

Evaluation of the Nordic Data Grid Facility (NDGF)

March 23, 2007

Summary

We will here present a summary of the panel's assessments and recommendations.

The evaluation committee is in general satisfied and impressed by NDGF and what has been achieved by the organization within a fairly short time of operation.

The evaluation committee recommends that the management team of NDGF should be given more freedom and decisive power in order to increase the speed of the project. The committee recommends that the NDGF- Steering Board takes a political role and leaves more of the strategic decisions to the NDGF management team. In the current phase, a certain amount of pragmatism is necessary in order to get NDGF as fast as possible into operational mode.

NDGF may today be interpreted as a "CERN activity". Similarly, the NDGF middleware seems to be very much directed by the requirements derived from CERN. This is a concern shared by both NDGF and the evaluation committee. The evaluation committee recommends that NDGF, within 2010, should extend their activity to at least two new scientific areas. The development of the organization and the magnitude of effort towards new targeted areas should match the CERN activity today. The new application areas should have somewhat different and complementary requirements, compared to the batch oriented processing model adopted for CERN. Ideally, this should drive the NDGF software towards a more general purpose grid middleware platform. If successful, this would increase the value of the NDGF platform, by being a significant research vehicle for other scientific domains as well. Hence, by targeting other scientific areas NDGF may become a driving force with respect to the adoption of grid technology within Europe.

NDGF is currently located in Copenhagen, Denmark and co-located with NORDUnet. The evaluation committee finds the choice of location and host institution overall flexible and good. Consequently, we recommend that no changes are made in this respect.

The NDGF activity has faced dead periods due to coordination problems across many different institutions, participants and nations. To reduce the risk of similar situations in the future, the evaluation committee would like to stress the importance of having well coordinated financial sources in the future — in particular across national funding agencies. A stable financial situation is of crucial importance in order to secure long term success of NDGF. We recommend that the funding agencies in the Nordic countries as soon as possible guarantee a stable funding for NDGF until 2010. On behalf of the Nordic countries and as a part of the European grid community, NDGF want to achieve something significant. This must be supported.

1 Introduction

This report presents the results of an evaluation of the NDGF (Nordic Data Grid Facility). The evaluation committee met with the NDGF management team during 8th - 9th February 2007. The meetings were hosted by NDGF in their offices in Copenhagen, Denmark.

The evaluation report starts with a summary of the assessments and recommendations made by the panel. This section provides a short introduction and describes the structure of the report.

The rest of the evaluation report is structured as follows. First an overview of the evaluation procedure is given. The following three sections then address the three areas of investigation, as specified in the mandate for the evaluation committee — organization and management, technical progress, and the choice of host institution. A section related to recommendations to the NDGF Steering Board follows. Finally, the meeting agenda as well as references are provided at the end of the document.

2 The evaluation procedure

The evaluation was initiated by NOS-N (Nordisk samarbeidsnemnd for naturvitenskap) in a letter from NOS-N to NDGF, dated 19. December 2006 [15]. The purpose of the evaluation is to assist the NDGF Steering Board with an evaluation of the activities in NDGF. The evaluation should finish during first half of 2007. Due to time limitations, a comprehensive midterm evaluation was considered unfeasible. Consequently, NOS-N suggested NDGF Steering Board to initiate a self-evaluation. Note that an in depth evaluation is planned in 2009.

2.1 The evaluation mandate

The mandate of the self-evaluation work was also given in the letter from NOS-N to the NDGF Steering Board. The evaluation committee was specifically asked to further investigate:

- Organization and management model, including cost-efficiency
- Technical progress, with particular attention to the recommendations made by the Evaluation Panel in their report of 01.06.2005 [1]
- Choice of host institution

2.2 The evaluation schedule

The plan for the evaluation work was determined by NDGF. The schedule and the deadlines for the work taking place after the meeting with the evaluation committee and NDGF management is as follows:

- A draft of the evaluation report to NDGF Steering Board, 5. March 2007
- Comments from NDGF Steering Board to the evaluation committee, 15. March 2007
- Delivery of evaluation report to NDGF Steering Board, 26. March 2007
- Submission of report to NOS-N, 28. March 2007.

