

Executive Summary

Many scientific disciplines are presently undergoing technological revolutions that lead to a common challenge: managing a distributed data explosion. Detectors, medical imaging instruments, micro-arrays, and multi-sensor instruments are producing amounts of data that are rapidly exceeding the capacities of their current local data storage and computing environments. In many cases these ‘exploding data’ are distributed from the very start, being produced by different research groups or distributed sensor networks. Consider examples like genome and protein analysis data produced by many research labs in the world, biobanks containing patient data from a variety of hospitals, biodiversity data collected at the banks of the river Waal, historical archives and text corpora in many different places. Combining these datasets allows for completely new forms of research. Moreover, experiments generating petabytes of data per year, such as LOFAR in radio-astronomy and CERN in particle physics, need more data processing power than ever can be located in a single facility, with data utilized by researchers all over the world.

From an ICT perspective, these data have similar properties: all require reliable storage, comprehensive archiving, secure coupling and sharing. We propose to build and roll out a nationwide grid-based e-Science infrastructure, BIG GRID, that strengthens the international position of the Netherlands in many scientific areas. BIG GRID encompasses data storage facilities and data processing services, enabled by grid services, for a requested budget of 30 M€ over a four-year period.

The science case for this proposal is the integral of many different science cases, reflecting the broad scientific community base. The realization of BIG GRID is crucial to the success and continuity of many Dutch research communities, covering important areas such as life sciences, astronomy, particle physics, meteorology, and climate research, water management, to name just a few. However, the very nature of the new infrastructure, a multidimensional collaboration enabler and accelerator, allows for direct participation of also social sciences, humanities, and even addresses communities in administrative domains, like digital academic repositories.

One basic ingredient for the proposed infrastructure is the network. The Netherlands are already in an excellent position, due to the world-class network services provided by SURFnet, the upgrade of which has been secured from GigaPort-NG project. BIG GRID provides opportunities for enhanced international visibility. Dutch participation in international generic grid developments is already prominent (in flagship projects like EGEE and DEISA) and are on a national scale very well covered by the VL-e project. Coordinated by the Netherlands Genomics Initiative, NBIC is the key player for enabling informatics methodology for life sciences.

While the Netherlands is a leading player in the development of the grid, and has considerable expertise in bio-informatics, distributed sensors networks, and particle physics, the large-scale infrastructure to fully exploit this leading position is missing. The purpose of this proposal is to realize a science-wide national grid infrastructure. This puts the Netherlands at the forefront of grid developments, enabling many national ambitions. It enhances the excellent position of Dutch academic hospitals in patient data collections using the grid for biobanking. It enables major advances in drug discovery through combining data and through availability of massive compute resources for modelling. It allows industrial research labs, such as Philips, to both contribute to and profit from the available resources for engineering sciences. It positions LOFAR as the European centre for serving a variety of scientific communities using LOFAR data and the Netherlands as one of the Tier-1 sites for CERN’s LHC experiments.

This proposal is a collaborative effort of NCF, NBIC and NIKHEF.