



netherlands

eScience center

Call for Proposals

Accelerating Scientific Discovery in the Arts and Humanities (ADAH)

September 2016

The purpose of this Call for Proposals is to enable researchers working in the Arts and Humanities to address compute-intensive and/or data-driven problems within their research and to contribute to a generic and sustainable eScience infrastructure.

The call aims to fund proposals that pursue new scientific domain challenges and that enhance and accelerate the process of scientific discovery within the Arts and Humanities using eScience technologies. To achieve this, submitted proposals should be characterized by the usage, adaptation, and integration of existing methods and tools, as made available through the CLARIAH and NLeSC software infrastructures. Newly developed tools should be made available through the eScience Technology Platform (eStep) of the Netherlands eScience Center (NLeSC) for potential use in other studies.

CLARIAH and NLeSC invite researchers whose main area of expertise is in the Arts and Humanities to submit a proposal in this call.

Den Haag / Amsterdam, August 2016 - September 2016
CLARIAH / Netherlands eScience Center

1 Introduction

1.1 Background

There is an urgent need for methods and tools that help manage and exploit the rapidly increasing wealth of available scientific data. This urgency is no longer felt solely in headline-grabbing and inherently large-scale initiatives such as the Human Genome Project or the Large Hadron Collider, but is rapidly permeating virtually all levels and directions of (experimental) research. Having the methods and tools available to effectively handle very large quantities of research data, the process of scientific discovery – from initial idea to new scientific breakthrough – can be accelerated significantly (also referred to as eScience).

eScience and NLeSC

eScience is a multi-disciplinary pursuit concerned with bridging the gap between the need for modern data- and compute-intensive research on the one hand, and the capabilities of state-of-the-art ICT on the other. This ranges from high-performance computing and networking, to large-scale data management, and novel data analysis methods. The challenge of eScience is not only to ensure that most value is gained from new research endeavors, but also to exploit the often problem- or domain-independent nature of the underlying eScience requirements. As such, eScience aims to deliver sustainable and discipline overarching solutions, made available for reuse and tailored to problem-specific needs. Also, eScience aims to not reinvent the wheel, but rather to utilize and creatively combine existing methods and tools as much as possible.

The Netherlands eScience Center (NLeSC) supports and reinforces multi-disciplinary compute- and data-intensive research in the Netherlands through creative and innovative use of eScience in all its manifestations. To achieve this, NLeSC fosters research collaborations aiming to accelerate scientific breakthroughs by effectively utilizing advanced ICT technologies, and by making large-scale data analysis possible across multiple disciplines. NLeSC collaborates closely with partners such as SURFsara and SURFnet, academic and non-academic institutes, various commercial technology providers, and universities.

Humanities and CLARIAH

The availability of massive quantities of digital sources (textual, audio-visual and structured data) for research is revolutionizing the Arts and Humanities. Top-quality humanities scholarship of today and tomorrow is therefore only possible with the use of sophisticated ICT tools. The Common Lab Research Infrastructure for the Arts and Humanities (CLARIAH) aims to offer humanities scholars a 'Common Lab' that provides them access to large collections of digital resources and innovative user-friendly processing tools, thus enabling them to carry out groundbreaking research to discover the nature of human culture.

Information on culture and on how culture and cultural identities evolved in the past is hidden in enormous amounts of unstructured data, such as texts written and printed over a period of more than 20 centuries, artefacts created over an even longer period, and audio-visual documents produced since the last quarter of the 19th century. In addition to

these unstructured data, we have (semi-)structured data, partly structured-born (such as vital records), partly made structured by the labor of generations of researchers in the humanities and social sciences.

Until recently, research addressing questions about culture and identity depended on experts' abilities to identify potentially relevant pieces of information in archives, libraries and museums. Digital data and sophisticated ICT tools will enable humanities researchers to efficiently incorporate huge quantities of data into their research, thus boosting the empirical basis of their research, enabling them to address old research questions in completely new ways, and possibly even formulate completely new research questions.

