

## The Dutch National e-Infrastructure

---

### Jeff Templon\*

NIKHEF, Science Park 105, 1098 XG Amsterdam, NL

E-mail: [templon@nikhef.nl](mailto:templon@nikhef.nl)

### Jan Bot

SURFsara

E-mail: [jan.bot@surfsara.nl](mailto:jan.bot@surfsara.nl)

The Dutch National e-Infrastructure for Research (DNI) was reorganised in 2013 in order to increase the integration between the infrastructure platforms offered. The SURF Cooperative was charged with leadership of this integrated ecosystem. Since then, there has been much progress and activity towards increased utility, ubiquity, and ease-of-use of the DNI. Significant effort has also been expended in the area of user support and outreach, national alignment between several data repositories, and innovation of the infrastructure. eScience research and engineering, best practices and policies regarding data stewardship and software sustainability have been addressed by the Netherlands eScience Center which is closely associated with the DNI. This paper will cover these topics, as well as outline the current challenges faced in improving the reach and quality of the infrastructure, as well the DNI's position in the international e-infrastructure ecosystem.

**Disclaimer:** Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the SURF Cooperative.

*International Symposium on Grids and Clouds (ISGC) 2016*

*13-18 March 2016*

*Academia Sinica, Taipei, Taiwan*

---

\*Speaker.

## 1. Introduction

Scientists are increasingly relying on IT resources to carry out their research. Where in the past e-infrastructures were mostly useful to a small subset of the research community, they have now become an integral part of the scientific workflow, spanning from environmental modeling to literature analysis.

Furthermore, interest in research data has increased from both the scientific community as well as the agencies sponsoring the research, and from the business world (*e.g.*, interest from the pharmacological world in bioinformatics results). Funding agencies envision research data being hosted in a unified environment accessible to the entire scientific community.

Finally, these data, when so available, need to be accessed from and processed on, an ever-increasing diversity of e-infrastructure platforms. This either becomes an administrative burden for researchers, or the data and platforms must be integrated as much as possible in some unified environment, along the lines of what has been called the “GAFA model” (Google, Amazon, Facebook, Apple) [1]. In this model, a single credential is used to access a set of cross-integrated data and computing services. Many researchers already use GAFA, without the blessing of their funding agencies; unless our national/public e-infrastructures offer the same ease-of-use and level of integration, they will be experienced as “failures” by our scientist-users.

All these considerations require scientific institutes, funders and e-infrastructure providers to rethink the e-infrastructures and the way they are being offered.

The Netherlands has been pioneering e-infrastructures for research for the last 15 years. The 14 universities, 8 university medical centers and several research institutes have been collaborating on e-infrastructure provisioning for many years. The relatively small size of the country and well-organized scientific community has made it possible to create a Dutch National e-Infrastructure for Research (the DNI) accessible to all researchers in the Netherlands. Our vision is to integrate the components of this infrastructure along the GAFA model, and extend past this model to incorporate and integrate research platforms not available from GAFA. This paper describes our infrastructure, the sustainability model behind it, and the progress towards achieving our vision.

## 2. Scope & Target users

The mission of the Dutch National e-Infrastructure is to enable scientific research in the Netherlands by provisioning ICT infrastructures and services. Provision is meant here in a broad sense and includes targeted developments of new infrastructures and services, as well as support and training related to effective use of the DNI. Any person who is eligible to request support from the The Netherlands Organisation for Scientific Research (NWO) is eligible to be supported by the DNI; there are also cases in which post-secondary educational institutions can make use of the DNI.

Astronomy, life sciences, quantum chemistry, particle physics, theoretical physics, various flavors of “omics”, climate, water management/morphodynamics, and astrophysics are some of the research areas that register prominently in our usage statistics. During a recent three-week period, the following incoming resource requests were received:

- 3 from morphodynamics

- 2 from structural biology
- 1 each from complexity science, social media analysis, deep learning, and theoretical high-energy physics

The degree of technical ICT knowledge skills, amount of support required, and research group sizes vary widely across the research communities and resource requests. This has consequences for the DNI support organization described later in the paper.

The type of infrastructure being requested also varies widely per community and per request. Our HPC Cloud service is a popular choice, accounting for 46 of the 59 requests arriving during the first 10 weeks of 2016. Grid computing follows with only 8 requests; although less popular, grid requests far exceed cloud in terms of total core-hours requested. Big data, visualization services, and collaborative storage had 3, 3, and 2 requests respectively. It was not possible to make a comparable breakdown for the National Supercomputer or the National Cluster, as requests for allocation on those still proceed via a request to NWO for the moment.