2.3 Comments regarding the evaluation mandate

The relevance of the previous evaluation report is not very strong due to changes of the NDGF set up during the past year. Consequently, the evaluation committee finds it hard to comment on the technical progress based on the previous report. In this respect, the committee has not adhered strictly to the mandate. However, it should be noted that this issue was raised several times by the committee during the meeting, and also acknowledged by the steering board during the meeting at NDGF.

2.4 NDGF evaluation panel members

The evaluation committee consists of three persons. The following two persons were suggested by the NOS-N as external experts:

Morten Dæhlen

Professor, Head of department

Department for Informatics, University of Oslo, Norway

mortend@ifi.uio.no

Karl-Fredrik Berggren

Professor of Theoretical Physics at IFM

Linköpings Universitet, Sweden

kfber@ifm.liu.se

In addition to the external experts, the following person was invited by NDGF to act as scientific secretary for the evaluation committee:

Viktor S. Wold Eide

Post Doc. Networks and Distributed Systems department

Simula Research Laboratory, Norway

viktore@simula.no

None of the committee members have been involved in any of the NDGF work, neither in the past nor present, which should contribute to the independence of the evaluation work.

3 Organization and management model

This section discusses the organization and management model of NDGF, including recruitment and cost efficiency.

3.1 Recruitment

Putting together a management team consisting of persons with cooperative skills has been of paramount importance. When the recruitment process started in second quarter of 2006, the goal was to employ five leaders which work well together. In such a small team, team-workers are more important than self promoters.

NDGF has used professional and highly qualified assistance for the recruitment task. This task was handled by a professional recruitment company, which has specialized in recruitment of persons to management and specialist positions.

Currently, the management team of NDGF consists of the following positions and employed persons:

- Director — Rene Buch
- Development Manager — Lars Fischer
- Technical coordinator — Michael Grønager
- Software coordinator — Josva Kleist
- CERN coordinator — Oxana Smirnova

Our impression is that the NDGF team consists of highly qualified persons who seem to work well together. Despite the relatively short time frame, the NDGF director has put in a lot of work and also been willing to spend a fair amount of money in the recruitment process. The evaluation committee is impressed by both the recruitment efforts and the outcome. Altogether, the evaluation committee finds that NDGF has been successful regarding recruitment.

3.2 Slim organizational structure

The number of people in the NDGF management team is rather low. In spite of the relatively thin administration, the committee finds it adequately for the current tasks. However, if and when NDGF enters other scientific areas, as discussed in the following, additional resources are necessary to strengthen the management.

3.3 Distributed organization

NDGF as a whole is realized as a distributed organization. Although the NDGF top management is co-located, the different node leaders are scattered throughout the data centers in the Nordic countries. Similarly, the software developers are not co-located either.

The distributed organization of NDGF creates both some challenges and benefits. To overcome the difference in location, NDGF employees utilize messaging applications and also engage in weekly conference calls. Additionally, face-to-face meetings are held

once a month. However, the decentralized structure also allows, for example, the node leaders to get closer to the real users of the NDGF products and services.

Currently, NDGF contributes to grid activities at 10 sites. According to the technical coordinator this number of sites is manageable.

Regarding software development, only 100% positions are used. Part time positions are avoided in order to avoid conflicts of interest.

3.4 Cost efficiency

Having a centralized NDGF is very difficult, since a centralized unit will require a localization discussion among the Nordic countries. Consequently, NDGF assumes a decentralized approach that more easily adapts to the different national systems of the Nordic countries.

With respect to human organization, the non-centralized model gives additional cost, in particular for physical meetings and associated travel expenses.

With respect to software organization, the cost efficiency of the planned grid facility is closely related to the efficiency of the grid software being developed and its ability to keep the computers busy working at all times. By pooling resources, the efficiency can potentially be improved, but the efficiency experienced is inherently coupled to the efficiency of the grid software.