CLARIAH is a national project that is designing, constructing and exploiting the Dutch parts of the European CLARIN and DARIAH infrastructures. CLARIAH covers the humanities as a whole but has three core discipline areas: linguistics, media studies, and socio-economic history.

1.2 Available Technological Competences

NLeSC

NLeSC develops, scouts and applies the ICT technologies and software best suited to solving research questions in various domains. The set of technologies applied at NLeSC is therefore inherently dynamic. However, NLeSC has identified a series of core technologies that underpin existing projects and that are likely to continue to be crucial for the foreseeable future. Maintaining at least operational expertise in these areas allows NLeSC to serve as a valuable and even essential partner in a large variety of research projects dealing with data- and compute-intensive problems.

Today, NLeSC's core technological competences include:

- **Optimized Data Handling**, incl. real-time data analysis, database optimization, data interoperability, combining structured and unstructured data;
- **Big Data Analytics**, incl. data exploration, analysis, data mining, machine learning, text analysis, natural language processing, statistics and visualization, structured and unstructured data, from data to information to insight;
- **Efficient Computing**, incl. use of high-performance and distributed computing (e.g. Grid, Cloud), heterogeneous computing, efficient algorithms, use of accelerator hardware (e.g. GPUs), and green computing.

CLARIAH

CLARIAH creates a research infrastructure which aims to incorporate a wide range of data and software that are relevant for Arts and Humanities researchers, incl.:

- **Data**: Textual, audio, audio-visual data, pictures and structured data, and combinations of these.
- **Advanced Enrichment Tools**: tools to enrich data automatically with all kinds of annotations, e.g. linguistic annotations, links to external knowledge sources, transcriptions of audio, interpretation of images, etcetera.

- **Intelligent Search** in data and their metadata, e.g. search for complex linguistic annotations, semantic search, topic-based search, search in structured data, including linked open data. Both aggregated and federated search in data.
- **Analysis and Visualization:** Advanced analysis tools for analyzing the data and their metadata and the results of search actions. Tools to visualize these analyses in multiple ways, including geographic visualization.

Many data and tools with the functionality described above are already available for CLARIAH through the CLARIN-NL, CLARIAH-SEED, Clio-infra and HSN projects.

1.3 Purpose of this call

The call focusses on the *acceleration of the process of scientific discovery* in **the Arts and Humanities**.

CLARIAH and NLeSC welcome project proposals in the Arts and Humanities with significant compute- and/or data-intensive challenges at their heart, and with clear relevance to the software infrastructures currently developed by CLARIAH and by NLeSC. A competitive proposal aims to undertake novel scientific discovery in one of the humanities discipline areas. The process of scientific discovery must be driven by a clearly defined set of scientific research questions originating from that discipline area. Solving these questions is made possible (only) by the utilization, combination and/or development of scalable, robust, and sustainable methods and tools for optimized data handling, Big Data analytics, efficient computing, or beyond. As such, a competitive proposal combines a clear, credible, and novel scientific goal with a technical solution providing the computing capabilities, data management systems and/or analytics required to support that goal.

A competitive proposal stresses the expected impact of the proposed solutions on the stated research questions. Moreover, it will indicate in what way the proposed solutions may have value for other research problems, in the context of CLARIAH, and preferably for other scientific domains as well. Newly developed ICT technologies and software solutions should be made available through NLeSC's eScience Technology Platform, eStep, for potential use in other studies. Next to this potential for reuse and generalization, a successful proposal provides a sustainability plan, describing the measures taken to ensure the usability and availability of the provided solutions beyond the duration of the project itself.

CLARIAH and NLeSC specifically request proposals that aim to develop (standards-based) open source/open access solutions¹, where possible extending or working in concert with the technological competences, methods, and software tools of CLARIAH and NLeSC (see Appendices A and B).

¹ Applicants are asked to endorse and follow the [NLeSC Strategy towards Publishing, Licensing, and IP](#). For alternative agreements, contact NLeSC before proposal submission.