Looking at the entire spectrum of resource requests, we can summarize the types of infrastructures requested and how these correlate with the research-ICT problem to be solved:

**storage** various types are requested, from small-scale storage with flexible features for cross-institute collaboration, all the way to platforms for large-scale data-driven science

**cloud** predominately viewed as an “easy” way to scale workflows out to medium scales

**big data** large-scale analyses of unstructured, heterogeneous data

**grid** popular amongst savvy users and/or large collaborations, especially when a large data component is present; capable of reaching the largest scales

**cluster** high performance, little demand for collaborative features; another “easy” scale-out platform

**supercomputing** niche infrastructure for the toughest computations

**dedicated network connectivity** specialized infrastructure for transferring large amounts or sensitive data.

The takeaway message here is that, for a truly national e-infrastructure, one size definitely does not fit all. Serving the entire scientific community will require a rather diverse palette of infrastructure components.

### 3. Outreach & Support

The centralized nature of the DNI means that the distance to the end-user (scientists) is relatively large compared to locally-offered infrastructures. The consequence of this is that many scientists and research supporters don’t know about the existence of DNI resources they could be using. To close this gap, the DNI undertakes a number of outreach activities.

### 3.1 End-user support

Support of the DNI is arranged per infrastructure, with each infrastructure having its own group of system administrators and consultants. Most consultants can provide advice and hands-on support on multiple infrastructures which enables us to direct researchers to the infrastructure best suited for their task. The support ranges from gathering system requirements and advising on which infrastructure to use to full project participation where one or more consultants become part of a project team and help port or optimize an application for a particular infrastructure.

The support model differs per infrastructure. For example users of the DNI HPC Cloud are expected to be self-reliant, creating and maintaining (including security updates) their own virtual machines; users of the Cartesius supercomputer however, can request software to be compiled and installed by DNI consultants. The support model follows the IaaS, PaaS, SaaS hierarchy; as most services fall between PaaS and SaaS, most supporting software will be installed and tested by DNI staff on the users' behalf, though the users are responsible for efficiently running their own application.

### 3.2 Support4research

Interviews held with people selected from a cross-section of the research community showed that a majority of scientists prefer local support, and are more likely to engage with research supporters when these are located at their own institute. Hence the Support4research project was initiated, to involve the ICT staff of research institutes in providing support for the national e-infrastructure. On-location workshops and masterclasses are regularly held to familiarize support staff at the institutes with the DNI resources available to the researchers they support, to a sufficient extent that they can advise their users regarding the use of DNI and also provide first-line support for our infrastructures. This improves the scalability of user support as the institutes themselves can take care of handling user requests. To further promote this, Support4Research also facilitates integration of the service portfolio of the DNI into the institutes' websites, and actively solicits input on how to improve both the DNI request and user-support procedures.

### 3.3 Training activities

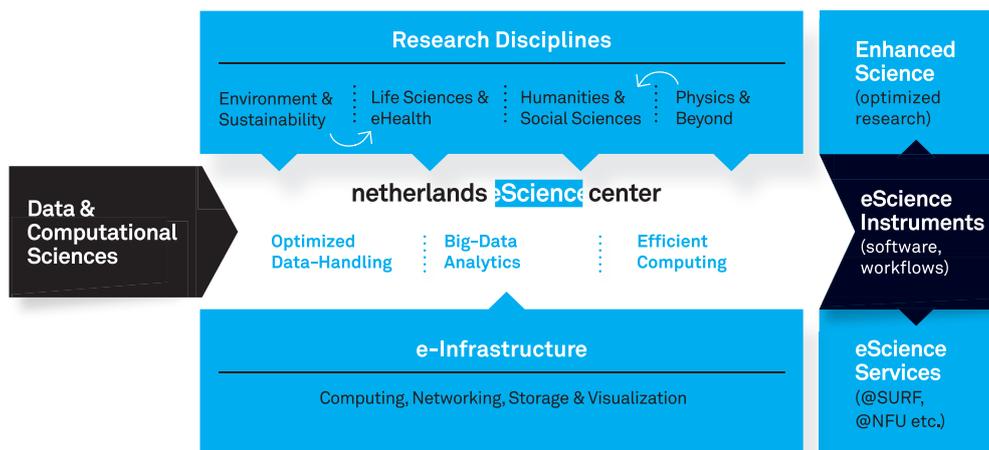
To engage directly with scientists, DNI also provides a wide range of training opportunities directed at scientists. These training sessions range from short 'crash courses' of a few hours on using a particular infrastructure to full courses which take an entire week and cover the use of multiple infrastructures. The latter group of courses are often organized in collaboration with a PhD school which motivates students to participate.