4 Technical progress

This section discusses the technical progress, with particular attention to the recommendations made in the evaluation report of 01.06.2005 [1].

4.1 NDGF and CERN

The evaluation committee would like to honor the cooperation with CERN, which has created interest in the CERN community for the NDGF approach. More specifically, NDGF has taken a more distributed approach compared to other efforts. The current NDGF approach has shown, for example, the best responsiveness regarding storage. As a result, it seems that CERN expects fairly significant contributions from NDGF.

4.2 Alignment of expectations and intentions

At one hand these expectations are positive. However, there may be a potential mismatch between the expectations from CERN and what NDGF intends to deliver. In this respect the evaluation committee would like to stress the importance of an alignment of the expectations from CERN with what NDGF both can and will deliver.

4.3 Processing model

The NDGF middleware requirements derived from the CERN activities seem to be very much batch oriented. The batch oriented processing approach directed by CERN is

rather conservative. Additionally, according to NDGF technical people the processing model for the CERN activities ATLAS, ALICE, and CMS are not really that different¹.

In the evaluation committee's point of view, the batch oriented processing model is unsuited for many other important application areas. As an example, some applications may require workflow oriented processing. Such processing is more complicated, since several resources have to be scheduled in a coordinated fashion.

4.4 Technical progress

CERN will start generating data at the end of 2007. With respect to technical progress, it seems like NDGF is fairly well positioned to accommodate the needs of CERN and in a timely fashion. Most notably, the monitoring and accounting system is still not ready, but integration work is ongoing.

The evaluation committee is impressed with the pace at which things have progressed since the contract with NDGF-SB, medio May 2006.

4.5 Uncertainties related to lack of resources

It should be noted that there are some uncertainties related to the hardware which will eventually be placed at the different sites in the Nordic countries. In some countries the necessary budgets are in place, but that is not the case for all countries. In particular the money required for the investments in hardware have not even been approved in Finland.

In general, it has proven hard to get tier-1 resources committed from the different institutions. NDGF can not force, neither people nor institutions to participate. Similarly, NDGF does not decide what institutions commit. The node coordinator in each country is responsible for getting resources such as CPUs, disks, and tapes online. Resources not online are considered non existent.

For the upcoming experiments, the equipment must run 24/7. However, some slack may be tolerated due to some planned redundancy. Since NDGF relies on physical equipment located in data centers scattered around the Nordic countries, the success of the NDGF to a large extent depends on local system administrators being able to keep the equipment up and running. In other words, getting resources for having system administrators more or less continuously operating the equipment is crucial.

In some sense, the lack of control over resources may be seen as a weakness of the NDGF model, where NDGF does not own or directly control the hardware. In order to handle the uncertainty, stronger commitments from the involved institutions and long term contracts are required.

4.6 Grid software considerations

Regarding the software development, NDGF strives towards modularity, in order to improve the potential for reusing the software in other settings. The developed software can, for example, be used in settings different from what is imposed by the upcoming CERN experiments.

¹There are some differences, for example, related to how software is updated.

5 Choice of host institution

The committee finds the choice of host institution an overall flexible and good choice, which has allowed for a quick startup of the NDGF.

From a Nordic point of view, NDGF is centrally placed in Copenhagen, Denmark. The offices also have good location with close proximity to Kastrup airport and the railway system. Being close to the airport is important due to relatively frequent meetings with people from the Nordic countries and the rest of Europe, for example CERN.

NDGF is co-located with NORDUnet, which is advantageous since some persons are employed by both NDGF and NORDUnet. Additionally, this arrangement allows NDGF to leverage on the NORDUnet infrastructure. Hence, the need for costly investments is reduced and the project can more smoothly be reorganized in the future if needed.

The evaluation committee sees no reason to make any changes regarding host institution. The committee therefore recommends that NDGF sticks to both the current choice of host institution and location.