As part of this call, we stimulate the link between scientific research and the capabilities of advanced e-Infrastructure (e.g., high-performance computing, large-scale data storage, lightpath connectivity; see also Appendix C). For this reason, further infrastructural support services may be provided by SURF to proposals with clear e-Infrastructure needs.

Typically, the project leader will be working in (and have leading expertise in) the Arts and Humanities. However, given the nature of many of today's compute- and data-intensive research problems, proposals of multidisciplinary teams to achieve their scientific goals are specifically welcomed. Public-private collaborations are positively valued, but the inclusion of one or more industrial partners as part of the research team is not a necessary requirement.

1.4 Available budget

Three grants are available in this call, each consisting of two parts: 1) an 'in cash' contribution for the employment of local research personnel and other expenses, and 2) an 'in kind' contribution in the form of highly skilled eScience Research Engineers employed by NLeSC, whose time is allocated to the awarded projects.

The total available budget in this call consists of 300K Euros in cash funding and 4.5 FTE² in kind funding. With three projects to be awarded, a typical project will receive € 100,000 and 1.5 FTE contribution from one or more eScience Research Engineers.

1.5 Validity of the call

The deadline for the mandatory proposal abstract is **15 November 2016, 14:00 CET** and the deadline for full proposals is **16 December 2016, 14:00 CET**.

2 Guidelines for applicants

2.1 Who can apply?

Proposals can be submitted by researchers from any Dutch university or research institute that is affiliated with NWO or KNAW, as well as by researchers from established publicly-funded non-profit knowledge institutions in the fields of the Humanities, like the Koninklijke Bibliotheek, RKD, Nationaal Archief and Beeld en Geluid.³ Each proposal is to be formally submitted by a single named researcher (the 'principle investigator' or PI) on behalf of a team comprising researchers from one or more institutes. The PI is also the proposed project leader. The PI will be responsible for scientific progress, reporting, and financial accounting. A copy of the proposal must be submitted to the director or dean of your institute. Proposals are accepted on the understanding that your institute has been

² FTE = Full Time Equivalent; 1.0 FTE represents 1680 hours.

³ If in doubt, interested PIs can contact us to check their eligibility for funding in this call (see Section 4).

informed of the proposal, and that it accepts the conditions relating to any grant awarded in this competition.

Further requirements:

- the PI must have a PhD;
- the PI must have a proven track record and research experience relevant to the proposal;
- the PI may submit only one proposal in the capacity of PI in this call;
- the PI holds a contract (period of appointment) for at least the duration of the research for which the grant is being requested;
- the simultaneous submission of identical/very similar proposals is not permitted;
- proposals submitted to more than one competition can be awarded only once;
- the PI can only submit a full proposal in the second phase of this call if he/she also submitted an abstract in the first phase.

It is encouraged to have links with industry to enlarge the potential for project valorization. As such, co-applicants may be employed by a non-academic partner.

2.2 What can be applied for?

A project grant consists of two parts: 1) an 'in cash' contribution for the employment of local research personnel and other expenses, and 2) an 'in kind' contribution in the form of highly skilled eScience Research Engineers employed by NLeSC. The maximum project duration is 24 months.

A standard project budget will be as follows:

- **K€100** for local personnel (e.g. postdoc, UD or UHD positions, or a partial PhD student position), project-related equipment, software, and (non-NLeSC) travel expenses. Budget for local personnel should follow the most recent NWONWO-VSNU agreement, wherever applicable.⁴ It is possible to request multiple positions within the budget.
- Part of this sum may be used for material expenses, for example equipment or software, or travel expenses. The necessity of these expenses in connection with the project must be specifically justified.
- **1.5 FTE** in terms of support provided by personnel employed by NLeSC, with each 1.0 FTE representing 1680 hours. Of the total requested FTEs, 15% comprises internal overhead. The remaining 85% covers all activities performed on behalf of the project by one or more eScience Research Engineers in particular, and - to a maximum of 5% - by the assigned NLeSC Coordinator who oversees the project at NLeSC, and communication personnel.