Over the years many different training activities have been tried, with varying degrees of success. Hands-on training sessions for specific infrastructures have worked well, depending on which infrastructure was being taught. A MOOC on grid computing was developed and run; while we were pleased with the number of participants, it was not sufficient to justify the workload associated with running the course. The focus is now on local topical workshops, aimed at groups of scientists with similar computational challenges. This allows the material to be tailored to demands of a scientific discipline. Where possible we reuse materials created by others; the training materials provided by the Data Carpentry [2] organization is a specific example.

To conclude this section, we remark that our national mandate makes it possible for us to organize a support ecosystem of the scale and scope required, once again profiting from economies of scale. That being said, taken across all of science, much research-ICT support can be satisfied at the local level (see also section 7). This means that local research-ICT staff at universities and institutes play a crucial role in this support ecosystem, as they are in the best position to realize when their researchers need to scale out to the national level.

#### 4. eScience Research and Developments

Modern data-driven and computing-intensive research approaches require researchers to rapidly develop ICT skills that may currently be foreign to them. ICT skills will increasingly underpin research in the way that mathematics and statistics currently do. For many scientists, the prospect of developing the skills needed to engage increasingly diverse and complex e-infrastructure is daunting. eScience can provide discipline focused scientists the tools, personnel and wider support needed to lower this barrier and ensure developments made in computing and data-science can be readily applied in more fields.



**Figure 1:** Graphical representation of the Netherlands eScience Center strategy. New science is enabled by bridging gaps between research disciplines and the DNI by making advances in optimized data handling, efficient computing, and big-data analytics. These advances are captured when possible as new eScience instruments to be reused by other communities.

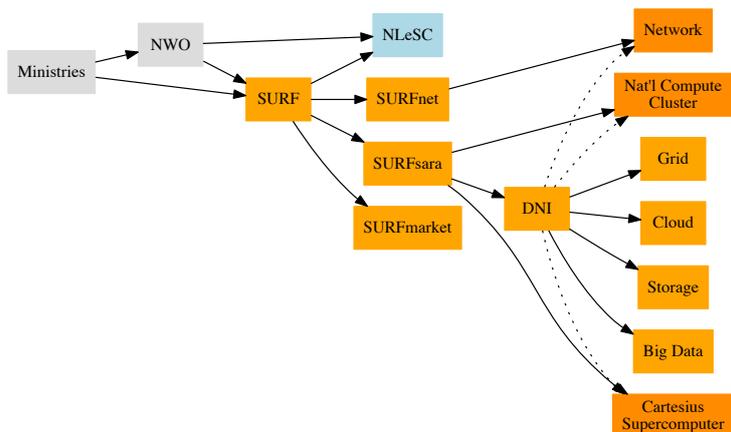
The commission charged with designing the integrated Dutch National e-Infrastructure foresaw [3] a separate organization concerned with innovation at the boundary between the infrastructure and the science. Quoting from the commission’s report:

... found an eScience Center at which the existing universities and institutions cooperate on research aimed at innovation of ICT infrastructures and scientific applica-

tions. This will make it possible to continue the valuable multi-disciplinary collaboration between universities, institutions, and businesses. The Netherlands will continue to be an attractive location for knowledge-intensive industry and top researchers. The new centre will also raise the international profile of the Netherlands as a country of open, international science and scholarship, with highly developed connectivity.

The Netherlands eScience Center (NLeSC) was launched in 2011 as a response to this charge. The Center is a unique collaboration between NWO (p. 1) and SURF (p. 5), occupies an important place in the Dutch eScience landscape. NLeSC projects are driven by domain science questions, fueling innovation in research software and informing planning for the eInfrastructure. The “eScience Integrators” advisory board and the ePLAN coordination activity, both run by the NLeSC, function as additional channels for feedback from the research community back into the eInfrastructure ecosystem. The NLeSC mission is to enable digitally enhanced research through efficient utilization of data, software and e-infrastructure. This mission is pursued via a four-pronged approach: enabling scientific breakthroughs, collaborating in problem-driven projects, developing versatile cross-disciplinary eScience tools, and coordinating eScience activities.