6 Recommendations to NDGF-SB

The evaluation committee recommends that the management team of NDGF should be given more freedom and more decisive power in order to increase the speed of the project. Hence, the evaluation committee recommends that the NDGF-Steering Board takes a political role and leaves room for more proactive work to the NDGF management team, for example, to identify new fields of action. In the current phase, a certain amount of pragmatism is necessary in order to get NDGF as fast as possible into operational mode.

Moreover, NDGF- Steering Board should as soon as possible establish a roadmap for the planned activities. An important part of the roadmap is when and how NDGF should enter new scientific areas, see the following discussion.

6.1 Entering new areas

The NDGF management team sees a danger of being too much oriented towards CERN activities. The evaluation committee very much shares this concern. Today, NDGF can almost be interpreted as a “CERN activity”. By targeting other scientific areas, NDGF may be a driving force with respect to the adoption of grid technology within Europe.

The evaluation committee recommends that within 2010 NDGF should extend their activity to at least two new scientific areas. With respect to organization and effort, these targeted areas should match the CERN effort today. In order to make this goal realistically achievable, the evaluation committee recommends that the activities within these upcoming areas have to be initiated in 2007 and 2008, respectively. It should be noted that in the NDGF draft budget 2006 – 2007, [24], Euro 340.000.- has been set aside for specific projects outside the current core user community.

According to how the CERN activity is organized and handled today, with a designated coordinator, it seems reasonable to assign a separate coordinator for each

new addressed area. From what was said at the evaluation committee meeting, it appears that NDGF management would like to hire a new coordinator within 12 months.

6.2 Criteria for selecting new areas

The evaluation committee retains from specifying the exact criteria for selecting the new areas. However, the committee believes that emphasis should be given to the following.

The new scientific areas targeted should be chosen based on researchers interests and scientific quality. The areas should also have strong support in at least two of the countries funding the NDGF. Additionally, the likely relevant applications within the new areas should challenge the NDGF middleware, not only as a batch processing platform.

The NDGF middleware currently developed seems to be very much directed by the requirements derived from the CERN activities. By choosing several application areas with somewhat different and complementary requirements, the NDGF software should ideally be driven towards a more general purpose grid middleware platform. Examples include requirements for workflow, streaming, and real-time processing. If successful, this would increase the value of the platform, by being a significant research vehicle for other scientific domains as well.

The research areas to consider include, but are not limited to: material science, medicine, bio-informatics, biomedical sciences, astronomy, atmosphere research, meteorology, climate research, northern lights (aurora borealis) research, and environmental surveillance.

Some of these areas have requirements which do not fit the batch oriented approach for processing, as targeted by the NDGF middleware currently.

The evaluation committee agrees with the NDGF Steering Board that the outcome of the current Nordic Computational Grand Challenge Survey may provide valuable information regarding which areas to enter. The evaluation committee understands that the results from the survey will be available within short time.

6.3 Marketing towards potential research communities

With respect to visibility NDGF plans a new content management system. Keeping the information updated is a continuous task which requires resources. NDGF would like the updating process to be distributed to the different stakeholders.

In order to get attention from potential research communities, visibility at conferences and invitations to participate may not be sufficient. A more active approach seems necessary, for example, hands on participation in new projects. Experiences from other fields of science show that such an active approach is necessary to attract new users into new areas like grid computing.

The NDGF-Steering Board should encourage pilot projects and incubator projects within the chosen areas. In the NDGF draft budget 2006 – 2007, [24], Euro 175.000.- has been set aside for special development / incubator activity. In this respect, it is important that NDGF promote the grid by interacting with highly motivated key persons

within the potential new areas for expansion. These persons hopefully will play an important role in further promoting the grid, both internally and externally.

The evaluation committee acknowledges that it would be highly beneficial, if not necessary, to have some available computing resources before encouraging research communities to start taking advantage of the NDGF.