Flexibility in the requested budget is possible, however. Applicants may request an 'in kind' NLeSC contribution between 1.0 FTE (min) and 2.0 FTE (max). For every 0.1 FTE requested above the standard 1.5 FTE, the standard in cash budget of K€100 must be

⁴ See: <http://www.nwo.nl/financiering/hoewerkt-dat/Salaristabellen> .

lowered by K€8. Alternatively, for every 0.1 FTE requested below the standard 1.5 FTE, the standard in cash budget may be raised by K€8.

eScience Research Engineers are scientists who work at the interface of various scientific disciplines and advanced ICT. They will become an integral part of the proposed research team and will focus on the development and implementation of eScience methods, technologies and software. Primarily, they will ensure that the research team will be able to make easy and effective use of the envisioned technological solutions. They will also help to interpret the results of the research and to make the delivered eScience tools useable for a broad range of users. The eScience Research Engineers perform their project activities both at NLeSC in Amsterdam and at the project locations (typically at the institute of the PI; a guideline is 60% of their time available for the project). In this way, they contribute directly to the project team whilst also providing a direct link to the expertise available throughout NLeSC and its wider networks.

The **NLeSC Coordinator** is an experienced project manager, responsible for the daily supervision of the NLeSC engineer(s) assigned to the project, and - together with the PI - in charge of monitoring progress and the delivery of project results.

Finally, NLeSC will contribute an additional 0.5 FTE in kind contribution jointly to the 3 projects rewarded. This engineer will be tasked to contribute to generalizing the software for further use in the arts and humanities and beyond and contribute to disseminating this software via eStep. This contribution does not need to be included in the budget or description of the project proposal.

2.3 When can applications be submitted?

The round consists of two phases:

1. A mandatory abstract phase, in which the main research ideas and projected outcomes are outlined. The closing date for the submission of abstracts is **15 November 2016, 14:00 CET**. We need the abstract for two reasons. In this phase the evaluation (e.g. contacting reviewers) will start already based on the abstracts which will speed up the procedure. The abstracts and names of 3 potential referees/non-referees should be submitted online at the website of the Netherlands eScience Center (see section 2.5). Also, a pre-selection will be made if the number of abstracts exceeds 20. If a pre-selection will be made, this will be done on the basis of the full proposal and after **16 December 2016**. In either case all applicants will have to submit a full proposal. A slightly adapted time-table will apply when a pre-selection phase is introduced.
2. The full proposal phase, in which a detailed application is submitted. The closing date for the submission of full proposals is **16 December 2016, 14:00 CET**. The template is provided online on the website of the Netherlands eScience Center and proposals should be submitted by email to adah@esciencecenter.nl (see section 2.5).

Further procedural information, including a timetable, is found in Section 3.1.

2.4 Specific conditions

Awarded projects must commence within six months of the award date. If the project has not started within that period, the NLeSC Board and a representative from the CLARIAH board may decide to withdraw the grant.

2.5 Preparing and submitting an application

An abstract consists of three parts: a fact sheet, the abstract form, and a list of suggested referees/non-referees. A full proposal has two parts: a fact sheet, and the application form.

- The mandatory abstract should be submitted online at the website of the Netherlands eScience Center, <http://www.esciencecenter.nl/project-calls>.
- The full proposal must be submitted as a PDF document and required application documents are to be submitted (as PDF files only) using the following email address: adah@esciencecenter.nl.
- The application documents should arrive no later than the deadlines set in Section 2.3.
- The correct templates for all application documents must be used for preparing the proposal. The template forms can be obtained from the NLeSC and CLARIAH webpages for this call.
- The list of possible referees should be added online when submitting the project abstract. Please give their full names, email address and web address. You should not propose anyone with whom you have recently collaborated or with whom you intend to collaborate in the near future, whether as co-authors or in other forms of joint undertaking. Only referees who are not directly involved in the research project and research team to which your application refers can be considered. Moreover, suggested referees must not currently hold an appointment in The Netherlands. In addition, it is also possible (but not mandatory) to give names of up to three people who should NOT be approached as referee.
- Possible letters of intent from (for example) industrial partners can be attached as separate PDF files.
- Applications must be completed in English. The layout of the full proposal should facilitate its readability. Use a font size of at least 10 points.