### 5. Organization



**Figure 2:** The chain of responsibility (and funding) from the ministry level down to hardware on the floor. Access to the dark-orange-colored infrastructures still follow historical routes, hence the dotted lines from to DNI.

contributions coming from other subsidiaries of SURF as well as partner institute Nikhef. Figure 2 shows how both the funding and responsibility for scientific computing is organized in the Netherlands.

Aside from Nikhef, the Rijksuniversiteit Groningen is a second partner; these two institutes operate substantial portions of the DNI Grid infrastructure, and are both involved in prototyping the next generation DNI Cloud infrastructure. They both contribute to user support for the infrastructures they operate. Nikhef and RUG contribute to the executive decisions concerning the

POS (ISGC 2016) 020

infrastructure as a whole via their representation on the DNI “Executive Team” which meets every six weeks to discuss high-level issues concerning the DNI.

The association within the DNI of several major providers of ICT resources makes the DNI the natural partner, in many cases, to participate in various international initiatives. In some cases a direct participation by one of the partner institutes is warranted, however issues regarding the interaction between such projects and the DNI are generally discussed in the Executive Team meetings.

## 6. Resources

Over the years the number of e-infrastructures has grown, starting with a super computer and gradually adding cluster, cloud and big data resources. This organic growth was due to specific e-infrastructure needs as described by various research communities and often funded through different sources. The DNI is working on the harmonization of access procedures to provide the infrastructure best suited for a project, regardless of scientific discipline or institute a researcher might be associated with. In this section we describe the available e-infrastructures and the way scientists can gain access to them.

### 6.1 National e-Infrastructures

**NL Grid and T1 site** Nikhef and SURFsara together form the Dutch T1 site for the WLCG. Together they provide 12 000 cores and 7 PB of storage capacity to the LHC community serving the ALICE, ATLAS and LHCb experiments. The same resources together with grid cluster located at RUG CIT are made available to other researchers in the Netherlands, regardless of which area of research they are in. The main usage of grid resources comes from large user communities which either provide their own services which hide the underlying scheduling system (*e.g.* in the form of portals) or by advanced users who can deal with the complexity of grid systems themselves.

**Life Science Grid** To better serve the life science community and create a sense of ownership, small clusters were placed at university medical centers and other institutes with a strong focus on life science research. These clusters are used as a stepping stone to the grid facilities by providing an environment where scientists can experiment with a remote system, optimizing their workflow for a batch scheduling environment and eventually making the leap to grid computing. These life science clusters help ensure local support and provide a good intermediary between locally maintained compute servers and more powerful shared systems. Scientists working on the life science grid can access all twelve small clusters, as well as the NL Grid resources described above.

**HPC Cloud** The IaaS HPC Cloud first came online in 2011 [4], being the first multi-tenant IaaS system specifically designed for scientists. The current HPC Cloud system is an OpenNebula-based platform, focusing on interactive usage. The ‘self-service’ nature of this infrastructure attracts users that need a flexible environment and tend to use individual virtual machines as powerful workstations. The infrastructure has developed to keep up with this usage pattern: it provides machines with a relatively large amount of cores, memory and disk space.

**Compute cluster LISA** LISA is a general purpose compute cluster which targets users with moderate compute needs. It is used mainly for batch processing applications, running many serial tasks in parallel. It is one of the workhorses of the DNI, providing resources to a diverse set of users.

The spectrum of users across the HPC Cloud and LISA are quite similar; in most cases, the choice between the two is driven by which technology the user happens to prefer.

**Big Data infrastructure** The Big Data facilities of SURFsara enable scientists to do large-scale data processing, using the same tools as technology companies such as Yahoo!, LinkedIn and Facebook. The core component of the service is an Hadoop cluster of approximately two hundred nodes. These nodes are part of both the distributed storage system and the compute cluster, with a scheduler that optimises for processing with data-locality. This lowers network bandwidth usage in the cluster and improves IO throughput. Software tools on the cluster include YARN, Spark, Pig, Flink, HBase and many other open-source frameworks.

The Hadoop cluster also hosts a number of ready-to-use datasets used by scientists from multiple institutes. Examples are Wikipedia, CommonCrawl, TREC and ClueWeb. While scientists still must seek permission for access from external parties, they need no longer arrange for staging of the datasets.

