic eScience strategy, and engage in dialog as to its later implementation.

Defining eScience[4]: eScience will in this context refer to the large scale science that will increasingly be carried out through distributed global collaborations enabled by the Internet. Typically, a feature of such collaborative scientific enterprises is that they will require advanced ICT infrastructure, access to very large data collections, very large scale computing resources and high performance visualisation back to the individual user scientists. eScience is about global collaboration in key areas of science, and the next generation of ICT infrastructure that will enable it.

It is understood that eScience will noticeably change the dynamics of the way science is undertaken in the twenty-first century.

The background for the proposal about a common Nordic eScience strategy takes its point of departure in the following rational:

- The Nordic countries are all facing an ever increasing globalization and pressures on economic competitiveness. Hence there is a growing understanding that one must move from a *physically-based* to a *knowledge-based* economy. A well educated population, strong research environments, more, ever shrewder innovation, and a very advanced information and telecommunications (ICT) infrastructure with accompanying application, are all thought to be crucial to such a shift.
- Focusing on the later – *eScience*: The Nordic countries are all among the most advanced countries in the world[5] in respect to proliferation, understanding and

[4] Source: Cambridge eScience Centre

[5] The World Economic Forum's Networked Readiness Index (NRI): http://www.forbes.com/technology/2005/03/09/cx_0309wefranking.html; Nordic Information Society Statistics 2005: http://www.norden.org/pub/sk/showpub.asp?pubnr=2005:562

usage of almost any king of ICT – namely the Internet and advanced Internet applications.

- The Nordic countries, therefore, all face the same worrying challenges and have more or less the same favorable preconditions to meet them. In addition, the Nordic countries are all relatively small and share a common cultural heritage, adding extra meaning to Nordic cooperation in the field of eScience – like is already the case within research, via the new NordForsk, and ICT infrastructure and Internet research, via the well established Nordunet-cooperation[6] on ICT infrastructure and Internet research. The Nordunet-cooperation gives a strong platform for collaborative computing.
- Hence, the role of ICT in propelling the shift to a knowledge-based economy is thought to be of paramount importance for the Nordic countries. Not just in terms of ICT *infrastructure*, but more importantly as an enabling factor for advanced ICT *applications*, which will serve not only research but also society at large. Examples range from R&D intensive *GRID services* linking together research infrastructure/facilities (telescopes, archives, databases etc), sharing ICT processing and storage capacity on a global scale, across to wider societal ICT applications like telemedicine, digital libraries, distance education, eGovernment services etc.
- All countries, especially the Nordic, are facing scientific challenges that can not be tackled by one single country alone, but by sharing responsibilities and efforts. Hence, experts call for an open access data repository policy that opens access to resources across both thematic and country borders. Technologically it can be done already, but it is politically complicated.

It is against this background that NCM would like to establish a common Nordic eScience strategy. If you agree to participate, we suggest that a time limited *ad hoc eScience workgroup* is formed, which is given the task of drafting the strategy, for later adoption. It is suggested that the group consists of:

- One representative from each of NordForsk, NOS-N, NOS-HS, NOS-M,
- Two representatives from industry, since eScience has significant societal implications and direct relevance to national competitiveness,
- One representative from The Nordic Advisory Committee on Higher Education (HØGUT), since eScience also has strong links to higher education
- One representative from European e-Infrastructures Reflection Group (e-IRG), since it is important that the Nordic cooperation does not duplicate, but strengthens, equivalent efforts at the European level.

In regard to implementing the strategy, it is essential to also draft:

- An initial action plan, including distribution of responsibility and a time frame,
- A tentative budget with budget distribution,

We therefore suggest that the parties name representatives to the *ad hoc eScience workgroup*, and finance their participation (i.e. travel and accommodation). Further, we suggest that the NCM secretariat contributes with the secretariat function to the group, and drafts a first version of the strategy (discussion paper), based on the first replies to this request. As secretary to the group we suggest NCM senior advisor Rene Belsø (RB@norden.org; mobile phone +45.29692951), whom you may contact regarding any questions. The eScience strategy draft can then be discussed in more detail at a later meeting in the group. We suggest the first meeting to be **4 December 2006** at 10:00 – 15:00 in the NCM secretariat, Copenhagen.

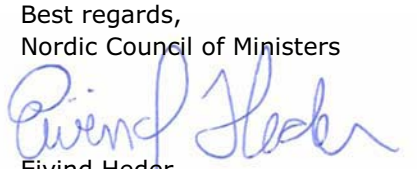
NCM can inform that we have received a letter from NOS-N with a suggestion, which is closely related to the suggestions forwarded here. We enclose the letter for your information, and regard the suggestions mentioned here as an answer.

---

[6] The Nordunet cooperation is Nordic cooperation on IT infrastructure for research (www.nordu.net) as well as research within Internet technologies (www.nordunet3.org).

NCM would appreciate your reply at your earliest convenience and preferably no latter then **21 November 2006**.

Best regards,
Nordic Council of Ministers



Eivind Heder,
Chair, Nordic Committee of Senior Officials for
Educational and Research Issues

Enclosed: Letter from NOS-N dated 2006.08. 17, regarding: *Nordic eScience cooperation.*

Front cover: Recently CSC's in-house multi-physics package Elmer has been applied to simulations of glaciers and ice-sheets. The picture shows the bedrock geometry of the Gorshkov crater glacier at Ushkovsky volcano, Kamchatka, Russia, with the computed result for the age of the ice.
© Visualization Thomas Zwinger, CSC, the Finnish IT center for science

Back cover: Using computational methods researchers can screen potential leads for drugs by studying their 3D conformations and binding properties. Silver colored drug raloxifene binds to a nuclear receptor, reducing the risk of endometrial cancer and osteoporosis.
© Visualization Jyrki Hokkanen and Tommi Nyrönen, CSC, the Finnish IT center for science