A copy of your proposal must be submitted to the director or dean of your institute. Proposals are accepted on the understanding that your institute has been informed of the proposal, and that it accepts the conditions of grants awarded in this call.

3 Assessment procedure

3.1 Procedure

Information event

To allow all interested applicants to get acquainted with the approach and structure of both CLARIAH and NLeSC, with the role of the eScience Research Engineers, and with the technologies being implemented and applied by both CLARIAH and NLeSC, an information event is organized at NLeSC (Science Park 140, Amsterdam) on **1 November 2016**.

Presence of at least one team member at the information event is highly recommended, but not mandatory.

Assessment committee

The assessment committee (AC) consist of independent and non-partial experts within the field of eScience and the scope of this call. The main task of the assessment committee is to review the proposals, the peer-reviews and the rebuttals, and draft an advice for the decision making board on the granting and rejection of proposals. In case of pre-selection, the assessment committee will also advise the decision making board on the pre-selected proposals, as described below.

(if applicable) Pre selection

If the number of proposals submitted in this call exceeds 20, NLeSC and Clariah will use a process of pre-selection. In the pre-selection process, the AC will assess all applications based on the criteria for this call, the proposals and the expertise of the assessment committee members. During the pre-selection process, the committee will assess whether a proposal has a chance of being granted or not. As a result, the committee will draft a motivated advice for each of the proposals on rejection and acceptance of the proposal for the full peer-review process. The advice and motivation will be communicated to the PI, which will receive the opportunity to write a reaction on the motivation. The reactions will be sent to the assessment committee, which will be asked whether the reaction leads to a reconsideration of their earlier advice. This will result in the final advice, on which the NLeSC Board supplemented by a member of the CLARIAH board will make its final decision.

Review of the proposals

All (pre-selected) proposals will be reviewed by international and independent peer-reviewers. The peer-reviewers are requested to assess the proposals based on the criteria mentioned in section 3.2.2. PI's will receive the opportunity to react to the peer-reviews in the form of a rebuttal. The assessment committee will assess the proposals in an assessment committee meeting on the quality of the proposals, using the proposal, the peer-reviews and the rebuttal. This will result in an advice to the decision making board, in the form of a motivated ranking of the proposals.

Proposals from all humanities sub-disciplines are welcomed but priority will be given to proposals from one of CLARIAH's three core disciplines: linguistics, media studies, and socio-economic history. If there are proposals of sufficient quality from one of these core disciplines, *at least one* of these will be awarded funding.

Granting Decision

The NLeSC board, supplemented with a board member from CLARIAH, will make the final decision on funding of submitted proposals, based on the recommendations of the Assessment Committee. The decision is based on consensus.

Timetable

1 November 2016	Information event
15 November 2016	Submission deadline abstracts
16 December 2016	Submission deadline full proposals

<i>End March 2017</i>	Applicants receive reviewer's comments and are given the opportunity to respond
<i>End April 2017</i>	AC evaluation and prioritization
<i>June 2017</i>	Formal granting decision by NLeSC and CLARIAH Boards
<i>June 2017</i>	Applicants informed of final decision

3.2 Admissibility and assessment criteria

3.2.1 Formal admissibility of applications

Applications are admissible only in case all of the following conditions are met:

- the proposal has been submitted by a researcher employed by a recognized institution (see Section 2.1);
- the proposal is consistent with the purpose of the call (see Section 1.3);
- the proposal was submitted using adah@esciencecenter.nl;
- the proposal was submitted before the deadline;
- the proposal meets the conditions and requirements of this call, as stated in chapter 2 of this call for proposals.