**Supercomputer Cartesius** The current Dutch super computer is Cartesius, a 41 000 core 1.6 PFLOP/s bullx system. This infrastructure is a capability system, providing large, fast storage, GPGPU nodes and infiniband interconnects. This system targets power users who can efficiently use large parts of this infrastructure for their applications. Cartesius is also used as a code- and algorithm-development stepping stone to the PRACE T0 systems. Codes and algorithms can be tested and optimized on Cartesius before applying for PRACE T0 time.

**SURFdrive** SURFdrive is an OwnCloud-based file-sync service positioned as an institute-approved alternative to DropBox. With currently around 15 000 active users it is one of the largest OwnCloud instances targeting academic users.

**Network** SURFnet offers Dutch education and research institutes a fast and reliable state-of-the-art Internet connection with fast links to international institutes and to the Amsterdam Internet Exchange. In addition, SURFnet offers lighthpaths to create fast secure, and reliable point-to-point connections between two locations.

As described in section 2, the diverse set of resources described above is what is requested, and in most cases required, by our scientific users. We can only offer such a broad palette of platforms, and achieve a reasonable degree of integration amongst them, because of the economies of scale afforded by our national scope and mandate.

## 6.2 Access routes to national e-infrastructure

A central DNI request portal is the starting point for obtaining access to those DNI platforms colored light-orange in fig. 2. Applicants fill out an online form indicating which platform they want to use and estimating their resource requirements (cores, terabytes, *etc*) on the platform(s)

specified, as well as the period over which they are needed. Requests are evaluated by platform-specific teams on technical merit, iterating with submitting party and/or other platform teams (in the tri-weekly meeting of representatives from all platforms) in order to choose the most suitable platform. For projects requiring long-term commitment of a substantial fraction of a platform, the request is also examined by the DNI executive team.

The concepts “National Supercomputer” and “National Computing Cluster” predate the Dutch National e-Infrastructure, and are not yet fully integrated into DNI. Requests for access to those require a separate route via NWO (see p. 2). Changing the access request for these infrastructures is a historical and political issue, which is more difficult to address than a technical one. Applications are judged through a peer review system, judgment of the technical feasibility is done by SURFsara. Small projects are treated as pilots and get access within weeks, large applications take a bit longer due to the review process.

## 7. Relationship to International Infrastructures and Initiatives

So far we have described the scope of the DNI, the researchers who make use of it, the organization of the support ecosystem enabling its use, and the infrastructure platforms themselves. In section 8 we describe the thinking around models for the economic sustainability of the DNI. All of these issues are predominantly national in character.

Furthermore, while large-scale international collaborations account for a large fraction of resource usage on the DNI, in terms of numbers of requests, most projects have no requirement for international collaboration or cross-border data access. We see from our connection to the universities and institutes, that there are many researchers who have even no need to expand outside of institute borders.

These considerations form the basis of our tiered approach to resource provisioning for research communities:

**Local if possible** It makes no sense to migrate scientific workflows that fit perfectly on a personal workstation with Excel installed, onto some pan-european integrated scientific research ICT system.

**National if needed** The first threshold in our vision is when resource usage gets large enough that specialized ICT support staff are required to operate the required systems. At this point, we (or for example NWO) can achieve economy of scale by providing allocations on the National Computer Cluster LISA instead of funding separate clusters – and separate cluster administrators – at multiple departments throughout the country. A second threshold is passed when research is done in collaboration; then it makes sense to migrate at least the research data to some federated (national) infrastructure in order to facilitate the collaboration.

**International when required** A final threshold is passed when there is significant international collaboration such as that seen in large-scale projects such as the LHC at CERN, the radio telescope LOFAR and the ELIXIR infrastructure for biological information.

For research passing this last threshold, the international dimension becomes very important. Also, research data contained in or described by scientific publications are relevant both nationally and internationally. Following the local-national-international tiered approach described above, the DNI participates in several organizations and projects that address international scientific collaborations, some of which address computing at large scales. Examples of participation of the DNI in international projects are:

**PRACE** The Netherlands are, with the participation of SURFsara, a member of PRACE, the Partnership for Advanced Computing in Europe. This EU project has created a pan-European supercomputing infrastructure providing access to computing and data management resources and services for large-scale scientific and engineering applications. The PRACE infrastructure consists of six multi-PFLOP/s systems. Dutch researchers can access the PRACE resources through a granting procedure where applications are judged through a peer review process.

