



eScience development and experiences in The Netherlands

Wilco Hazeleger · Tom Bakker
Rob van Nieuwpoort

Introduction

Historically, researchers have been aided by scientific instruments and methodologies to address their scientific questions. In the modern era, digital technology and methods are omnipresent, and their use is quintessential to current scientific discovery. It has been recognized that ever growing data sizes and increased availability of disparate data sets pose tremendous opportunities for research, but effective use of these data challenges scientific researchers [5]. Moreover, computer capabilities keep increasing, while computer architectures continuously change. Data and computer challenges are tightly coupled as analyses of large and disparate data sets require modern computational resources.

The “big” sciences, such as high energy physics, astronomy, materials science, earth and climate science, are prime examples of research domains that have developed strongly over the past decades due to the availability and development of computational resources [2, 3]. With the rise of data availability and with developments in novel applied mathematical techniques, such as machine learning, new domains are increasingly profiting from the digital revolution. In different aspects of the life sciences this started about a decade ago, drug discovery [7] and genomics being well-known examples, and increasingly also scholars in social sciences and humanities (e. g. [12]) are using digital technology and analytics techniques. In general, scientific research is becoming more data intensive and inductive of character. So, in modern research, nearly all researchers also become data scientists, the complexity of projects increases, and research is inherently multidisciplinary.

eScience effectively applies modern e-infrastructure (computing, storage, networks) and utilizes research software and data to address challenging scientific questions posed in research disciplines. eScience is developing into a scientific field itself as well. eScience has engineering aspects, similar to civil or chemical engineering, by applying novel technology and methods to other application domains. These approaches themselves, developed at the interface of domain challenges and computer and data science, result in questions that are addressed with a scientific approach. For instance, scaling up or scaling out using high performance and distributed data and computer resources is a typical eScience challenge. Moreover, new analytics methods and access to distributed data, developed in the domains of data and computer sciences, generate new scientific challenges when effectively applied.

In the Dutch research landscape, The Netherlands eScience Center (the eScience Center from now on) was set up in 2011 to effectively build an interface between domain research and e-infrastructure, as well as an interface between computer and data science and domain research. This need was recognized after successful large projects on eScience

<https://doi.org/10.1007/s00287-018-01132-3>

© The Author(s) 2018. This article is available on SpringerLink.com with Open Access.

Wilco Hazeleger · Tom Bakker · Rob van Nieuwpoort
Netherlands eScience Center,
Amsterdam, The Netherlands
E-Mail: {w.hazeleger, t.bakker, r.vannieuwpoort}
@esciencecenter.nl

Wilco Hazeleger
Wageningen University,
Wageningen, The Netherlands

Rob van Nieuwpoort
University of Amsterdam,
Amsterdam, The Netherlands

Abstract

The Netherlands eScience Center is the national expertise center for the development and application of research software. Collaborating with researchers from all academic disciplines, it extends the breadth and depth of research opportunities by exploiting the latest insights from computer and data science, as well as making optimal use of hardware, software, and data infrastructures. It does so through problem-driven research projects where eScience research engineers, employed by the eScience Center, collaborate with researchers in all disciplines at Dutch academic institutions. Project software is generalized and made available for reuse for other disciplines and goals. The center has three main technological competences: efficient computing, optimized data handling, and data analytics. Furthermore, on the national level it coordinates and contributes to science policies on computing, data, and applications thereof. With its two main assets, a staff of highly educated and multi-disciplinary eScience Research Engineers and an open online directory of research software tools and knowledge, it successfully contributes to the Dutch scientific landscape and enhances and accelerates all research in The Netherlands and beyond.

Its strategy and implementation will be discussed below, as well as examples of eScience projects.