Once declared admissible, the proposal will be processed. Proposals with serious errors or omissions may be disqualified.

3.2.2 Assessment of contents

Proposals will be assessed by the referees and by the Assessment Committee on the basis of the criteria below:

Scientific quality (25%)

- the proposed research should be at the forefront of the state-of-the-art within the research of Arts and Humanities, also at an international level;
- the research team should be of the highest quality, and – if possible – already be recognized as representative (and authoritative) with respect to the proposed research questions, direction, and long-term vision.

Scientific novelty and impact (25%)

- the proposed research should be novel and represent a number of essential steps towards the long term aim of solving a specific, major scientific challenge;
- the proposed research can be expected to lead to one or more significant results (in particular in terms of a significant advancement within the Arts and Humanities);
- the proposed research should potentially change the modus operandi of scientific practice within the domain at hand, in terms of broadness, scale, speed of delivery, or otherwise;
- the proposal must indicate which efforts are made to promote the results of the project (publications, demonstrations, workshops, training, etcetera).

Relevance to CLARIAH (25%)

- the proposal must explicitly specify its relevance to CLARIAH, e.g. use instruments that are already (planned to be) incorporated in CLARIAH, contribute to the design and

construction of the infrastructure, or show how the instruments fit in the CLARIAH infrastructure and how they will be incorporated in CLARIAH so that they can be used also after the project ended.

eScience state-of-the-art, lateral impact, re-use and sustainability (25%)

- the eScience technologies (e.g., software for data analytics, data management, efficient computing, etcetera) should be sufficiently state-of-the-art, meaning that no alternative (proven) technologies exist that could serve better in solving the domain specific research questions, lead to more significant breakthroughs, or serve better in the pursuit of entirely new research questions;
- the proposal must indicate how the proposed solutions will find use beyond the proposed work itself, in particular after finalization of the project and in the context of CLARIAH and of NLeSC's eStep (See appendix B);
- the proposed solutions and (software) deliverables must be open source/open access and permit use and/or interpretation by other researchers;
- the proposal must indicate how maintenance and sustainability of project results will be secured and managed;
- the proposal must indicate how the project will build further collaborations, in science, industry, or both; inclusion of concrete letters of intent from such foreseen partners will be valued positively, but is not required.

4 Contact details

4.1 Contact

If you have any Arts- and Humanities-oriented questions in relation to this call, or questions about CLARIAH, please contact:

Prof. dr. Jan Odijk, Program Director CLARIAH
Tel.: +31 (0)30 253 5745
Email: j.odijk@uu.nl

If you have any eScience-oriented questions in relation to this call, or questions about NLeSC, please contact:

Prof. dr. Wilco Hazeleger, Director NLeSC
Tel.: +31 (0)20 460 4770
Email: adah@esciencecenter.nl

Technical questions **about** the procedures of this call, please contact:

Drs. Joep van Wijk, Program Officer NWO
Tel.: + 31 (0)70 349 4459
Email: e-science@nwo.nl

Appendix A:

The CLARIAH software infrastructure

The CLARIAH infrastructure contains data and software related to generic infrastructure services, and to its core disciplines linguistics, social economic history, and media studies.

First of all, the generic infrastructure includes web-based REST-services and generic GUIs to ingest, map, convert, curate, harvest, query, explore, visualize and export structured humanities research data. It allows scholars to combine their own local specific project datasets with generic structured (meta)data collected from across the entire CLARIAH domain. This will also include a number of selected external data providers. Data, models, ontologies and vocabularies will primarily focus on describing 'who', 'what', and 'where'. The data served through this infrastructure will be sustainable and authoritative. It will feature e.g. persons, locations and the historical development of the properties of Dutch words. Other generic structured data from the core disciplines – like e.g. 'when' events - may be included at a later stage. All data and related software are based on the Linked Open Data (LOD) paradigm. Secondly, the generic infrastructure features reliable, open and documented web-based APIs. Research teams developing custom tools for specific projects, or other third-party software providers, can fully integrate with (parts of) the CLARIAH infrastructure. This includes secondary services for user identification, user (group) management, authentication and authorization.