**EGI** The European Grid Infrastructure is the federation of the National Grid Infrastructures in Europe. The Grid resources provided by the DNI are compatible with the EGI standards and are visible to all connected European scientists. Our policy on resource allocation to communities using EGI is that a Dutch scientist should make the request to the DNI; upon acceptance, access is granted to the collaborators named (specifically or by community affiliation) in the grant proposal. In special cases we allow certain communities with which the Netherlands has ties to consume idle resources when available. The DNI makes important contributions to various aspects of the Federated Authentication and Operational Security areas of EGI.

**WLCG** the DNI provides a Tier-1 computing center for the ALICE, ATLAS, and LHCb experiments at CERN and as such is an important partner in WLCG. In most cases, this contribution follows automatically by adhering to EGI standards as well as providing the agreed-upon resources. While the operation of the WLCG resources is funded by the DNI, the resources themselves (petabytes and computer cores) are paid for via an NWO grant to Nikhef. This system achieves economy of scale for both hardware (the supporting infrastructure like data centers, networks, head nodes and the like) and for operations manpower.

**EUDAT** EUDAT is a European research infrastructure that provides research data services, training and consultancy for researchers, research communities and research infrastructures & data centers. It provides a range of services to facilitate data storage and exchange, some in the form of web services and some as software stacks that can be used by the project's partners to host the services themselves. The DNI contributes expertise on infrastructures and outreach and integrates the EUDAT services in its own service portfolio.

**GÉANT** GÉANT is a European project that develops, delivers and promotes advanced networks and associated e-infrastructure services. It delivers network infrastructure, wireless services and authentication & authorization. SURFnet is the goto partner in the Netherlands for GÉANT which allows Dutch researchers to use international connectivity.

**EU-T0** the DNI is a partner in the EU-T0 Collaboration, a consortium of research agencies and research institutes that have traditionally been supporting large-scale data intensive science.

**HNSciCloud** the DNI (specifically SURFsara and Nikhef) participate in the pre-competitive procurement part of the HelixNebula Scientific Cloud project, aimed at exploring options for hybrid clouds based partially on infrastructures like the DNI, and partially on commercial cloud services.

**EOSC** The Netherlands follows the European Open Science Cloud (EOSC) developments very closely, partially within the context of the DNI. The Dutch Organization for Scientific Research NWO envisions a similar integrated scientific data-and-compute environment at a national level.

Our vision for the EOSC can serve as an example for the correct balance between national and international concerns. The European Commission's view for the EOSC is based on the FAIR access principle (Findable, Accessible, Interoperable, and Reusable) and while for this to work across national boundaries, international coordination is essential, the heart of such an EOSC will consist of the national Open Science Clouds, following the considerations given at the beginning of this section.

## 8. Sustainability and Economic Model

The integration under SURF in 2013 provided a welcome degree of sustainability for some infrastructure platforms (for example, the former "BiG Grid" Grid and HPC Cloud services) that were formerly dependent on project funds. Despite this, there is a gap between the funds available for eInfrastructure in the Netherlands, and the sum of user demand. Bridging this gap is perhaps the most challenging task the DNI presently faces. The difficulty stems from the widely differing sociological and cultural views on the issue. At the heart of the discussion is a choice between funding routed via the users *vs.* via the infrastructure. This section explores some initial ideas by the authors about various models that could be used to address this sustainability gap.

**via the users** In this (pay-per-use) model, scientists receive funds for computing directly from NWO or other research funders. They may spend this as they see fit. DNI would charge the scientists for use, although in practice this might not be the scientist herself, but her research community or her university or institute. This model possesses an attractive transparency ("follow the money") and also allows scientists flexibility in choosing the type of computing resource for their problem. This model lacks the advantage of economy of scale that some of the alternative mentioned later on do have, and also goes against stated policy by NWO and the EU to move towards a unified science cloud for all sciences, although we see an emerging trend towards an "adopt or explain" policy with respect to the DNI (see p. 11).