Strategy

In response to the changing research environment described in the Introduction, we developed a strategy that positions the eScience Center as an organization with interfaces to domain research, computer and data science, and e-infrastructure, while delivering eScience services and output. Its strategy is based on the mission, which is to *enable digitally enhanced research through efficient utilization of data, software, and e-infrastructure*. This translates into four tasks: to enable scientific breakthroughs, to collaborate on problem-driven projects, to develop versatile cross-disciplinary eScience tools, and to coordinate eScience activities (see the strategy: <https://www.esciencecenter.nl/about/strategy>).

It is crucial that the eScience Center takes the domain research perspective as a starting point and works in a problem-driven manner, primarily through calls for collaborative project proposals. Figure 2 summarizes the strategy and shows the different interfaces. It should be read as an arrow from left to right, with the eScience Center in the middle and with different domains and institutions feeding into the eScience Center. The right-hand side shows the outcome of our activities. The strategy is explained in detail below.

Interface with domain research

The eScience Center addresses all research disciplines (top part of Fig. 2), but distributes its efforts in the domains over Physics & Beyond (this includes high and low energy physics, fluid dynamics, materials sciences, and astronomy), Environment and Sustainability (this includes climate research, ecology, energy research), Life Sciences and eHealth (this includes genomics, proteomics, medicine), and Humanities and Social Sciences. The eScience Center's work is driven by scientific problems from these disciplines. We grant, in competition, collaborative

(e.g., VL-e; Virtual Laboratory on eScience) and the successful integration of academic e-infrastructure facilities and services of SURF. SURF (the collaborative organisation for ICT infrastructure in Dutch education and research) and NWO (The Netherlands Organization for Scientific Research, effectively the Dutch research council) jointly set up the eScience Center. The eScience Center is positioned between scientific research and e-infrastructure in the national landscape and connects the national e-infrastructure with academic disciplines (see Fig. 1).

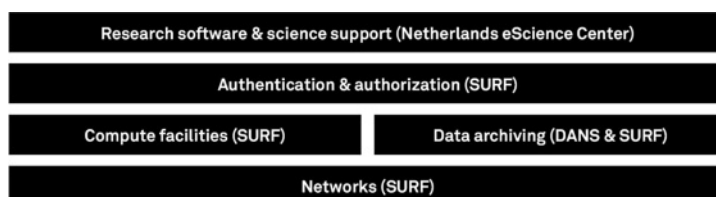


Fig. 1 Positioning of the Netherlands eScience Center in the national digital infrastructure for science and education

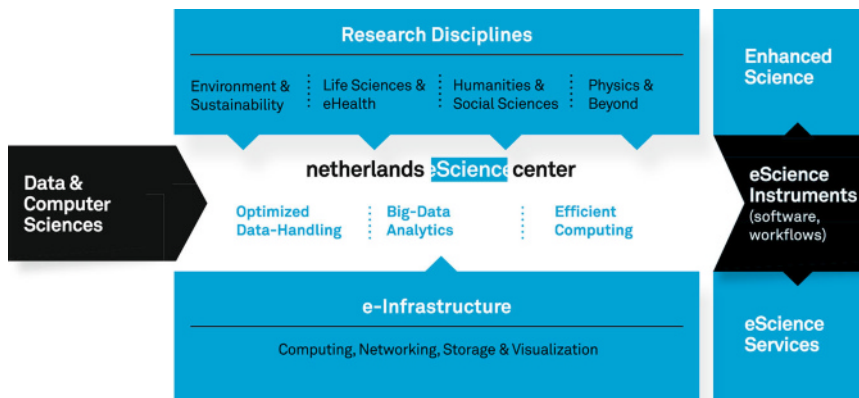


Fig. 2 Schematic of the strategy of The Netherlands eScience Center, focusing on the eScience technical competences (optimized data handling, big data analytics, and efficient computing) and the different interfaces

project proposals for 1 to 4 years, in which one or more eScience Research Engineers work with PhD students or postdocs at a university or academic institution (see “Implementation of the strategy” below). The four domains are deliberately chosen such that there is a balance in the level of literacy with digital technologies. That is, the four domains do not compete amongst each other for eScience support, but are all addressed. This also facilitates reuse of technologies and cross-fertilization of digital technologies among research disciplines.