The linguistics part includes software to collect and create data and associated metadata, to automatically enrich or manually annotate data with a variety of linguistic properties, to incorporate enriched data in search and analysis applications, and to search for linguistically interesting phenomena in the data and to analyze the search results. A wide variety of enriched linguistic data is covered by CLARIAH.

The social economic part incorporates a wide variety of databases with structured data on social economic history. These databases are being integrated by turning them into linked open data and by harmonizing vocabularies used and making their semantics explicit. Software includes services to facilitate the conversion of data into LOD, and to search in and analyze the data in the resulting (huge) LOD databases (triple stores).

The media studies part is developing a fully integrated Media Suite based on existing independent tools. By integrating and extending the functionalities offered by the individual tools in the Media Suite they become interoperable, and with appropriate user interfaces also more user-friendly. The Media Suite offers support for resource selection and criticism, for exploration and search, for analysis and visualization, and for annotation. The data dealt with include huge amounts of audio-visual data (provided by Beeld & Geluid) but also textual data such as newspapers (via the KB) and modern social media (e.g., Twitter), as well as the metadata associated with these data.

All software created in CLARIAH is open source and is expected to be compliant with the software guidelines being developed.

More information about specific services and datasets, as well as contact persons for specific data or software, can be found on the CLARIAH components web page:

<http://www.clariah.nl/en/projects/research-pilots/components>

Appendix B:

The eScience Technology Platform (eStep)

The eScience Technology Platform (eStep)⁵, developed by NLeSC, represents the core of NLeSC's eScience competences. It forms a *generic and overarching* platform, containing software, knowledge, and eScience research. 'Generic and overarching' means that all eStep components can be applied in a broader context than just a single project. The components should be of support to multiple research efforts, an entire scientific discipline, or even multiple disciplines. All projects executed in partnership with NLeSC are expected to contribute to eStep.

One of the key goals of eStep is to scout, adopt, study, develop, integrate and make available an



extensive and stable set of advanced scientific software technologies. These can take the form of algorithms, compute kernels, interfaces, libraries, tools, scientific workflows and applications. This way, eStep helps to manage large collections of data, utilize advanced compute infrastructures, storage facilities, high-speed networks, high-resolution visualization equipment, and instruments.

eStep explicitly aims to promote the exchange and re-use of best practices and to prevent fragmentation and duplication. The eStep software base is NLeSC's facility for managing and disseminating software and work-flows employed in its project portfolio, along with supporting documentation, demonstrations and training resources. All NLeSC eScience Research Engineers contribute to eStep by generalizing the technologies built in their projects.

eStep contains software that is developed in-house at NLeSC, but also externally developed software that we reuse and have expertise in applying to challenging scientific problems. This allows NLeSC to also provide integrated turn-key solutions for scientists, regardless of whether the software is developed in-house or externally. A key idea behind eStep is to have high-level, and sometimes domain-specific, solutions on top of generic low-level libraries, thus maximizing software re-use.

It must be stressed that the technological developments undertaken by and with NLeSC are not aimed at realizing benefits for NLeSC itself. All developments are in support of the scientific goals of the research project, with the additional aim to also serve other research communities as much as possible, now and in the future.

⁵ See also: estep.esciencecenter.nl and <http://www.esciencecenter.nl/technology/>

Appendix C:

In this Call for Proposals, all applicants are asked to indicate the project's e-Infrastructure needs, in terms of compute hours, data storage capacity, lightpath connectivity, or otherwise. NLeSC will apply a 'use-or-explain' policy, meaning that

- projects *without* e-Infrastructure needs are asked to give a brief explanation;
- projects with clear e-Infrastructure needs are expected to select the hardware resources and services as part of the Dutch National e-Infrastructure as first option, and to indicate the expected extent of use;
- projects with clear e-Infrastructure needs that aim to use international (e.g. PRACE, XSEDE, etcetera) or commercial (e.g. web, cloud, etcetera) hardware and services instead are required to give a brief explanation.