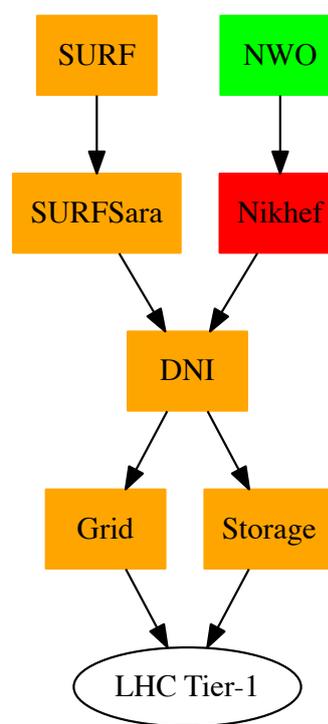
**via the institution** Research institutes are already purchasing block allocations for their faculties via a pay-for-use model. In essence, the charges for access are used to finance a part of the DNI which is allocated to the research staff. The institutes scientists no longer need to submit access requests as long as the usage fits within the block allocation. As the allocation is embedded within the DNI, fluctuations in demand can be dealt with without resorting to overprovisioning. With the increasing demand in computational and storage resources for research, we see a rise in the number of institutes that request this arrangement.

**via the infrastructure** In this scenario, scientists inform NWO (*e.g.* when writing grant applications) of how much computing is needed; NWO grants this portion of a request by allocation on the national infrastructure. The total capacities needed for the DNI can be inferred given the aggregate allocation and the observed usage patterns; funds are transferred directly from NWO to the infrastructure when replacing and/or expanding capacity. The same model applies to funding operations and support people. This latter part is already functioning in the DNI; operations people at partner sites are funded, there is an advisor team assisting scientists with mapping their problems onto the infrastructure. This model seems to be the most appropriate one for the “long tail of science” use cases, where a large infrastructure is needed to supply relatively small amounts of computing to a large number of small research groups or individual researchers.

**co-funding by power users** a fourth option exists for communities needing large amounts of computing and/or storage resources, who might have had their own infrastructures in the past. Such communities can (and in our opinion, should) include the computing costs as part of the scientific proposal to do the science. Such costs can be significantly reduced when these resources can be operated in the context of the DNI compared to constructing and operating an “own” infrastructure. This model has been used for the LHC Tier-1 in the Netherlands. A Nikhef proposal to NWO regarding upgrades to the experiments, included 3,6 M euro for computing (about 20% of the total budget). The money is being transferred into the DNI ecosystem to purchase hardware that will be operated as part of the DNI, hence the operations, data centers, and generic infrastructure to access the resources is 100% common with the rest of the DNI (and does not need to be paid for out of the high-energy physics budget).

In order to prevent fragmentation (especially for research data), the “adopt or explain” principle has been suggested, which would have data storage and computing provided via (and funded through) the DNI unless the requesting party can explain why the national infrastructure misses some important capabilities needed for the particular research; at this point a choice could be made to either charge DNI (or NLeSC) to develop the missing capacities, or to transfer funds directly to the researcher, allowing her to arrange for the computing and/or necessary developments herself.

The tension between research and commercial computing presents a final challenge. Many companies provide services that address some of the research-computing problem space. In cases like these, it would make sense to make use of such services instead of offering DNI-sponsored



**Figure 3:** Funding for the Netherlands LHC Tier-1. Funding via SURF is for operations manpower and for DNI Grid and Storage generic infrastructure; funding via Nikhef is for the computing power (not specific computers) and storage specifically allocated to the Tier-1.

equivalents. On the other hand, in many cases the commercial services only *seem* appropriate until inspected more closely. Dropbox is an example; it's widely used amongst researchers, who are generally unaware of (or don't take seriously) the legal problems associated with its use. SURF developed the *SURFdrive* service to avoid these problems, and offers this service to the Dutch research community.

Another example is cloud computing (*e.g.*, Amazon); while getting a bare-bones image from Amazon is quite cheap, it turns out that doing data-intensive distributed computing with such cloud services is quite expensive, as extra charges are assessed based on data storage and movement. This area is developing rapidly; as of this writing (end March 2016), Amazon has just announced that it will drop data egress charges [5] for academic use.

## 9. Challenges & future

The rapid pace of developments in the ICT world makes provisioning ICT services a challenge; in IT, four years is a long time. Aside from this intrinsic challenge in alignment of resources with changing user demand, we describe two areas of challenge for the future: integration amongst platforms, the data infrastructure, and the infrastructure's economic model.