Interface with computer and data science

Computer and data sciences are very quickly developing research disciplines. New opportunities in digital technologies and analytics techniques are investigated and developed in computer and data sciences. For the eScience Center to continue to be up to date with recent developments and scout relevant technologies and techniques for applications in domain research, collaborative projects are set up with computer and data scientists. These are also competitive multiyear projects, similar to projects

with application domains, but with the specific objective to strengthen our own capabilities in our three main technological competence areas: efficient computing, data analytics, and data management.

Interface with e-infrastructure

Computing, data, and network facilities are important tools in digitally enhanced research. The eScience Center does not have its own hardware facilities but uses state-of-the-art e-infrastructure facilities that match the research needs at hand. SURF is the main national e-infrastructure provider (with SURFsara providing high performance computational and storage facilities), but we also use local facilities at universities and commercial (cloud) providers. There is a preference to use national public facilities, but other facilities can be used when needed and when they match the scientific problem at hand better. In each project, a consultant from SURF is available for individual support and advice. In total, over half of the projects use SURFs e-infrastructure (Fig. 3), while there is also a lot of use of local facilities and commercial providers of

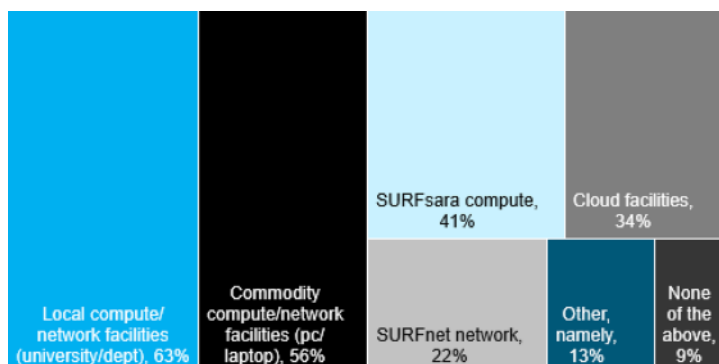


Fig. 3 Use of e-infrastructure in eScience projects based on a survey of 38 projects conducted with support from the eScience Center

infrastructure, in particular the different commercial cloud providers. There are large differences in e-infrastructure use in the different disciplines, so a tailored approach is needed.

Outcomes of the eScience Center

The outcomes of eScience Center projects are peer-reviewed papers and research software, both in the domain sciences and in eScience. A leading principle of the eScience Center is to develop and apply reusable and sustainable open-source software within multiple projects. To promote this principle and to increase the chance for growth and adoption, we set up an online research software directory to increase the adoption of research software that we develop, enhance, and have expertise with. Some of the projects reach beyond the realm of academic research and develop into services for communities (e.g., in the health domain, an infrastructure for translational research has been set up).

Implementation of strategy

Call strategy

The eScience Center fulfils its mission primarily by carrying out collaborative scientific research projects. These projects are selected after a call for proposals to the academic community. Potential project leaders from academia are invited to develop proposals, matching their purely scientific ambition with the digital technology and methods needed to achieve that. The typical project size is 500 k€, half of which is funding for a PhD or postdoc position at the academic partner institution, while the other half constitutes in-kind contribution to the project by one or more eScience Research Engineers employed by the eScience Center. The project proposals are evaluated on their scientific quality and the eScience quality. This last criterion includes the potential for impact on other disciplines, the sustainability of the tools beyond the duration of the project, and the innovation for the problem at hand. The current success rate of applications is low due to the very large demand (~10–15 % of the proposals are granted funding). Typically, one project in each of our four scientific “domains” is funded, while also two collaborative projects are granted with researchers from computer and data science. Recently, in collaboration with other funding programs of NWO, the number of granted projects has increased. By col-

laborating with existing funding programs, more and larger-impact projects can be set up in existing research communities. Recent examples include a joint call with CLARIAH, a humanities research infrastructure, and joint calls with the FOM/Shell program on computational sciences for energy research. Since 2011, up to 60 large projects have been conducted, distributed over all scientific domains, and covering almost all Dutch universities.