The use of the Dutch National e-Infrastructure is not a requirement, nor is it a formal review criterion. However, in all cases in which the Dutch National e-Infrastructure is not used, a justification should be provided.

The Dutch National e-Infrastructure

In this call, NLeSC defines the Dutch National e-Infrastructure as follows:

all publicly-funded hardware resources (e.g. compute, data, visualization, networking, etcetera) and directly connected support services (people, software), set up and maintained with the aim to support publicly-funded research in The Netherlands, and made available to either all or a selected subset of all researchers from a.o. Dutch universities and research institutes affiliated with NWO or KNAW.

The definition distinguishes between hardware resources and services available to all researchers in The Netherlands (Category I), and those made available to a selected subset (Category II). The Category I e-Infrastructure, outlined below, is formed by the hardware resources and services provided and maintained by SURFsara, SURFnet, DANS, and – in part – also by Nikhef and RUG-CIT.

The Category II e-Infrastructure is formed by all other hardware resources and services that are accessible to a selected group of researchers following thematic or geographic criteria. Examples of such infrastructures include the Life Science Grid (LSG), the Distributed ASCI Supercomputer (DAS), and the many stand-alone local facilities at various universities (e.g. the Peregrine cluster (RUG-CIT), the GPFS data storage facilities (Target), the WUR HPC Cluster (Wageningen), etcetera).

Overview: Category I e-Infrastructure

While it is impossible to provide a complete overview of all resources part of the Dutch National e-Infrastructure in this call text, the following provides entrance points to the major Category I e-Infrastructure resources and services. For more information, it is advised to contact the organizations and institutes responsible for these resources directly, in particular SURF:

<https://www.surf.nl/en/contact.html> .

Compute Resources and Services

- Cartesius: National supercomputer for maximum performance
<https://www.surf.nl/en/services-and-products/cartesius/index.html>
- HPC Cloud: Complete control over your own computing infrastructure
<https://www.surf.nl/en/services-and-products/hpc-cloud/index.html>
- Grid: Distributed computing system for fast processing of large data sets
<https://www.surf.nl/en/services-and-products/grid/index.html>
- Hadoop cluster: Big data processing and analysis
<https://www.surf.nl/en/services-and-products/big-data-services/index.html>

Data Resources and Services

- BeeHub: Easily save and share large volumes of data
<https://www.surf.nl/en/services-and-products/beehub/index.html>
- Data Archive: Secure long-term storage of research data on tape
<https://www.surf.nl/en/services-and-products/data-archive/index.html>
- DataverseNL: Store, share and register research data online
<http://www.dans.knaw.nl/en/about/services/archiving-and-reusing-data/DataverseNL>
- EASY: Online archiving, depositing and downloading of research data
<http://www.dans.knaw.nl/en/about/services/archiving-and-reusing-data/easy>

Networking Resources and Services

- Lightpaths (SURFlichtpaden): Ultra-fast and high capacity connectivity
<https://www.surf.nl/en/services-and-products/surflichtpaden/index.html>

Cloud and Collaborative Resources and Services

- SURFconext: Online collaboration and services in a single environment
<https://www.surf.nl/en/services-and-products/surfconext/index.html>
- SURFdrive; Personal and secure cloud storage, synchronization, sharing
<https://www.surf.nl/en/services-and-products/surfdrive/surfdrive.html>

Visualization Resources and Services

- Remote visualization: Visualize large datasets on your desktop
<https://www.surf.nl/en/services-and-products/visualisation/index.html>
- Collaboratorium: Sophisticated presentation and visualization aids
<https://www.surf.nl/en/services-and-products/collaboratorium/index.html>

SURF

For a complete overview of all Category I services provided by SURF, see:
<https://www.surf.nl/en/services-and-products>

DANS

For a complete overview of all Category I services provided by DANS, see:
<http://www.dans.knaw.nl/en/about/services>