6.4 User satisfaction and quality

It is always important to keep in mind why the Nordic countries are doing this through NDGF. It is not only to satisfy the needs of CERN and researchers using data from the CERN experiments. Equally important are the need for knowledge development and experience gathering so that the systems and competence being developed can be used for other purposes. The users gradually become more and more important.

The close relation to CERN provides suitable follow up routines towards the central organization in Switzerland. However, as soon as the real users start utilizing the system it is important to get good user feedback and some kind of approval of functionality and system quality. If not already done, the evaluation committee recommends that NDGF develops simple routines that monitor user satisfaction. This is of course also an important issue when entering new areas. Scientific achievements made possible by NDGF (high profile publications, PhD theses, etc.) should also be reported on regular basis.

6.5 Stable long term funding

The NDGF activity has faced dead periods due to coordination problems across many different institutions, participants, and nations. The committee would like to take the opportunity to highlight the risk that such situations may again occur in the future, unless care is taken.

The evaluation committee would like to stress the importance of having well coordinated financial sources in the future — in particular across national funding agencies.

The success of an undertaking such as the NDGF depends on the ability to plan some time ahead. Hence, a stable financial situation is required and efforts should be made to secure long term funding.

Agenda for the meetings 8-9 February

Thursday February 8th 2007

1730-1800 A discussion on the practical setup of the evaluation. Present: SB delegation, Director.

1800-1900 Presentation of NDGF by the director. Present: SB delegation, Director

1900-2000 Presentation of SB issues

2000 - Dinner arranged by NDGF at Blue Elephant SAS Radisson

Friday February 9th 2006

0900-1000 Interview with Director - Rene Buch

1015-1115 Interview with Technical Coordinator - Michael Grønager

1115-1145 Interview with CERN SUB Committee Chair - Ian Bearden

1150-1230 Interview with Development Manager - Lars Fischer

1230-1315 Lunch — all

1315-1350 Interview with Software Coordinator - Josva Kleist

1350-1415 Interview with CERN Coordinator - Oxana Smirnova

1415-1445 Rounding up — evaluation committee

References and background information

- [1] Evaluation of Nordic DataGrid Facility (NDGF). June 2005.
- [2] A Proposal for a Nordic DataGrid Facility. The Result of the NDGF Pilot project. August 2005.
- [3] NDGF SB Meeting. March 2006.
- [4] Memorandum of Understanding. April 2006.
- [5] Agreement on Hosting of the Nordic Data Grid Facility. April 2006.
- [6] NDGF High-Energy Physics Committee Constituting Meeting. May 2006.
- [7] NDGF SB Meeting. September 2006.
- [8] Oxana Smirnova. CERN activities in NDGF. September 2006.
- [9] Michael Grønager. Technical Plans and Milestones. Progress Report. September 2006.
- [10] Lars Fischer. NDGF Development Report. September 2006.
- [11] Josva Kleist. NDGF Software Development Status & Plans.
- [12] NDGF Status and Planning Report, Q4 2006. October 2006.
- [13] NDGF Tier-1 Network Proposal. October 2006.
- [14] Decision Minutes — NDGF Steering Board Meeting. November 2006.
- [15] NOS-N. Regarding the future Nordic Data Grid Facility. December 2006.
- [16] Michael Grønager. WLCG-NDGF Management Meeting. January 2007.
- [17] Michael Grønager. NDGF Tier 1 and assoc. Tier 2 report. January 2007.
- [18] Lars Fischer, Michael Grønager, Josva Kleist, and Oxana Smirnova. The Nordic DataGrid Facility Design and infrastructure for the distributed NDGF Tier-1. January 2007.
- [19] Michael Grønager. NDGF-EGEE Interoperation Status. January 2007.
- [20] NDGF Status and Planning Report, Q1 2007. January 2007.
- [21] Oxana Smirnova. NDGF CERN Project Planning. January 2007.
- [22] Call for proposals: Nordic Data Grid Facility Community Grid Projects. January 2007.
- [23] NDGF Self-Assessment. February 2007.
- [24] NDGF - Draft Budget 2006 - 2007 .