eScience Research Engineers

The *eScience Research Engineers* require special attention in this article, as they play an important role in the implementation of eScience in The Netherlands. They are digital specialists at academic level (MSc to PhD level up to associate professor level). They are employed by the eScience Center, where they follow a career track with a technological, research or managerial emphasis. They are specialists in one or more of our three major technological competence areas (efficient computing, data analytics, and data management). They typically spend two-thirds of their time at the academic research institution where eScience projects are executed, while the remaining time is spent at the eScience Center in Amsterdam. Here, they collaborate with other eScience Research Engineers, sharing best practices, tackling technological project challenges, and performing joint research and development activities. On purpose they have been labeled *research engineers*, as they perform not only software development and engineering tasks but are part of a research team. For instance, they jointly publish the results in peer reviewed literature. Each project has been assigned one or more eScience Research Engineers, depending on the skills and competences required. This allows us to effectively combine and contribute disparate skills to an individual project (e.g., data visualization and high-performance computing). For regular academic groups, this is difficult to achieve. The eScience Center currently (mid-2018) employs more than 50 eScience Research Engineers.

Other advantages of our approach are that the research engineers can identify and exploit opportunities for novel methodologies and the reuse of existing tools *across different disciplines*. Because they collaborate in many inter-disciplinary projects in different domains, they have a birds-eye view of technology in science, allowing them to make quick and efficient progress, using the right tool for the

task. Moreover, they develop software with general applicability and reuse in mind from the start.

eScience technologies

As mentioned before, the eScience Center's competences are focused on efficient computing, big data analytics, and optimized data management. The required skills largely depend on project demands, which requires a flexible and adaptive staff. This is accomplished through continuous training and working in teams leading to consistent internal knowledge management and through mobility of staff to and from the eScience Center.

eScience technologies vary from algorithms to the tools for utilization of advanced computer infrastructures and virtual research environments (e.g., computer kernels, libraries, scientific workflows, data tools and applications to use storage facilities, high-speed networks, visualization equipment). Another relevant development is the emergence of the user-friendly Jupyter Notebooks, which has a significant impact on reuse and reproducibility. This technology supports the open science development, which is stimulated by the Dutch government and European Committee, for instance as part of the European Open Science Cloud.

Reuse and generalization

An important aspect of the development and application of eScience technologies is generalizing tools and making them applicable and re-usable in different disciplines. The importance of academic software is increasingly recognized (e.g. [1]). To enable reuse of our software, we make it publicly accessible, engage in active online and offline promotion, and use open software licenses. To increase and maintain software quality, we use standard off-the-shelf cloud-based software testing and source code quality assessment tools. To maximize transparency, all code and test results are publicly available on our GitHub repositories. The eScience Center created a checklist¹ containing the essential steps for creating high quality software, which is applied to all tools developed. The software is – if possible – released under the permissive open-source Apache 2.0 license². For documentation and

data sets, we use the permissive Creative Commons licenses.

To further improve findability and re-usability, we set up the Research Software Directory: <https://www.research-software.nl> (Fig. 4). The aim is to provide an extensive and curated online portal of advanced scientific software, including (1) our own software, (2) external software to which we have made contributions, and (3) externally developed software that we have significant expertise with. The added value compared to our GitHub repositories is that it helps researchers to get a high-level overview of available relevant software and related eScience Center projects. Moreover, it has increased the visibility of our contributions to leading external software projects, such as ROOT, Amuse, OpenDA, and libLAS. We associate DOI's with all software releases, stimulating software citation. This is equally important for scientific reproducibility, and for giving credit to software as research output. The latter is key for the career perspectives of research software engineers.