### 9.1 Inter-platform integration

As explained in sec. 1, researchers desire a platform ecosystem integrated along the lines of GAFa. The cornerstone of such an integrated ecosystem is a single credential (or "login", "account") that suffices for access to all components. In the DNI, this single credential is based on the SURFconext system [6], which federates the home-institution credentials for educational and research institutions within the Netherlands. Most web-based services (such as SURFdrive) are already using the SURFconext system; support for other platforms is being developed *via e.g.* the AARC project, in which the Netherlands is a strong participant. Internationally, services allowing authentication and authorization via EduGAIN can be accessed using SURFconext credentials.

### 9.2 Innovation

DNI can remain effective by both following technology trends as well as listening closely to the scientists. The HPC Cloud and the Hadoop Big Data infrastructure at SURFsara are two excellent examples of this, the pilot infrastructures appearing in both cases in advance of the peak in demand for those infrastructures.

Innovation on the other hand is a challenging topic for a couple of reasons. Firstly, production infrastructures like ours are relied upon by many scientists in the Netherlands and elsewhere; these scientists expect a stable and reliable system. While this does not prevent innovation, careful planning is required to ensure that the innovations don't break user workflows. Secondly, much innovation is coming these days from the commercial world. These useful innovations are unfortunately plagued with disadvantages such as vendor lock-in, personal data storage outside the EU, and authentication/authorization frameworks incompatible with academic standards. Also most commercial software systems are not designed for use by multiple communities.

### 9.3 Data Infrastructure

Despite 15 years of EU Framework projects on data infrastructures, there is still no true data infrastructure for science in Europe or in the Netherlands. Models for storage, metadata, and access differ between scientific communities to a degree that makes the design, implementation and deployment of such a system very difficult. Hence we have within DNI alone, at least three different data infrastructures, and we work together with two other data infrastructures with a national mission within the Research Data Netherlands consortium.

Our scientist users have been charged by NWO to follow developing policy on data management, stewardship, publishing and preservation, which only increases the need to at least align the various infrastructures so that interoperability is possible (“data re-use”). SURF has founded the National Coordination Centre for Research Data Management to assist users, research communities, and organisations with a significant public research function such as universities and research institutes, in complying with this charge. This area is also being actively explored by the Netherlands eScience Center.

## 10. Conclusions

The Netherlands has responded to a charge from the government and organized the national e-infrastructure activities under the SURF cooperative. Much progress has been made since 2011 in integration amongst the various e-infrastructure platforms, organization of the support for the infrastructure and its connection to both local (research institute, department, or university) and international e-infrastructures. Current activities focus on yet stronger integration within the DNI, developments supporting the recent increased interest in data management and stewardship, and continual enabling of new science via targeted developments of eScience instruments and the palette of infrastructure platforms.

## 11. Acknowledgments

The authors would like to thank all colleagues from Nikhef, SURFnet, SURFsara and RUG CIT and all others who have provided input on this document. This work was carried out on the Dutch National e-Infrastructure with support of the SURF cooperative.

## References

- [1] R. Belsø, “*Researcher Perspectives on e-Infrastructure Provisioning*,” 2016, e-IRG Workshop, 9 March 2016, Amsterdam, The Netherlands. [Online]. Available: <http://e-irg.eu/documents/10920/304839/4.+Rene+Belsø.pdf>
- [2] T. K. Teal, K. A. Cranston, H. Lapp, E. White, G. Wilson, K. Ram, and A. Pawlik, “*Data Carpentry: Workshops to Increase Data Literacy for Researchers*,” *International Journal of Digital Curation*, vol. 10, p. 135, 2015.
- [3] ICTRegie, “*De ict-infrastructuur voor het wetenschappelijk onderzoek in nederland*,” 2008, english title: “The ICT-Infrastructure for Scientific Research in the Netherlands”, a report by the National Governance Board for ICT-research and -innovation.

- [4] SARA, “*Festive launch of the new HPC Cloud infrastructure,*”  
<https://www.surf.nl/en/agenda/2011/10/festive-launch-of-the-new-hpc-cloud-infrastructure/festive-launch-of-the-new-hpc-cloud-infrastructure.html>, 2011, [Online; accessed 31-March-2016].
- [5] Amazon, “*AWS Offers Data Egress Discount to Researchers,*”  
<https://aws.amazon.com/blogs/publicsector/aws-offers-data-egress-discount-to-researchers/>, 2016, [Online; accessed 31-March-2016].
- [6] P. van Dijk, “*SURFconext, showcasing a new collaboration paradigm,*” 2011, terena Networking Conference TNC2011, 16 - 19 May, Prague, Czech Republic. [Online]. Available: <https://tnc2011.terena.org/core/presentation/39>