A major recent development in this respect is the use of Jupyter Notebooks. These offer a web-based interface for interactive programming in Python, R, and a wide range of other languages. This has vastly improved accessibility to a wider audience with less experience in using digital tools. Furthermore, we often package our software. Because many programming languages have package managers that offer both a central location to find software and simple tools to install it, we release software via package managers such as PyPI and Anaconda for Python, NPM for JavaScript, and Maven for Java. Successful examples are ROOT Conda recipes, to which we contributed, and which have been very successful in the high-energy physics community. Finally, the eScience Center increasingly uses Docker virtualization to distribute end-user applications. "Dockerizing" software decreases the dependency problems often encountered in legacy applications.

In addition to the Research Software Directory, a shared knowledge base is maintained (<http://knowledge.esciencecenter.nl>) to share guides for software development, tutorials, and technical reports. One of its goals is to retain knowledge when projects end, or engineers leave the eScience Center.

Example eScience projects

In this section, we highlight a few of the eScience collaborative projects in different domains.

¹ <https://nlesc.gitbooks.io/guide/content/>.

² Apache 2.0 allows anyone to re-use our software in any way, including commercial application.

Research Software Directory

Encouraging the re-use of research software

The Research Software Directory aims to promote the impact, the exchange and re-use of research software. Please use our tools! [Read more](#)

Start typing here to search for software

Q

Sort by: Last updated

Tags -

- ☐ Big data (19)
- ☐ GPU (2)
- ☐ High performance computing (13)
- ☐ Image processing (5)
- ☐ Inter-operability & linked data (9)
- ☐ Machine learning (6)
- ☐ Multi-scale & multi model simulations (4)
- ☐ Optimized data handling (8)
- ☐ Real time data analysis (5)
- ☐ Text analysis & natural language processing (13)

Case Law App

Ca

Web application for exploring the citation network of Dutch court decisions.

6 days ago

★ Featured

Noodles

No

Task-based parallel programming model in Python. Run complex workflows on large computer clusters or parallelize codes on your laptop: Noodles offers the same intuitive interface.

6 days ago

★ Featured

nlppln

nl

For text mining researchers that want to build workflows, nlppln is the most flexible solution among NLP toolkits, because it allows you to

Kernel Tuner

Ke

Kernel Tuner greatly simplifies the development of highly-optimized and auto-tuned CUDA, OpenCL, and C code, supporting many advanced

Fig. 4 Screenshot of the research software directory

In the *Summer in the City* project, we co-developed a very detailed weather model of the city of Amsterdam in collaboration with Wageningen University, integrating data sources of the earth's surface and near surface atmosphere. The project used existing software (the WRF mesoscale model), with the eScience challenge to obtain realistic bottom boundary conditions (i.e., the land surface characteristics) by integrating data sets. These include satellite data, weather and hydrologic data, and socio-economic data. Such improved boundary conditions and higher resolution improved the quality of local weather forecasts at the sub-kilometer scale. The scientific results published within the project were followed up by a successful grant application by one of the researchers in the project, which ensures further scientific development of meteorology in an urban environment. The results

are described in detail in a recent paper by Ronda et al. [8].

In the *eSalsa* project, we enabled global ocean modeling at very high spatial resolution using GPU computing technologies in collaboration with Utrecht University. An existing numerical ocean model code was optimized for execution on the CPU-GPU Cartesius supercomputer of SURF-Sara [13]. Century scale simulations, up to 2100 under climate change, are extremely challenging at the eddy-resolving spatial resolution (equivalent to about 10 km horizontal scale; [4]). Also, natural variability arising in the climate system needs to be sampled to detect changes forced by anthropogenic activities. Hence, to accelerate the numerical simulations, using eScience efficient computing expertise is highly relevant. Scientific publications on regional sea level rise projections were published,

and the work on high resolution global coupled climate modeling is now followed up in the H2020 funded PRIMAVERA project. Also, the GPU expertise developed here, including a kernel tuner to find optimal set ups in a large parameter space, is reused in many other projects where accelerated computing is addressed, such as radio astronomy (e. g. [10]).

The FAIR data principles (FAIR = findable accessible interoperable re-usable) were initiated with eScience at the heart. An eScience Center sponsored workshop in 2014 resulted in a white paper focusing on data stewardship. Later, a paper appeared in Nature Scientific Data on the FAIR principles with eScience co-authorship [14]. We applied the FAIR approach in the *CANDYGENE* project funded by the eScience Center. Integrating different software tools, we set up a FAIR data port for biological data [11], from genotype to phenotype of tomatoes. Currently, we are implementing aspects of FAIR in many other domains, such as in the eScience funded *AAalert* project for radio astronomical data and in the *MAGIC* project, funded by Copernicus Climate Change Services (C3S), for climate data.

In the *Embodied Emotions* project, the eScience Center collaborated with VU University in Amsterdam and demonstrated that a multi-label text classification approach to learning complex emotion models on historical text is feasible. In particular, we analyzed theater texts with respect to the development of emotions used in theater plays. The performance of the Historic Embodied Emotion Model (HEEM) developed by the project team is similar to the performance for simpler emotion models. Comparing HEEM to LIWC, an existing dictionary-based sentiment analysis tool, HEEM yields finer-grained results. The project's novel approach in sentiment analysis has the potential to have a transformative impact on the humanities domain. The substantially more detailed and fine-grained analysis method enables researchers working in sentiment mining to apply the method to their own research questions and possibly generate new and more detailed theoretical insights. The analyses over time of the use of emotions were not possible before [6].

In the *eAstronomy* project, we accelerated the pipeline of signal processing in radio astronomy. We developed novel GPU algorithms for the real-time signal processing system used in LOFAR, the largest radio telescope in the world. In collaboration with

ASTRON, we also created new methods to remove interference from the signal in real time. The IBM Blue Gene supercomputer that was initially used could be replaced with a small GPU cluster, reducing operational cost and power usage, and making the instrument more sensitive. In addition, we developed a pulsar searching pipeline on GPUs, which is now used in production for Apertif in the Westerbork telescope. This pipeline is an order of magnitude faster than all earlier pipelines, allowing astronomers to survey a much larger part of the sky [9].

Coordinating activities

As part of the national e-infrastructure, the eScience Center is also engaged in coordinating and training activities. The eScience Center has set up and leads ePLAN (<https://escience-platform.nl/>), the national platform for eScience and data research centers. This brings together institutions engaged in data research and eScience to share insights and have a single voice in the policy arena. Example activities include a survey among Dutch researchers on e-Infrastructure needs³, a survey on needs for the European Science Clouds, and workshops on, for instance, software sustainability, and the FAIR principles in different domains.

A similar activity has been initiated at the European scale. PLAN-E is the platform of European eScience Centers (<https://plan-europe.eu/>). Its members form a vital link between domain researchers and e-infrastructure providers in particular. Members from over 20 countries are represented. PLAN-E is a stakeholder to the European Committee and, for instance, advises the Committee on the development of the European Open Science Cloud. It is crucial that its view on current and future e-infrastructures is based on the perspectives from multiple scientific disciplines. Members meet at least once per year to discuss ongoing matters and to align their activities.

The eScience Center is strongly involved in the national science programming arena, in particular when it is related to digitization of research. It co-leads the National Science Agenda activities on big data and jointly works with SURF on implementing the FAIR principles in scientific research.

Training the next generation scientists on data science and data stewardship is done by university

³ https://www.esciencecenter.nl/pdf/Infrastructuur_duurzaam_op_maat_versie_1.0B_DEF_03-03-2016.pdf.

partners. The eScience Center contributes by offering Data and Software Carpentry workshops where essential data and programming skills are trained. Furthermore, we provide tutorials of research software packages that we develop and maintain at the eScience Center.

Beyond Europe, the eScience Center is visible at domain and computer science conferences, as well as the international eScience Conference. With IEEE, we are co-organizing the 14th International Conference on eScience in Amsterdam, where both eScience applications in specific scientific fields and advancements in computer and data science will be addressed.

Summary and conclusions

The Netherlands eScience Center is the Dutch national expertise center for the development and application of research software to pioneer new scientific horizons. It has been successful in collaborating with domain scientists since its inception in 2011, building upon earlier activities that were more programmatic of nature. Crucial is that it takes the application-domain perspective by translating research needs into software that takes advantage of available technology and data from all disciplines, computer and data science, and providers of hardware infrastructure. An important part of the success is the close collaboration of eScience Research Engineers, employed by the eScience Center, with researchers at Dutch academic institutions. The eScience Research Engineers have a skill set that will be increasingly required to achieve scientific advancements in all academic fields that depend on digital methods. Combining the physical central location where staff work and learn, together with distributed local environments where engineers collaborate with academic partners, has contributed to the success of the eScience Center. The demand for eScience projects substantially exceeds our financial capacity and has led to the advice to the national research council to augment the role of the center. The eScience Center offers career perspectives to eScience Research Engineers within and outside the center. Given the increased use of digital tools and the increased multidisciplinary character of scientific practice, such a career development is essential for the future development of scientific research. In this context, software citation is equally important for scientific

reproducibility, and for giving credit to software as research output.

eScience crucially contributes to open science through its application of open science standards (open access and open source of data, software, and publications), but also through the way it works. Re-using software, workflows, and data is at the heart of the center. The tools and applications developed are generalized and made available for reuse. The reuse and the sustainability of research software is a major challenge, though. While data management, data stewardship, and related data reuse are high on research policy agendas and accepted by many researchers, the sustainability and reuse of software and workflows is still underdeveloped. Software sustainability, although supported by the eScience Center, does not have a clear enough place in the national and international e-infrastructure, while it is very much needed for open and reproducible science.

The eScience Center has been very successful in enhancing and accelerating research in a wide variety of scientific disciplines, based on the steep rise of scientific publications to which eScientists contributed. However, it remains a challenge to have a much wider impact beyond individual scientists. Setting up multi-year large projects with one research group that is well connected to other researchers may partially remedy this. A more coordinated international effort is needed as well, though.

The link with the e-infrastructure is crucial. While the national e-infrastructure is used extensively, which can be considered to be a success, the feedback from the research to the e-infrastructure development itself needs attention. By taking the domain perspective and making work problem driven, eScience can provide feedback about e-infrastructure development. Coordinated activities, such as ePLAN and PLAN-E, do so, but this needs to be strengthened. For instance, the development towards data-intensive science and the increased use of AI needs to be rapidly responded to by the national e-infrastructure developments.

Further opportunities beyond academic research exist. Although the focus is on scientific research, collaboration with the private sector may further increase the impact of eScience. While the eScience Center has been engaged in some activities, this is still in development.

In summary, the Dutch eScience activities, culminating in the activities of The Netherlands

eScience Center, have successfully contributed to digitally enhanced scientific research. It is unique as a national institution and expertise center, employing its own highly skilled staff that collaborates with academia. With research software at its heart, it has developed as a crucial part of the Dutch national research infrastructure.

Open Access. This article is distributed under the terms of the Creative Commons Attribution Non-commercial License (<http://creativecommons.org/licenses/by/4.0/deed.de>) which permits any non-commercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Allen A, Aragon C, Becker C, Carver J, Chis A, Combemale B, Croucher M, Crowston K, Garijo D, Gehani A, Goble C, Haines R, Hirschfeld R, Howison J, Huff K, Jay C, Katz DS, Kirchner C, Kuksenok K, Lämmel R, Nierstrasz O, Turk M, van Nieuwpoort RV, Vaughn M, Vinju JJ (2017) Engineering Academic Software (Dagstuhl Perspectives Workshop 16252). Dagstuhl Manifestos 6. <https://doi.org/10.4230/DagMan.6.1.1>
- Bauer P, Thorpe A, Brunet G (2015) The quiet revolution of numerical weather prediction. *Nature* 525:47–55, <https://doi.org/10.1038/nature14956>
- Bird I (2011) Computing for the Large Hadron Collider. *Annu Rev Nucl Part Sci* 61:99–118, <https://doi.org/10.1146/annurev-nucl-102010-130059>
- Brunnabend SE, Dijkstra HA, Kliphuis MA, Bal HE, Seinstra FJ, van Werkhoven B, Maassen J, van Meersbergen M (2017) Changes in extreme regional sea level under global warming. *Ocean Sci* 13:47–60, <https://doi.org/10.5194/os-13-47-2017>
- Hey T, Tansley S, Tolle K (2009) *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Microsoft Research, Redmond, WA, ISBN 978-0-9825442-0-4
- Leemans I, van der Zwaan JM, Maks I, Kuijpers E, Steenbergh K (2017) Mining embodied emotions: a comparative analysis of sentiment and emotion in Dutch texts, 1600–1800. In: *Digital Humanities Quarterly* 11 <http://www.digitalhumanities.org/dhq/vol/11/4/000343/000343.html>
- Lusher SJ, McGuire R, van Schaik RC, Nicholson CD, Vlieg J de (2014) Data-driven medicinal chemistry in the era of big data. *Drug Discov Today* 19:859–868, <https://doi.org/10.1016/j.drudis.2013.12.004>
- Ronda RJ, Steeneveld GJ, Heusinkveld BG, Attema JJ, Holtslag AAM (2017) Urban finescale forecasting reveals weather conditions with unprecedented detail. *Bull Amer Meteor Soc* 98:2675–2688, <https://doi.org/10.1175/BAMS-D-16-0297.1>
- Sclocco A, Bal HE, van Nieuwpoort RV (2015) Finding Pulsars in Real-Time. In: 2015 IEEE 11th International Conference on e-Science, pp 98–107, <https://doi.org/10.1109/eScience.2015.11>
- Sclocco A, van Leeuwen J, Bal HE, van Nieuwpoort RV (2016) Real-time dedispersion for fast radio transient surveys, using auto tuning on many-core accelerators. *Astron Comput* 14:1–7, <https://doi.org/10.1016/j.ascom.2016.01.001>
- da Silva Santos LB, Wilkinson MD, Kuzniar A, Kaliyaperumal R, Thompson M, Dumontier M, Burger K (2016) FAIR data points supporting big data interoperability. In: Zelm M, Doumeingts G, Mendonca JP (eds) *Enterprise Interoperability in the Digitized and Networked Factory of the Future*. ISTE, London, pp 270–279, ISBN 978-1-8470404-4-2
- Warwick C, Nyhan J, Terras MM (eds) (2012) *Digital Humanities in Practice*. Facet, London, ISBN 978-1-8560476-6-1
- van Werkhoven B, Maassen J, Kliphuis MA, Dijkstra HA, Brunnabend S-E, van Meersbergen M, Seinstra FJ, Bal HE (2014) A distributed computing approach to improve the performance of the Parallel Ocean Program (v2.1). *Geosci Model Dev* 7:267–281, <https://doi.org/10.5194/gmd-7-267-2014>
- Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, Blomberg N, Boiten J-W, da Silva Santos LB, Bourne PE, Bouwman J, Brookes AJ, Clark T, Crosas M, Dillo I, Dumon O, Edmunds S, Evelo CT, Finkers R, Gonzalez-Beltran A, Gray AJG, Groth P, Goble C, Grethe JS, Heringa J, Hoen PAC 't, Hooft R, Kuhn T, Kok R, Kok J, Lusher SJ, Martone ME, Mons A, Packer AL, Persson B, Rocca-Serra P, Roos M, van Schaik RC, Sansone S-A, Schultes E, Sengstag T, Slater T, Strawn G, Swertz MA, Thompson M, van der Lei J, van Mulligen E, Velterop J, Waagmeester A, Wittenburg P, Wolstencroft K, Zhao J, Mons B (2016) The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3:16–18, <https://doi.org/10.1038/sdata.2016.18>